Measurement of the Lund Jet Plane Using Charged Particles in 13 TeV Proton-Proton Collisions with the ATLAS Detector

Bajic, Milena; ATLAS Collaboration

Published in:
Physical Review Letters

DOI:
10.1103/PhysRevLett.124.222002

Publication date:
2020

Document version
Publisher's PDF, also known as Version of record

Citation for published version (APA):
https://doi.org/10.1103/PhysRevLett.124.222002
Measurement of the Lund Jet Plane Using Charged Particles in 13 TeV Proton-Proton Collisions with the ATLAS Detector

G. Aad et al.*
(ATLAS Collaboration)

(Received 8 April 2020; revised manuscript received 6 May 2020; accepted 13 May 2020; published 4 June 2020)

The prevalence of hadronic jets at the LHC requires that a deep understanding of jet formation and structure is achieved in order to reach the highest levels of experimental and theoretical precision. There have been many measurements of jet substructure at the LHC and previous colliders, but the targeted observables mix physical effects from various origins. Based on a recent proposal to factorize physical effects, this Letter presents a double-differential cross-section measurement of the Lund jet plane using $139 \text{ fb}^{-1}$ of $\sqrt{s} = 13$ TeV proton-proton collision data collected with the ATLAS detector using jets with transverse momentum above 675 GeV. The measurement uses charged particles to achieve a fine angular resolution and is corrected for acceptance and detector effects. Several parton shower Monte Carlo models are compared with the data. No single model is found to be in agreement with the measured data across the entire plane.

DOI: 10.1103/PhysRevLett.124.222002

Jets are collimated sprays of particles resulting from high-energy quark and gluon production. The details of the process that underlies the fragmentation of quarks and gluons with quantum chromodynamic (QCD) charge into neutral hadrons is not fully understood. In the soft gluon (“eikonal”) picture of jet formation, a quark or gluon radiates a haze of relatively low energy and statistically independent gluons [1,2]. As QCD is nearly scale invariant, this emission pattern is approximately uniform in the two-dimensional space spanned by $\ln(1/z)$ and $\ln(1/\theta)$, where $z$ is the momentum fraction of the emitted gluon relative to the primary quark or gluon core and $\theta$ is the emission opening angle. This space is called the Lund plane [3]. The Lund plane probability density can be extended to higher orders in QCD and is the basis for many calculations of jet substructure observables [4–7].

The Lund plane is a powerful representation for providing insight into jet substructure; however, the plane is not observable because it is built from quarks and gluons. A recent proposal [8] describes a method to construct an observable analog of the Lund plane using jets, which captures the salient features of this representation. Jets are formed using clustering algorithms that sequentially combine pairs of protojets starting from the initial set of constituents [9]. Following the proposal, a jet’s constituents are reclustered using the Cambridge/Aachen (C/A) algorithm [10,11], which imposes an angle-ordered hierarchy on the clustering history. Then, the C/A history is followed in reverse (“declustered”), starting from the hardest protojet. The Lund plane can be approximated by using the softer (harder) protojet to represent the emission (core) in the original theoretical depiction. For each proto-jet pair, at each step in the C/A declustering sequence, an entry is made in the approximate Lund plane (henceforth, the “primary Lund jet plane” or LJP) using the observables $\ln(1/z)$ and $\ln(R/\Delta R)$, with

\[
z = \frac{p_T^{\text{emission}}}{p_T^{\text{emission}} + p_T^{\text{core}}} \quad \text{and} \quad \Delta R^2 = (y_{\text{emission}} - y_{\text{core}})^2 + (\phi_{\text{emission}} - \phi_{\text{core}})^2,
\]

where $p_T$ is transverse momentum [12], $y$ is rapidity, $R$ is the jet radius parameter, and $\Delta R$ measures the angular separation. Using this approach, individual jets are represented as a set of points within the LJP. Ensembles of jets may be studied by measuring the double-differential cross section in this space. The substructure of emissions, which may themselves be composite objects, is not considered in this analysis. To leading-logarithm (LL) accuracy, the average density of emissions within the LJP is uniform [8]:

\[
\frac{1}{N_{\text{jet}}} \frac{d^2 N_{\text{emissions}}}{d \ln(1/z) d \ln(R/\Delta R)} \propto \text{constant},
\]

where $N_{\text{jet}}$ is the number of jets. This construction of the plane is selected to separate momentum and angular
which have so far only been studied with the jet mass \( m_{\text{jet}} \) [19,20] (which is itself a diagonal line in the LJP: \( \ln 1/z \sim \ln m_{\text{jet}}^2/p_T^2 - 2 \ln R/\Delta R \) and groomed jet radius \[21,22\]. The number of emissions within regions of the LJP is also calculable and provides optimal discrimination between quark and gluon jets [5].

This Letter presents a double-differential cross-section measurement of the LJP, corrected for detector effects, using an integrated luminosity of 139 fb\(^{-1}\) of \( \sqrt{s} = 13 \) TeV proton-proton \((pp)\) collision data collected by the ATLAS detector. A unique feature of this measurement is that contributions from various QCD effects such as initial-state radiation, the underlying event and multiparton interactions, hadronization, and perturbative emissions are well localized in the LJP. This factorization is shown in Fig. 1(a), which qualitatively indicates the regions

FIG. 1. (a) Schematic representation of the LJP. The line \( z \theta \leq \Lambda_{\text{QCD}} \) roughly indicates the transition between regions where either perturbative \( (z \theta > \Lambda_{\text{QCD}}) \) or nonperturbative \( (z \theta < \Lambda_{\text{QCD}}) \) effects are expected to dominate. “UE/mpi” denotes the region where sources of nearly uniform radiation are relevant. (b) The ratio of the Lund jet plane as simulated by the HERWIG7.1.3 MC generator with either an angle-ordered parton shower or a dipole parton shower. (c) The ratio of the Lund jet plane as simulated by the SHERPA2.2.5 MC generator with either the AHADIC cluster-based or Lund string-based hadronization algorithm. (d) The ratio of the LJP as simulated by either the POWHEG+PYTHIA8.230 or PYTHIA8.230 MC generators. The inner set of axes indicate the coordinates of the LJP itself, while the outer set indicate corresponding values of \( z \) and \( \Delta R \).
populated by soft vs hard, wide-angle vs collinear, and perturbative vs nonperturbative radiation. Since different regions are dominated by factorized processes, the LJP measurement can be useful for tuning non-perturbative models and for constraining the model parameters of advanced parton shower (PS) Monte Carlo (MC) programs [23–26].

The ATLAS detector [27–29] is a general-purpose particle detector which provides nearly 4π coverage in solid angle. The inner tracking detector (ID) is inside a 2 T magnetic field and measures charged-particle trajectories up to |η| = 2.5. The innermost component of the ID is a pixeled silicon detector with fine granularity that is able to resolve ambiguities inside the dense hit environment of jet cores [30], surrounded by silicon strip and transition radiation detectors. Beyond the ID are electromagnetic and hadronic calorimeters, from which topologically connected clusters of cells [31] are formed into jets using the anti-kt algorithm with radius parameter R = 0.4 [32,33]. The jet energy scale is calibrated so that, on average, the detector-level jet energy is the same as that of the corresponding particle-level jets [34].

Events are selected using single-jet triggers [35,36]. The leading and subleading jets are used for the measurement and are required to satisfy \( p_T^{\text{leading}} > 675 \text{ GeV} \) and \( p_T^{\text{subleading}} < 1.5 \times p_T^{\text{leading}} \). This jet-\( p_T \) balance simplifies the interpretation of the final state in terms of a 2 → 2 scattering process. Both jets must be within the ID acceptance (|η| < 2.1). About 29.5 million jets satisfy these selection criteria.

