# Quarks and gluons in the Lund plane(s)

Gregory Soyez, with Frederic Dreyer, Andrew Lifson, Gavin Salam and Adam Takacs

based on arXiv:1807.04758, arXiv:2007.06578 and arXiv:2112.09140

IPhT, CNRS, CEA Saclay

CERN, June 3 2022

# Motivation

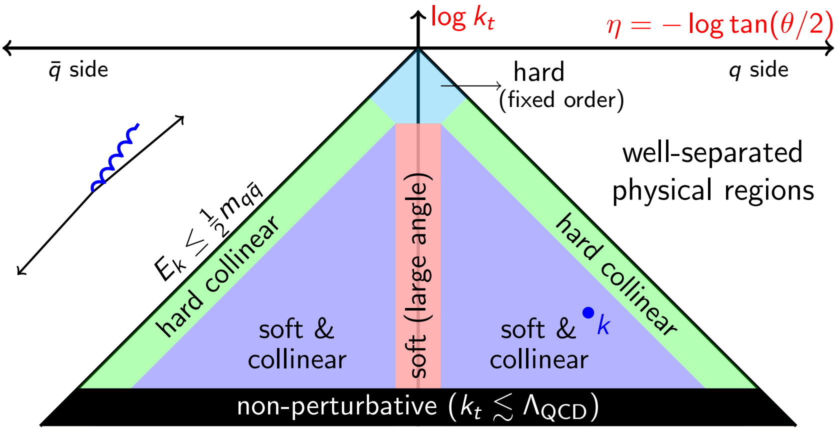
Your mere presence probably means you know this... but just in case:

- study (Lund-plane) tagging performance for “simple” (“1-prong”) objects
- hope to get better control than for more complex systems ($W/Z/H$, $t$, ...)
- many potential pheno applications (BSM searches, VBF, $H \to gg$, ...)

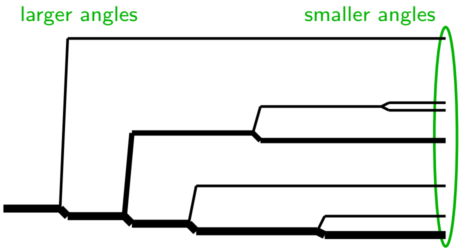
# The tools: Lund planes and trees diagrams

# Basic features of QCD radiations: the Lund plane

Lund plane: natural representation uses the 2 "log" variables $\eta$ and $\log k_\perp$

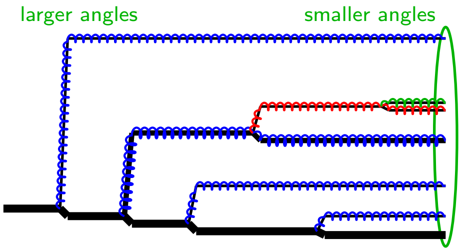


# The Lund plane(s) in practice: cluster with Cambridge/Aachen



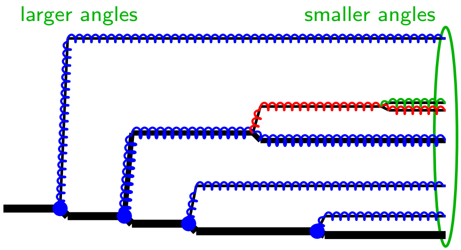

- closely follows our beloved angular ordering

# The Lund plane(s) in practice: cluster with Cambridge/Aachen

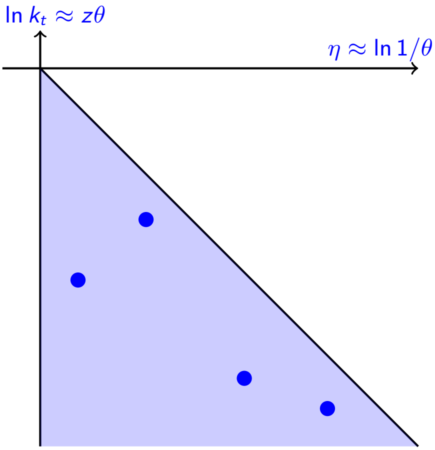



- closely follows our beloved angular ordering
- i.e. mimics partonic cascade

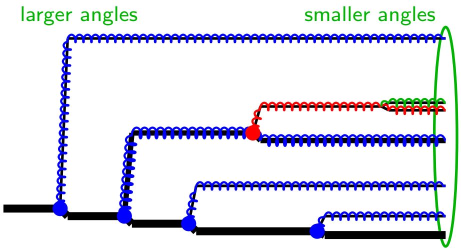
# The Lund plane(s) in practice: cluster with Cambridge/Aachen



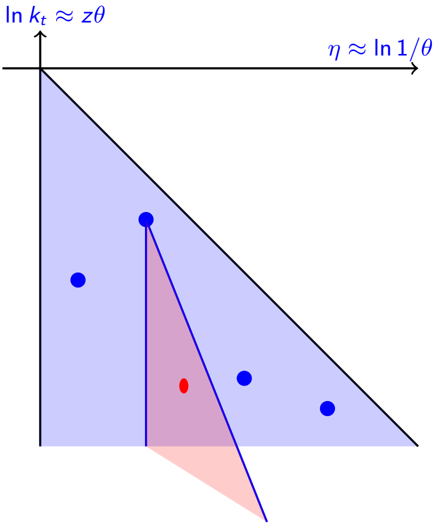

- closely follows our beloved angular ordering
- i.e. mimics partonic cascade
- can be organised in Lund planes
  - primary



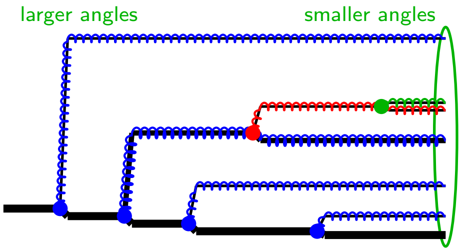
# The Lund plane(s) in practice: cluster with Cambridge/Aachen

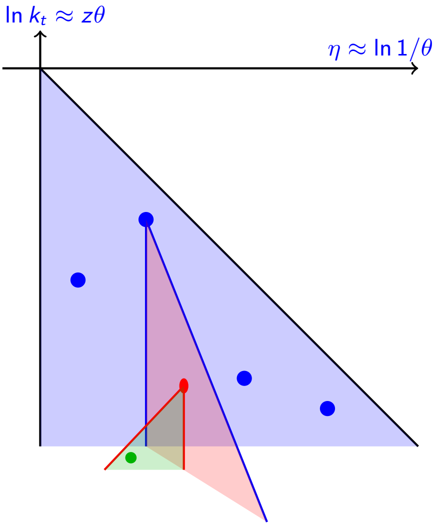



- closely follows our beloved angular ordering
- i.e. mimics partonic cascade
- can be organised in Lund planes
  - primary
  - secondary



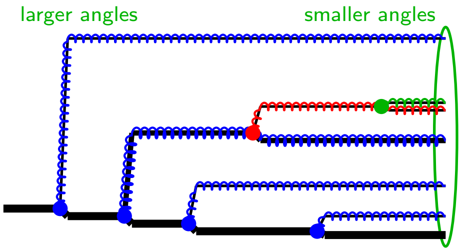
# The Lund plane(s) in practice: cluster with Cambridge/Aachen



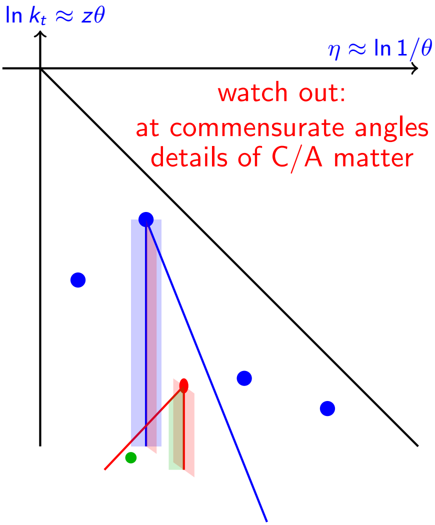



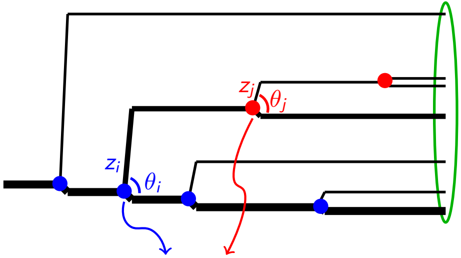

- closely follows our beloved angular ordering
- i.e. mimics partonic cascade
- can be organised in Lund planes
  - primary
  - secondary
  - ...

# The Lund plane(s) in practice: cluster with Cambridge/Aachen




- closely follows our beloved angular ordering
- i.e. mimics partonic cascade
- can be organised in Lund planes
  - primary
  - secondary
  - ...






$\mathcal{T}_i \equiv \{\theta_i, k_{t,i}, z_i, \psi_i, m_i, \dots\}$

for jets in $pp$: (similar for $ee$ events)

$\eta = -\ln \Delta R$

$k_t = p_{t,\text{soft}} \Delta R$ $\quad z = \frac{p_{t,\text{soft}}}{p_{t,\text{parent}}}$

$\psi \equiv$ azimuthal angle

**Two different Lund ($\mathcal{L}$) structures**

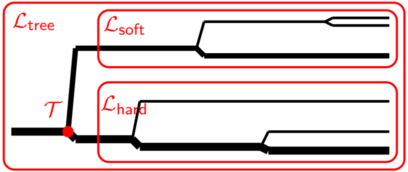
"primary plane" (follow hard branch)

$\mathcal{L}_{\text{prim}} \equiv \{\mathcal{T}_i\}$

OR

full (de-)clustering tree

$\mathcal{L}_{\text{tree}} \equiv \{\mathcal{T}, \mathcal{L}_{\text{hard}}, \mathcal{L}_{\text{soft}}\}$




Recall: $k_t > t_{t,\text{min}} \rightarrow$ **perturbative**

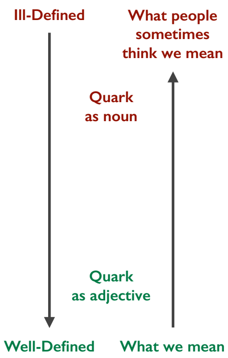
# Quark/gluon discrimination

**Goal: using the Lund declustering info (primary or full-tree) can we say if a jet is quark- or gluon-initiated?**

# Quark v. gluon jets: 0. basic considerations

## What is a Quark Jet?

*From lunch/dinner discussions*




A quark parton

A Born-level quark parton

The initiating quark parton in a final state shower

An eikonal line with baryon number 1/3 and carrying triplet color charge

A quark operator appearing in a hard matrix element in the context of a factorization theorem

A parton-level jet object that has been quark-tagged using a soft-safe flavored jet algorithm (automatically collinear safe if you sum constituent flavors)

A phase space region (as defined by an unambiguous hadronic fiducial cross section measurement) that yields an enriched sample of quarks (as interpreted by some suitable, though fundamentally ambiguous, criterion)

[Les Houches Phys at TeV colliders, 2017]

### pedestrian summary

- there is no such thing as a "quark" or a "gluon" jet
- well-defined: tagging process A ("quark-enriched"$^{(*)}$) against process B ("gluon-enriched"$^{(*)}$)

$^{(*)}$ ambiguous

### Our approach(es)

- discuss process-independent aspects (at least analytically)
- probe changes for different processes

# Quark v. gluon jets: I. approach

**Optimal discriminant** (Neyman–Pearson lemma)

$$\mathbb{L}_{\text{prim,tree}} = \frac{p_g(\mathcal{L}_{\text{prim,tree}})}{p_q(\mathcal{L}_{\text{prim,tree}})}$$

# Quark v. gluon jets: I. approach

**Optimal discriminant** (Neyman–Pearson lemma)

$$\mathbb{L}_{\text{prim,tree}} = \frac{p_g(\mathcal{L}_{\text{prim,tree}})}{p_q(\mathcal{L}_{\text{prim,tree}})}$$

**Approach #1**

Deep-learn $\mathbb{L}_{\text{prim,tree}}$

LSTM with $\mathcal{L}_{\text{prim}}$ or Lund-Net with $\mathcal{L}_{\text{tree}}$

# Quark v. gluon jets: I. approach

## Optimal discriminant (Neyman–Pearson lemma)

$$\mathbb{L}_{\text{prim,tree}} = \frac{p_g(\mathcal{L}_{\text{prim,tree}})}{p_q(\mathcal{L}_{\text{prim,tree}})}$$

## Approach #1

Deep-learn $\mathbb{L}_{\text{prim,tree}}$

LSTM with $\mathcal{L}_{\text{prim}}$ or Lund-Net with $\mathcal{L}_{\text{tree}}$