Particle-level charged hadrons and their reconstructed tracks are used for this measurement because individual particle trajectories can be precisely identified with the ID. As the LJP observables are dimensionless and isospin is an approximate symmetry of the strong force, the difference between the LJP observables constructed using all interacting particles and charged particles is small [21]. Tracks are required to have \( p_T > 500 \text{ MeV} \) and be associated with the primary vertex with the largest sum of track \( p_T^2 \) in the event [37]. Tracks within \( \Delta R = 0.4 \) of the cores of selected jets are used to construct the LJP observables by clustering them using the C/A algorithm and populating the plane by iterative declustering. The fiducial region of the measurement spans 19 bins in \( \ln(1/\zeta) \) between \( \ln(1/0.5) \) and 8.4 \times \ln(1/0.5), and 13 bins in \( \ln(R/\Delta R) \) between 0.0 and 4.33. The maximum \( \Delta R \) is the jet radius and the minimum \( \Delta R \) is comparable to the pixel pitch. The maximum \( z \) is 0.5 and the minimum is 500 MeV/\( p_T^{\text{jet}} \).

Samples of dijet events were simulated in order to perform the unfolding and compare with the corrected data. The nominal sample was simulated using PYTHIA8.186 [38,39] with the NNPDF2.3 LO [40] set of parton distribution functions (PDF), a \( p_T \)-ordered PS, Lund string hadronization [41,42], and the A14 set of tuned parameters (tune) [43]. Additional samples were simulated by PYTHIA8.230 [44] with the NNPDF2.3 LO PDF set and the A14 tune, using either the PYTHIA LO matrix elements (MEs) or NLO MEs from POWHEG [45–48]; SHERPA2.1.1 [49] with the CT10LO PDF set, a \( p_T \)-ordered PS [50], an ME with up to three partons (merged with the CKKW prescription [51]) and the AHADIC (A Hadronization model in C++) cluster-based hadronization model [52,53]; SHERPA2.2.5 with the CT14NNLO PDF set [54] including 2 → 2 MEs and either the AHADIC hadronization model or the Lund string model; and HERWIG7.1.3 [26,55,56] with the MMHT2014NLO PDF set [57] and either the default angle-ordered (Ang. ord.) PS or a dipole PS and cluster hadronization [52]. Further details of these samples may be found in Ref. [58]. The PYTHIA8.186 and SHERPA2.1.1 events were passed through the ATLAS detector simulation [59] based on GEANT4 [60]. The effect of multiple \( p_T \) interactions in the same and neighboring bunch crossings (pileup) was modeled by overlaying the hard-scatter event with minimum-bias \( p_T \) collisions generated by PYTHIA8 with the A3 tune [61] and the NNPDF2.3 LO PDF set. The distribution of pileup vertices was reweighted to match data, which have an average of 33.7 simultaneous interactions per bunch crossing.

Figures 1(b)–1(d) illustrate the kinematic domains of various physical effects in the LJP using ratios at charged-particle level between pairs of MC simulations where one component of the simulation is varied. Varying the PS model in HERWIG7.1.3 [Fig. 1(b)] results in differences of up to 50% in the perturbative hard and wide-angle emissions entering the lower-left region of the LJP. Changing the hadronization model in SHERPA2.1.1 [Fig. 1(c)] causes variations up to 50% in a different region of the plane, populated by softer and more collinear emissions at the boundary between perturbative and nonperturbative regions. Varying the ME from LO (PYTHIA8.230) to NLO (POWHEG+PYTHIA8.230) [Fig. 1(d)] causes small changes of up to 10% in the region populated by the hardest and widest-angle emissions.