## Approach #2

**Use pQCD to** *calculate* $p_{q,g}(\mathcal{L}_{\textbf{prim,tree}})$

- Consider $k_t \geq k_{t,\text{cut}}$ to stay perturbative
- Resum logs to all orders in $\alpha_s$, up to double logs
  - Each primary radiation comes with a factor $\frac{2\alpha_s(k_t)C_R}{\pi}$
  - Each subsidiary radiation comes with a factor $\frac{2\alpha_s(k_t)C_A}{\pi}$
- Probabilities: $p_{q,g} = \prod_{i\in\text{prim}} \frac{2\alpha_s(k_{ti})C_{F,A}}{\pi} \prod_{i\in\text{others}} \frac{2\alpha_s(k_{ti})C_A}{\pi}$ (up to a negligible Sudakov)
- The ratio largely cancels: $\mathbb{L}_{\text{prim,tree}} = \left(\frac{C_F}{C_A}\right)^{n_{\text{prim}}}$ [C.Frye,A.Larkoski,J.Thaler,1704.06266]
- The optimal discriminant is the primary multiplicity i.e. the Iterated SoftDrop multiplicity

# Quark v. gluon jets: I. approach

**Optimal discriminant (Neyman–Pearson lemma)**

$$\mathbb{L}_{\text{prim,tree}} = \frac{p_g(\mathcal{L}_{\text{prim,tree}})}{p_q(\mathcal{L}_{\text{prim,tree}})}$$

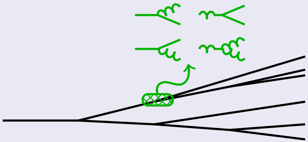
**Approach #1**

Deep-learn $\mathbb{L}_{\text{prim,tree}}$

LSTM with $\mathcal{L}_{\text{prim}}$ or Lund-Net with $\mathcal{L}_{\text{tree}}$

**Approach #2**

**Use pQCD to** *calculate* $p_{q,g}(\mathcal{L}_{\textbf{prim,tree}})$



- Consider $k_t \geq k_{t,\text{cut}}$ to stay perturbative
- Resum logs to all orders in $\alpha_s$, up to single logs
  - single logs from “DGLAP” collinear splittings

$$P_q(\mathcal{L}_{\text{parent}}) = S_q(\Delta_{\text{prev}}, \Delta)\left[\tilde{P}_{qq}(z)p_q(\mathcal{L}_{\text{hard}})p_g(\mathcal{L}_{\text{soft}}) + \tilde{P}_{gq}(z)p_g(\mathcal{L}_{\text{hard}})p_q(\mathcal{L}_{\text{soft}})\right]$$

$$p_g(\mathcal{L}_{\text{parent}}) = S_g(\Delta_{\text{prev}}, \Delta)\left[\tilde{P}_{gg}(z)p_g(\mathcal{L}_{\text{hard}})p_g(\mathcal{L}_{\text{soft}}) + \tilde{P}_{qg}(z)p_q(\mathcal{L}_{\text{hard}})p_q(\mathcal{L}_{\text{soft}})\right]$$

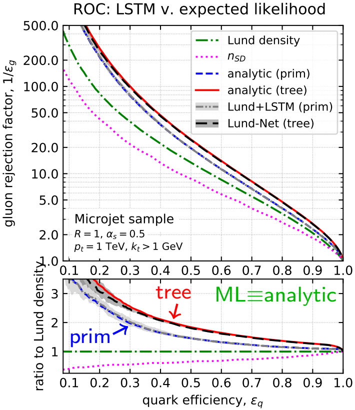
  - some single logs for emissions at commensurate angles
    Note: all-order not tractable analytically; we resum any *pair* of commensurate-angle emissions
  - running coupling (in the Sudakov)

# Quark v. gluon jets: II. ML validation

our analytic discriminant is exact/optimal in the dominant collinear limit $\theta_1 \gg \theta_2 \gg \cdots \gg \theta_n$

$\Rightarrow$ ML expected to give the same performance our analytic discriminant is exact/optimal in the dominant collinear limit $\theta_1 \gg \theta_2 \gg \cdots \gg \theta_n$

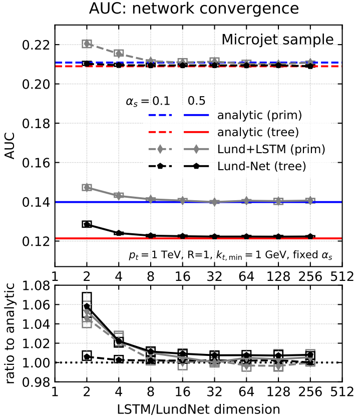
$\Rightarrow$ ML expected to give the same performance




ROC curves agree

Microjet
≡
exact pure-collinear

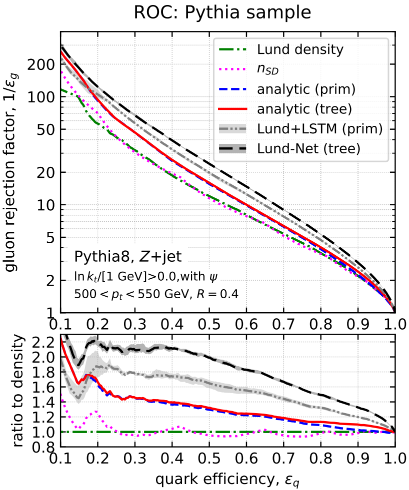
[M.Dasgupta,F.Dreyer G.P.Salam,G.Soyez, 1411.5182]




Converges for large-enough networks

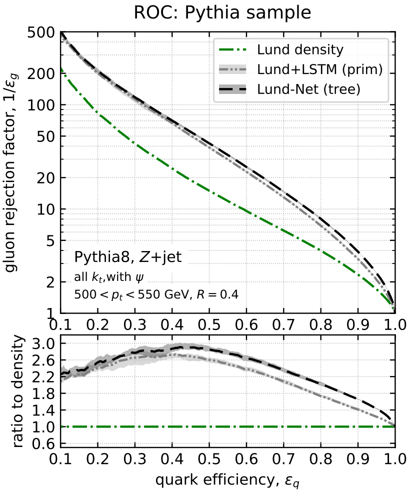
# Quark v. gluon jets: III. performance

$pp \to Zq$ v. $pp \to Zg$ $\quad (p_t \sim 500$ GeV, $R = 0.4)$




- clear performance ordering:
  1. Lund+ML > Lund analytic > ISD
  2. tree > prim