Selected data are unfolded to correct for detector bias, resolution, and acceptance effects by applying iterative Bayesian unfolding [62] with four iterations implemented in RooUnfold [63]. The MC generator used to unfold the data is PYTHIA8.186. The number of iterations was chosen to minimize the total uncertainty. The unfolding procedure corrects the LJP constructed from detector-level objects to charged-particle level, where jets and charged particles are defined similarly to those at detector level: jets are reconstructed using the same anti-\( k_T \) algorithm with detector-level stable (\( ct > 10 \text{ mm} \)) nonpileup particles, excluding muons and neutrinos, as inputs. The same kinematic requirements as for detector-level jets are imposed on these jets; charged particles with \( p_T > 500 \text{ MeV} \) within \( \Delta R = 0.4 \) of the cores of particle-level jets are used to populate the charged-particle-level LJP.
Emissions at detector level and charged-particle level are uniquely matched in $\eta$-$\phi$ to construct the response matrix. The matching procedure follows the order of the C/A declustering, starting from the widest-angle detector-level emission and iterating towards the jet core. The closest charged-particle-level match with angular separation $\Delta R < 0.1$ takes precedence. Unmatched emissions from tracks not due to a single charged particle (detector level) and from nonreconstructed charged particles (charged-particle level) are accounted for with purity and efficiency corrections. Corrections are applied before (purity) and after (efficiency) the regularized inversion of the response matrix. Both the purity and efficiency corrections are about 20% for wide-angle, hard emissions (lower-left quadrant of the LJP), increasing to 80% for the most collinear splittings and 50% in the lowest-$z$ bins. For matched emissions, the $\ln(1/z)$ and $\ln(R/\Delta R)$ bin migrations between particle and detector levels are largely independent. Furthermore, since the differential cross section varies slowly across the LJP, the purities and efficiencies are approximately the same across the entire LJP. The $\ln(R/\Delta R)$ migrations in a given $\ln(1/z)$ bin are less than 60% for the smallest opening angles and decrease to less than 40% for the widest angles. The $\ln(1/z)$ migrations decrease from about 50% for the softest to about 20% for the hardest emissions, with some degradation for the softest emissions at small opening angles. Migrations for both observables are nearly symmetric except for $\ln(R/\Delta R) > 3$, where harder-to-resolve small opening angles are measured with asymmetric resolution. In less than 10% of these cases, particle-level and detector-level emissions are mismatched and therefore measured with the wrong $\ln(1/z)$. While the $\ln(R/\Delta R)$ migrations are nearly the same when $\ln(1/z)$ migrates by one bin, the $\ln(1/z)$ migrations increase by about 30% when $\ln(R/\Delta R)$ migrates by one bin.

The unfolded distribution is normalized to the number of jets that pass the event selection, rendering the measurement insensitive to the total jet cross section. After normalization, the integral of the LJP is the average number of emissions within the fiducial region.

Experimental systematic uncertainties are evaluated by applying variations to each source, propagating them through the unfolding procedure, and taking the difference between the modified and nominal results. Theoretical uncertainties arise from jet fragmentation modeling. Different systematic uncertainties are treated as being independent. The size of various sources of uncertainty within selected regions of the LJP is displayed in Fig. 3.

Uncertainties in the jet energy are determined using a mixture of simulation-based and in situ techniques [34]. These uncertainties cause the migration of jets into or out of the fiducial acceptance, and are typically above 3% in total, reaching at most 7%. Uncertainties related to the reconstruction of isolated tracks and tracks within dense environments are considered by modifying the measured $p_T$ of individual tracks or removing them completely [30,64]. These uncertainties are small, contributing less than 0.5%. Other experimental uncertainties related to the modeling of pileup and the stability of the measurement across data-taking periods are less than 1% except for the most collinear splittings, where they reach 5%. A data-driven nonclosure uncertainty is determined by unfolding the detector-level distribution following a reweighting based on a comparison of the corresponding simulated detector-level distribution with the data [65]. This uncertainty is less than 1% except for the most collinear splittings, where it approaches 5%. An uncertainty for the matching procedure between emissions at detector and charged-particle levels is determined by repeating the unfolding and iterating through the C/A declustering sequence in reverse (from collinear to wide-angle emissions), taking the change in the result as an uncertainty. This uncertainty is less than 1% everywhere.