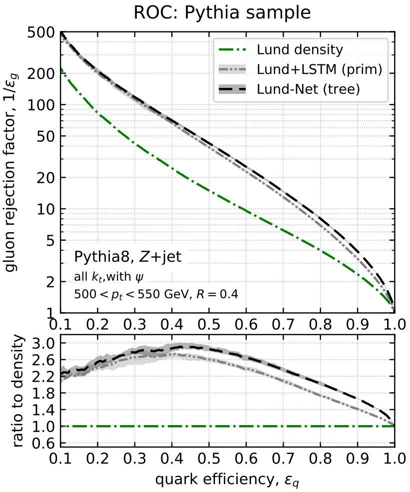
$pp \to Zq$ v. $pp \to Zg$ $\quad (p_t \sim 500$ GeV, $R = 0.4)$




- clear performance ordering:
  1. Lund+ML > Lund analytic > ISD
  2. tree > prim
- larger gains with no $k_t$ cut

# Quark v. gluon jets: III. performance

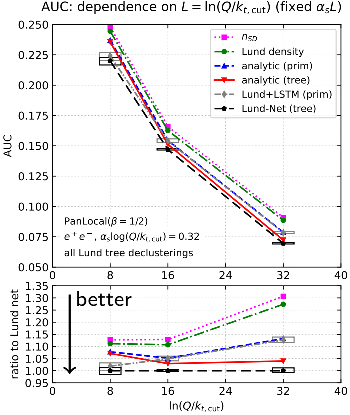
$pp \to Zq$ v. $pp \to Zg$ $\quad (p_t \sim 500$ GeV, $R = 0.4)$



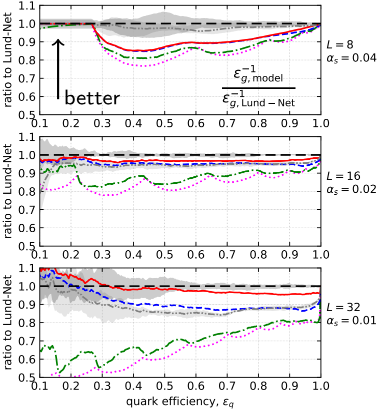

- clear performance ordering:
  1. Lund+ML > Lund analytic > ISD
  2. tree > prim
- larger gains with no $k_t$ cut
- Interesting questions:
  - Analytic approach to NP?
  - Apply analytics to other systems ($W/Z/H$, top)

# Quark v. gluon jets, part IV: towards asymptotics

Ares Under Curve:
lower is better




gluon rejection:
higher is better

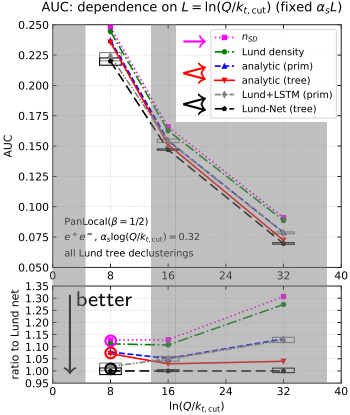



**Idea**

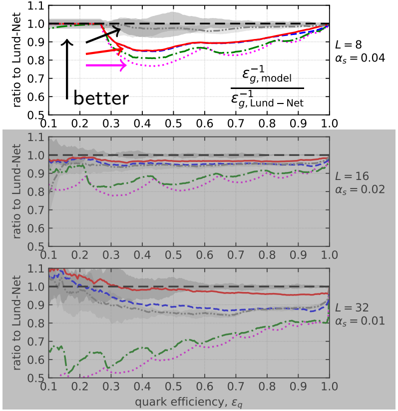
Asymptotics towards NLL

$\alpha_s L = \mathrm{cst},\ \alpha_s \to 0\ (L \to \infty)$

Ares Under Curve:
lower is better




gluon rejection:
higher is better




**Idea**

Asymptotics towards NLL

$\alpha_s L = \text{cst}$, $\alpha_s \to 0$ ($L \to \infty$)

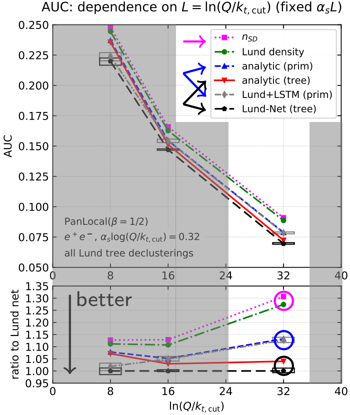
**Larger $\alpha_s$ (lower $L$)**

ML > analytics > $n_{SD}$

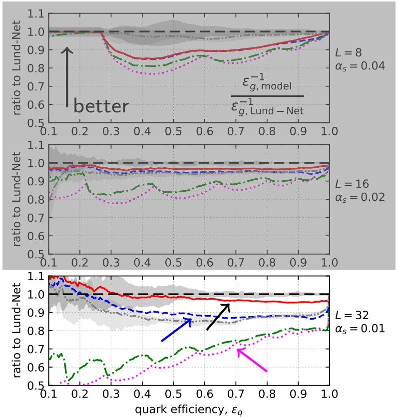
little help beyond primary

# Quark v. gluon jets, part IV: towards asymptotics

Ares Under Curve:
lower is better




gluon rejection:
higher is better




## Idea

Asymptotics towards NLL
$\alpha_s L = \text{cst}$, $\alpha_s \to 0$ ($L \to \infty$)

## Larger $\alpha_s$ (lower $L$)

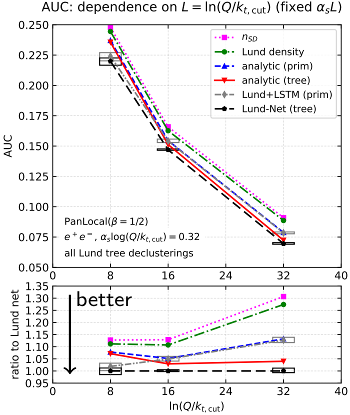
ML > analytics > $n_{SD}$
little help beyond primary

## Larger $\alpha_s$ (lower $L$)

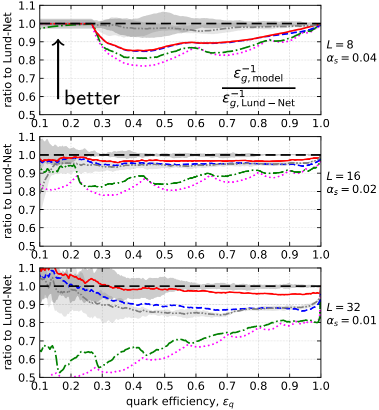
tree > primary > $n_{SD}$
ML ≈ analytics

# Quark v. gluon jets, part IV: towards asymptotics

Ares Under Curve:
lower is better




gluon rejection:
higher is better




**Idea**

Asymptotics towards NLL
$\alpha_s L = \text{cst}$, $\alpha_s \to 0$ ($L \to \infty$)

**Larger $\alpha_s$ (lower $L$)**

ML > analytics > $n_{SD}$
little help beyond primary

**Larger $\alpha_s$ (lower $L$)**

tree > primary > $n_{SD}$
ML ≈ analytics

develop accurate
parton-showers for ML

# Resilience (1/2)

Question: is your tagger resilient to uncontrolled effects?