Theoretical uncertainties arise mainly from the accuracy of jet fragmentation modeling. Variations in jet fragmentation can impact the result through a combination of sources: efficiency or purity corrections, response matrix, and unfolding prior. These contributions are estimated by repeating the unfolding with SHERPA2.2.1. As the correlation between the uncertainty sources is unknown, an envelope of the 100% and 0% correlation hypotheses is taken as the total modeling uncertainty. This uncertainty ranges between 5% and 20% depending on the region (larger for soft-collinear splittings) and is the largest single source of uncertainty. Experimental uncertainties are found to be comparable to those arising from modeling in some regions of the LJP.
The total systematic uncertainty varies across the LJP; an uncertainty between 5% and 20% is achieved. The uncertainty is found to increase as $k_t = z \Delta R$ decreases: the bin with the smallest $k_t$ is also measured least precisely, and has a total uncertainty of about 20%.

The unfolded LJP is shown in Fig. 2. A triangular region with $k_t \gtrsim \Lambda_{\text{QCD}}$ is populated nearly uniformly by perturbative emissions, agreeing with the LL expectation [Eq. (1)]. A large number of emissions are found at the transition to the nonperturbative regime, as $\alpha_s$ is enhanced for small values of $k_t$. Emissions beyond the transition fall within the nonperturbative region of the LJP ($k_t \lesssim \Lambda_{\text{QCD}}$), and are suppressed. The average number of emissions in the fiducial region is measured to be $7.34 \pm 0.03(\text{syst}) \pm 0.11(\text{stat})$. The uncertainty is estimated by propagating uncertainties from the measurement in an uncorrelated and symmetrized

![Graph](image)

**FIG. 3.** Representative horizontal and vertical slices through the LJP. Unfolded data are compared with particle-level simulation from several MC generators. The uncertainty band includes all sources of systematic and statistical uncertainty. The inset triangle illustrates which slice of the plane is depicted: (a) $0.67 < \ln(R/\Delta R) < 1.00$, (b) $1.80 < \ln(1/z) < 2.08$, (c) $3.33 < \ln(R/\Delta R) < 3.67$, and (d) $5.13 < \ln(1/z) < 5.41$. 222002-5
manner. The corresponding average emissions for PYTHIA8.230 is 7.64 and 7.67 for POWHEG+PYTHIA8.230. The average value for SHERPA2.2.5 is 6.90 for AHADIC hadronization and 7.30 for Lund string hadronization. The average value for HERWIG7 is 7.41 for the dipole PS and 7.37 for the angle-ordered PS. While a similar bracketing of the data by PYTHIA and SHERPA with AHADIC hadronization was noted in Ref. [66], the particle multiplicity inside jets has not previously been decomposed into perturbative and non-perturbative components.

Figure 3 shows data from four selected horizontal and vertical slices through the LJP, along with a breakdown of the systematic uncertainties [67]. The data are compared with predictions from several MC generators. While no prediction describes the data accurately in all regions, the HERWIG7.1.3 angle-ordered prediction provides the best description across most of the plane. The differences between the PS algorithms implemented in HERWIG7.1.3 are notable at large values of $k_t$, where the two models disagree most significantly for hard emissions reconstructed at the widest angles [Fig. 3(a) and 3(b)]. The POWHEG+PYTHIA and PYTHIA predictions only differ significantly for hard and wide-angle perturbative emissions, where ME corrections are relevant. The hadronization algorithms implemented in SHERPA2.2.5 are most different at small values of $k_t$, particularly for soft-collinear splittings at the transition between perturbative and non-perturbative regions of the plane. The ability of the LJP to isolate physical effects is highlighted in Fig. 3(b), where as emissions change from wide angled to more collinear, the distribution passes through a region sensitive to the choice of PS model, and then enters a region which is instead sensitive to the hadronization model. Figures 3(c) and 3(d) show regions dominated by nonperturbative effects. The PYTHIA samples describe the data in the collinear region of the jet core well, but all simulations fail to describe the softest, widest-angle emissions, which are characteristic of contributions from the underlying event. The PYTHIA8.186 and SHERPA2.2.1 predictions are not shown, but are consistent with the PYTHIA8.230 and SHERPA2.2.5 (Lund string hadronization) predictions, respectively. These observations indicate that the LJP may provide useful input to both perturbative and nonperturbative model development and tuning.