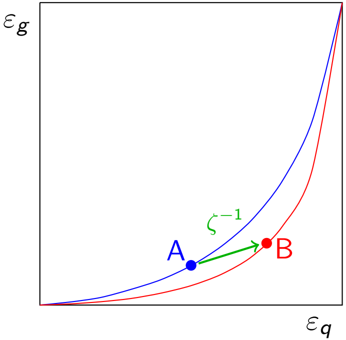
One has:

- a reference sample A
  (e.g. network trained+tested w Pythia)
- an alternate sample B
  (e.g. network tested w Herwig)

We want (for a given working point)

$$\zeta = \left[\left(\frac{\Delta\varepsilon_q}{\langle\varepsilon_q\rangle}\right)^2 + \left(\frac{\Delta\varepsilon_g}{\langle\varepsilon_g\rangle}\right)^2\right]^{-1}$$

as large as possible.




(would probably deserve a study on its own)

**Question: is your tagger resilient to uncontrolled effects?**

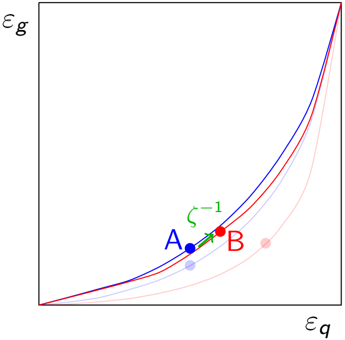
One has:

- a reference sample A
  (e.g. network trained+tested w Pythia)
- an alternate sample B
  (e.g. network tested w Herwig)

We want (for a given working point)

$$\zeta = \left[\left(\frac{\Delta\varepsilon_q}{\langle\varepsilon_q\rangle}\right)^2 + \left(\frac{\Delta\varepsilon_g}{\langle\varepsilon_g\rangle}\right)^2\right]^{-1}$$

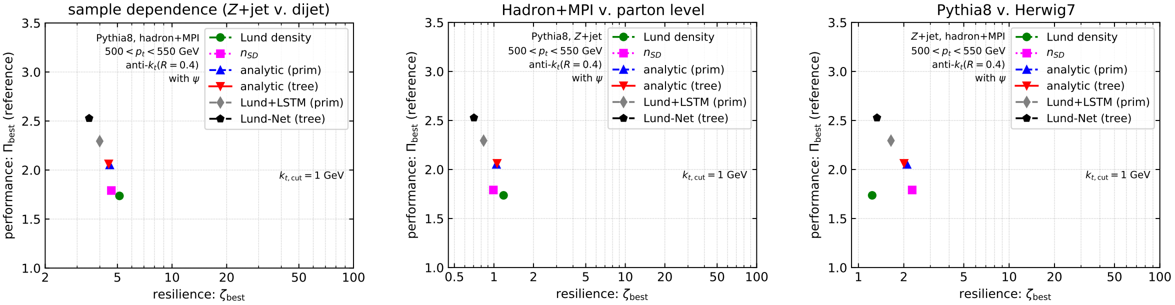
as large as possible.




Less performant
More resilient

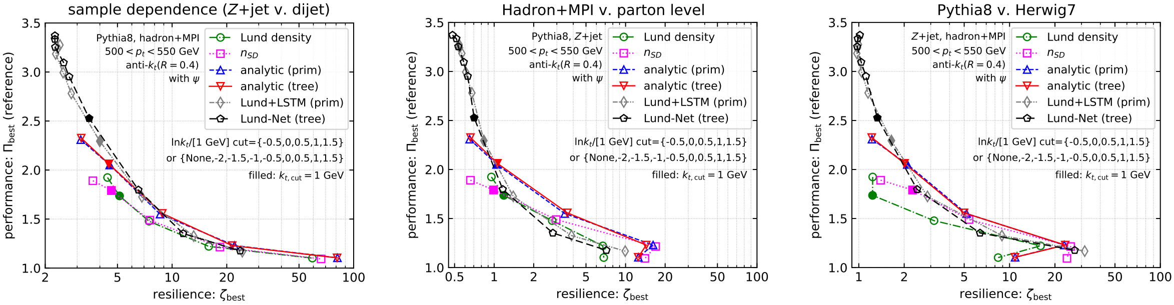
(would probably deserve a study on its own)

# Resilience (2/2)



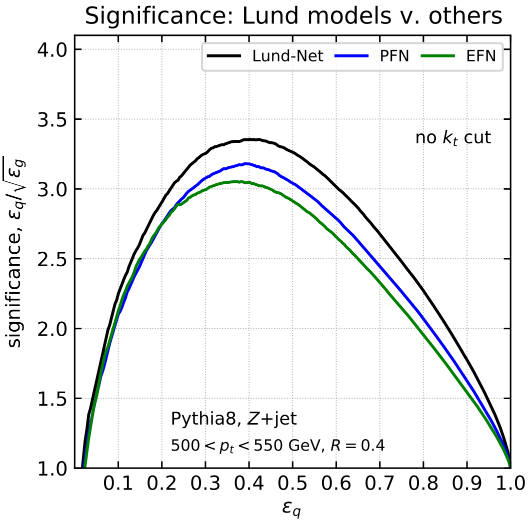

- performance $= \varepsilon_q / \sqrt{\varepsilon_g}$
- working point: $k_{t,\text{cut}} = 1$ GeV, optimal performance (reference: Pythia, hadron+MPI, $Z$+jet)
- 3 studies: sample ($Z$+jet v. dijets), NP effects (hadron v. parton), generator (Pythia v. Herwig)
- performance: same ordering as before
- resilience: network-based < Lund analytics $\lesssim n_{SD}$

# Resilience (2/2)




- same, varying $k_{t,\mathrm{cut}}$
- for each curve: "standard" trade-off between performance and resilience
- Overall: better behaviour for the new Lund-based approaches:
  - At "large" resilience: better envelope for the Lund analytic approaches
  - At "small" resilience: ML performance gain pays off

# Comparison to other approaches: ML-based

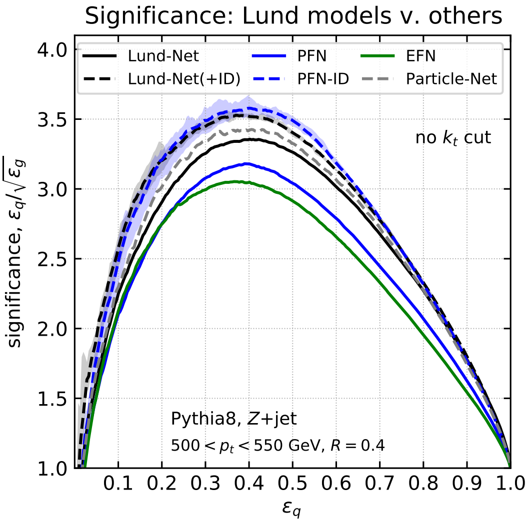



**Approaches:**

- Lund-Net (full tree)
- Particle-flow network
- Energy-flow network

- small performance gain for Lund
- differences might come from details

# Comparison to other approaches: ML-based

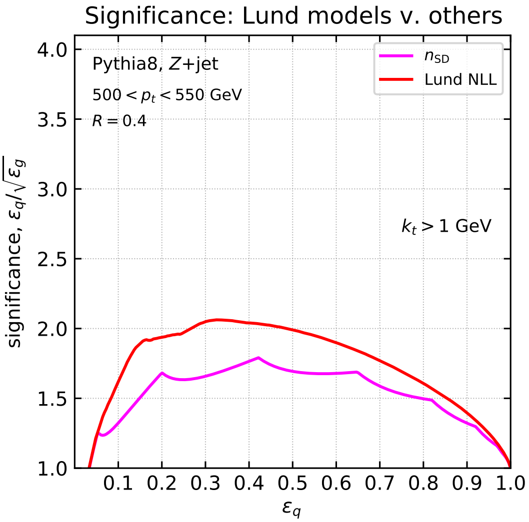



**Approaches:**

- Lund-Net (full tree)
- Particle-flow network
- Energy-flow network
- Dashed: with PDG-ID
- Particle-Net

- small performance gain for Lund
- differences might come from details
- with PDG-ID: PFN$\sim$Lund$\gtrsim$PNet

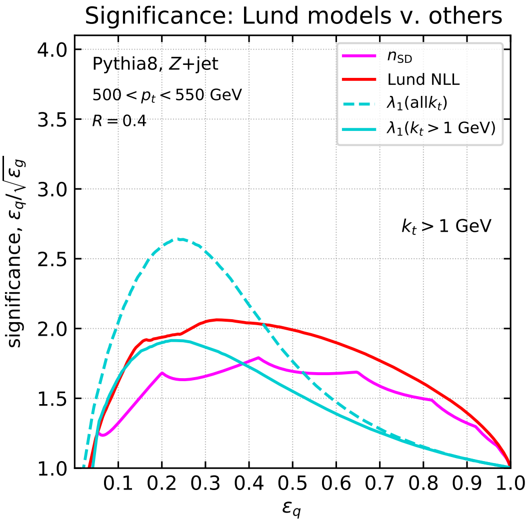
# Comparison to other approaches: analytics/shapes




**Approaches:**

- ISD mult ($n_{SD}$)
- Lund (full tree, analytic)

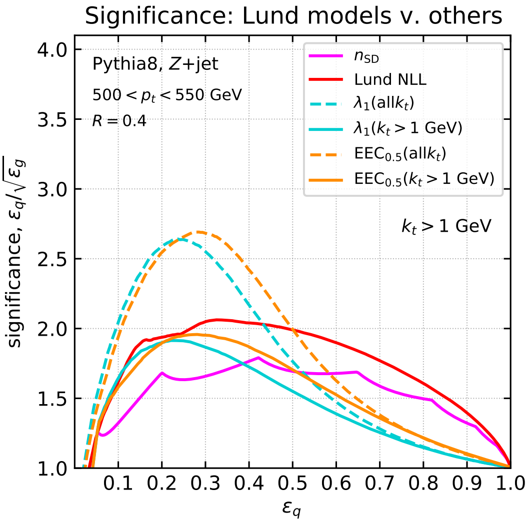
▶ clear gain from our analytic approach




**Approaches:**

- ISD mult ($n_{SD}$)
- Lund (full tree, analytic)
- width ($\sum_i p_{ti} \Delta R_i$)
- Dashed: use subjets with $k_t > 1$ GeV

- clear gain from our analytic approach
- Different behaviour for shapes
- Lund (expectably) better for same info



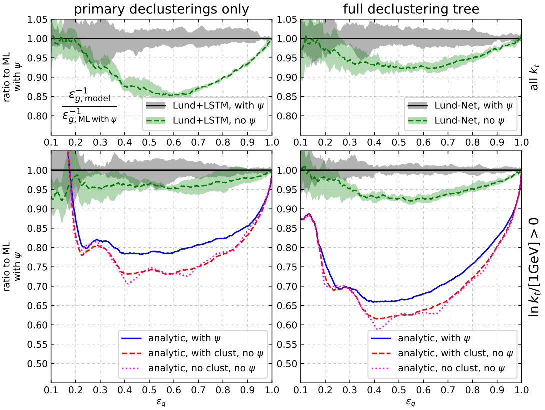

**Approaches:**

- ISD mult ($n_{SD}$)
- Lund (full tree, analytic)
- width ($\sum_i p_{ti} \Delta R_i$)
- EE correlation ($\sum_{i,j} p_{ti} p_{tj} \Delta R_{ij}^{\beta}$)
- Dashed: use subjets with $k_t > 1$ GeV

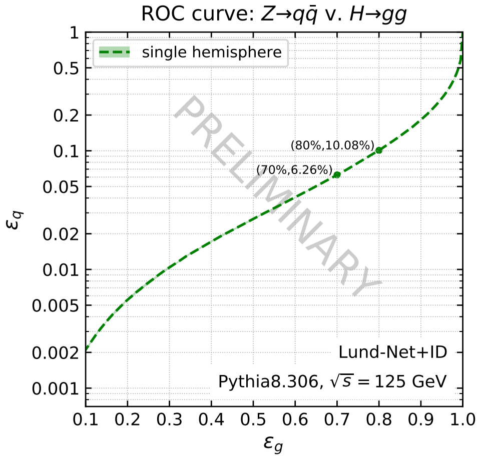
- clear gain from our analytic approach
- Different behaviour for shapes
- Lund (expectably) better for same info

# Effect of phi (& clustering logs)

- Just simple points (partially connected to the discussion yesterday)
- Some gain obtained by including $\phi$ info
- analytics:
  appear for commensurate angles
  $\leftrightarrow$ clustering logs
  only partially taken into account