In summary, a measurement of the jet substructure based on the Lund jet plane is reported. The analysis dataset corresponds to an integrated luminosity of 139 fb$^{-1}$ of 13 TeV LHC proton-proton collisions recorded by the ATLAS detector. The measurement is performed on an inclusive selection of dijet events, with a leading jet $p_T > 675$ GeV. Selected jets are reconstructed from topological clusters using the anti-$k_t$ algorithm with $R = 0.4$, and their associated charged-particle tracks are used to construct the observables of interest. The data are presented as an unfolded double-differential cross section, and compared with several Monte Carlo generators with various degrees of modeling accuracy. This measurement illustrates the ability of the Lund jet plane to isolate various physical effects, and will provide useful input to both perturbative and nonperturbative model development and tuning.

We thank CERN for the very successful operation of the LHC, as well as the support staff from our institutions without whom ATLAS could not be operated efficiently. We acknowledge the support of ANPCyT, Argentina; YerPhI, Armenia; ARC, Australia; BMWF and FWF, Austria; ANAS, Azerbaijan; STSC, Belarus; CNPq and FAPESP, Brazil; NSERC, NRC and CFI, Canada; CERN; CONICYT, Chile; CAS, MOST and NSFC, China; COLCIENCIAS, Colombia; MSMT CR, MPO CR and VSC CR, Czech Republic; DNRF and DNSRC, Denmark; IN2P3-CNRS and CEA-DRF/IRFU, France; SRNSFG, Georgia; BMBF, HGF and MPG, Germany; GRS; Greece; RGC and Hong Kong SAR, China; ISF and Benoziyo Center, Israel; INFN, Italy; MEXT and JSPS, Japan; CNRST, Morocco; NWO, Netherlands; RCN, Norway; MNiSW and NCN, Poland; FCT, Portugal; MNE/IFA, Romania; MES of Russia and NRC KI, Russia Federation; JINR; MESTD, Serbia; MSSR, Slovakia; ARRS and MIZŠ, Slovenia; DST/NRF, South Africa; MINECO, Spain; SRC and Wallenberg Foundation, Sweden; SERI, SNSF and Cantons of Bern and Geneva, Switzerland; MOST, Taiwan; TAEK, Turkey; STFC, United Kingdom; DOE and NSF, USA. In addition, individual groups and members have received support from BCKDF, CANARIE, Compute Canada and CRC, Canada; ERC, ERDF, Horizon 2020, Marie Skłodowska-Curie Actions and COST, European Union; Investissements d’Avenir Labex, Investissements d’Avenir Idex and ANR, France; DFG and AvH Foundation, Germany; Herakleitos, Thales and Aristeia programmes co-financed by EU-ESF and the Greek NSRF, Greece; BSF-NSF and GIF, Israel; CERCA Programme Generalitat de Catalunya and PROMETEO Programme Generalitat Valenciana, Spain; Göran Gustafssons Stiftelse, Sweden; The Royal Society and Leverhulme Trust, United Kingdom. The crucial computing support from all WLCG partners is acknowledged gratefully, in particular from CERN, the ATLAS Tier-1 facilities at TRIUMF (Canada), NDGF (Denmark, Norway, Sweden), CC-IN2P3 (France), KIT/GridKA (Germany), INFN-CNAF (Italy), NL-T1 (Netherlands), PIC (Spain), ASGC (Taiwan), RAL (UK) and BNL (USA), the Tier-2 facilities worldwide and large non-WLCG resource providers. Major contributors of computing resources are listed in Ref. [68].