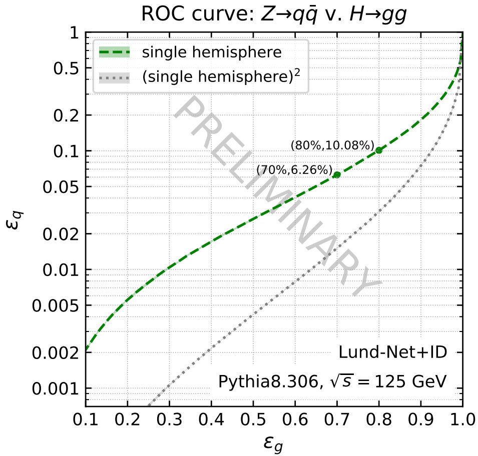
$$e^+e^- \to Z \to q\bar{q} \text{ v. } e^+e^- \to H \to gg \qquad (\sqrt{s} = 125 \text{ GeV, no ISR})$$




observed performance:

- for reference: $g$-tag on a single hemisphere

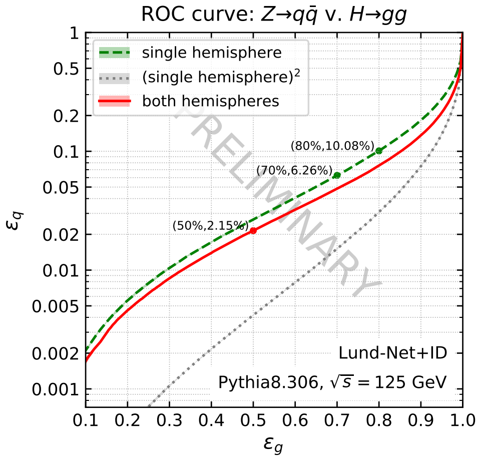
$e^+e^- \to Z \to q\bar{q}$ v. $e^+e^- \to H \to gg$ $\qquad (\sqrt{s} = 125$ GeV, no ISR)




observed performance:

- for reference: $g$-tag on a single hemisphere
- for reference: 2 hemispheres assumed independent

$e^+e^- \to Z \to q\bar{q}$ v. $e^+e^- \to H \to gg$ $\quad (\sqrt{s} = 125$ GeV, no ISR)



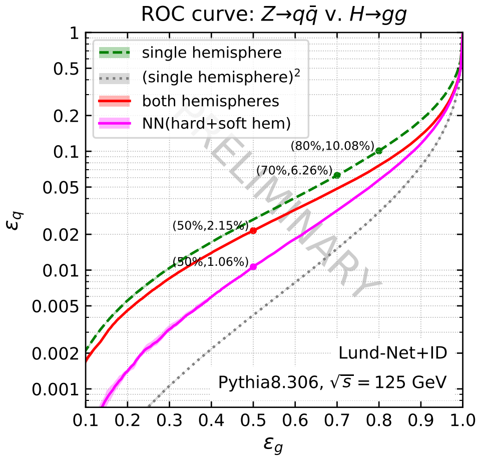

observed performance:

- for reference: $g$-tag on a single hemisphere
- for reference: 2 hemispheres assumed independent
- $g$-tag on both hemispheres
  i.e. both jets should be tagged

  full event clearly worse that (jet)$^2$

# Towards full-event tagging

$e^+e^- \to Z \to q\bar{q}$ v. $e^+e^- \to H \to gg$ $\quad (\sqrt{s} = 125$ GeV, no ISR)



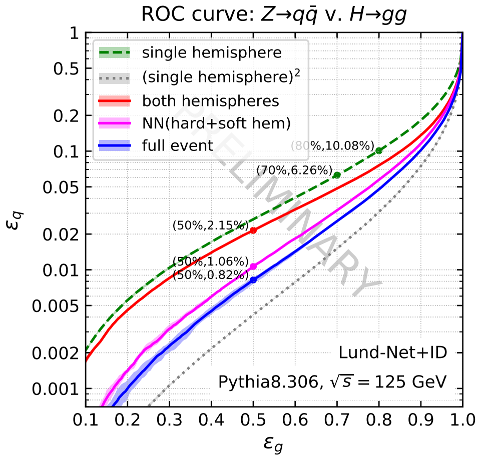

observed performance:

- for reference: $g$-tag on a single hemisphere
- for reference: 2 hemispheres assumed independent
- $g$-tag on both hemispheres
- ML on the 2 hemisphere LundNet scores
  train separately on hard & soft hemispheres
  use another NN (or MVA) to combine the two

  clear performance gain

# Towards full-event tagging

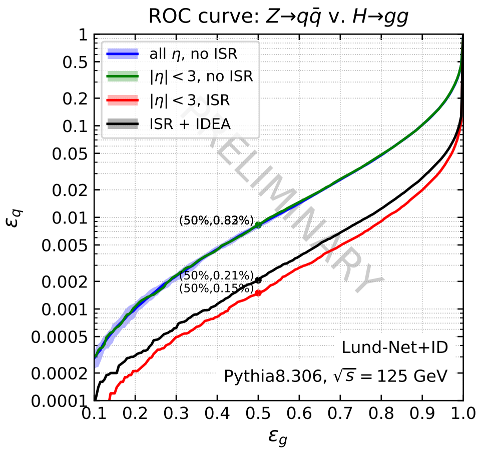
$e^+e^- \to Z \to q\bar{q}$ v. $e^+e^- \to H \to gg$ $\quad(\sqrt{s} = 125$ GeV, no ISR)




observed performance:

- for reference: $g$-tag on a single hemisphere
- for reference: 2 hemispheres assumed independent
- $g$-tag on both hemispheres
- ML on the 2 hemisphere LundNet scores
- Lund-Net for the full event

$$e^+e^- \to Z \to q\bar{q} \text{ v. } e^+e^- \to H \to gg \qquad (\sqrt{s} = 125 \text{ GeV, no ISR})$$

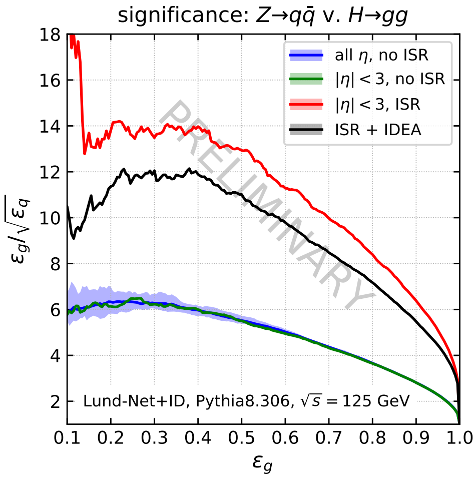



observed performance:

- for reference: $g$-tag on a single hemisphere
- for reference: 2 hemispheres assumed independent
- $g$-tag on both hemispheres
- ML on the 2 hemisphere LundNet scores
- Lund-Net for the full event
- incl. ISR and detector ⇒ further improvement

# Towards full-event tagging

$e^+e^- \to Z \to q\bar{q}$ v. $e^+e^- \to H \to gg$ $\quad (\sqrt{s} = 125$ GeV, no ISR)

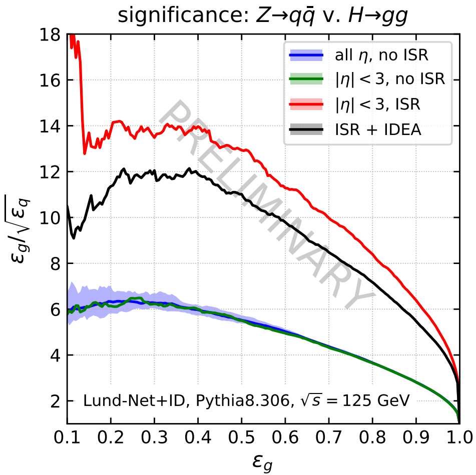



observed performance:

- for reference: $g$-tag on a single hemisphere
- for reference: 2 hemispheres assumed independent
- $g$-tag on both hemispheres
- ML on the 2 hemisphere LundNet scores
- Lund-Net for the full event
- incl. ISR and detector $\Rightarrow$ further improvement
- All in all: significance gain $\sim 12$

# Towards full-event tagging

$e^+e^- \to Z \to q\bar{q}$ v. $e^+e^- \to H \to gg$ $\quad (\sqrt{s} = 125$ GeV, no ISR)




observed performance:

- for reference: $g$-tag on a single hemisphere
- for reference: 2 hemispheres assumed independent
- $g$-tag on both hemispheres
- ML on the 2 hemisphere LundNet scores
- Lund-Net for the full event
- incl. ISR and detector ⇒ further improvement
- All in all: significance gain $\sim 12$

- Clear gain from full-event tagging
- Applications to other cases (e.g. at the LHC)?

# Conclusions

1. q/g tagging can be addressed both analytically and with ML tools
   - rich structures in both cases
   - overall a detailed degree of understanding emerging
   - analytic: single-log gives a systematic improvement over ISD multiplicity
   - deep-learning: Lund-Net shows very good performance (also for $W$ and top tagging)
2. Future directions:
   - Analytic approach for other cases than $q/g$?
     more complex (e.g. how does one treat the mass resolution for heavy bosons?)
     $b$-jet tagging might be interesting/easier
   - Towards event-wide tagging
   - higher accuracy, e.g. through more accurate (parton) showers
   - improved understanding of non-perturbative contributions

# Final words

## Conclusions from a Lund talk at CERN a year ago:

1. Lund diagrams have helped thinking about resummation and MCs
   Now they can be reconstructed in practice
   - They provide a view of a jet/event which mimics angular ordering
   - They provide a separation between different physical effects

2. Broad spectrum of applications:
   - Wide range of possible (p)QCD calculations
     Main limitation: (non-global) clustering logs; can we apply grooming-like techniques?
   - Large scope for crafting new observables ((p)QCD calculations, MC devel/validation)
   - More connections to deep learning, heavy-ion collisions, ...

...

**This connects very well to the nice list of talks we have had throughout the week!**

**Thanks to all for the participation!**

# Backup

# Promoting to a practical tool

**Construct the Lund tree in practice: use the Cambridge(/Aachen) algorithm**

Main idea: Cambridge(/Aachen) preserves angular ordering

## $e^+e^-$ collisions

1. Cluster with Cambridge ($d_{ij} = 2(1 - \cos\theta_{ij})$)
2. For each (de)-clustering $j \leftarrow j_1 j_2$:

$$\eta = -\ln\theta_{12}/2$$
$$k_t = \min(E_1, E_2)\sin\theta_{12}$$
$$z = \frac{\min(E_1,E_2)}{E_1+E_2}$$
$$\psi \equiv \text{some azimuth},\ldots$$

## Jet in $pp$

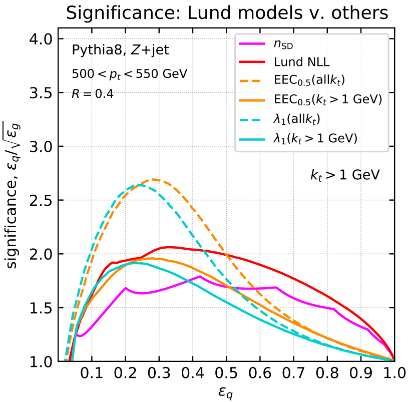
1. Cluster with Cambridge/Aachen ($d_{ij} = \Delta R_{ij}$)
2. For each (de)-clustering $j \leftarrow j_1 j_2$:

$$\eta = -\ln\Delta R_{12}$$
$$k_t = \min(p_{t1}, p_{t2})\Delta R_{12}$$
$$z = \frac{\min(p_{t1},p_{t2})}{p_{t1}+p_{t2}}$$
$$\psi \equiv \text{some azimuth},\ldots$$

## Primary Lund plane

**Starting from the jet, de-cluster following the "hard branch" (largest $E$ or $p_t$)**

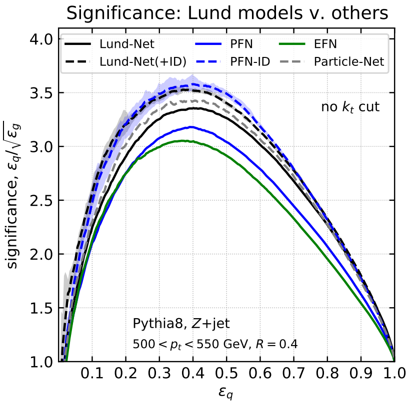
$pp \to Zq$ v. $pp \to Zg$ $(p_t \sim 500$ GeV, $R = 0.4)$




- Analytic approach shows gains for $k_t > 1$ GeV
  (shapes improve at small $\varepsilon_q$ by adding smaller $k_t$)

# Quark v. gluon jets: III. performance v. others

$pp \to Zq$ v. $pp \to Zg$ $\quad (p_t \sim 500$ GeV, $R = 0.4)$




- Analytic approach shows gains for $k_t > 1$ GeV
  (shapes improve at small $\varepsilon_q$ by adding smaller $k_t$)
- ML performance on par with PFN, slightly better than Particle-Net
  (treatment of PDG-ID could maybe be improved)