Calculating Track-Based Observables for the LHC

Hsi-Ming Chang,1 Massimiliano Procura,2 Jesse Thaler,3 and Wouter J. Waalewijn1

1Department of Physics, University of California at San Diego, La Jolla, California 92093, USA
2Albert Einstein Center for Fundamental Physics, Institute for Theoretical Physics, University of Bern, CH-3012 Bern, Switzerland
3Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

(Received 3 April 2013; published 3 September 2013)

By using observables that only depend on charged particles (tracks), one can efficiently suppress pileup contamination at the LHC. Such measurements are not infrared safe in perturbation theory, so any calculation of track-based observables must account for hadronization effects. We develop a formalism to perform these calculations in QCD, by matching partonic cross sections onto new nonperturbative objects called track functions which absorb infrared divergences. The track function $T_i(x)$ describes the energy fraction $x$ of a hard parton $i$ which is converted into charged hadrons. We give a field-theoretic definition of the track function and derive its renormalization group evolution, which is in excellent agreement with the PYTHIA parton shower. We then perform a next-to-leading order calculation of the total energy fraction of charged particles in $e^+ e^- \rightarrow$ hadrons. To demonstrate the implications of our framework for the LHC, we match the PYTHIA parton shower onto a set of track functions to describe the track mass distribution in Higgs plus one jet events. We also show how to reduce smearing due to hadronization fluctuations by measuring dimensionless track-based ratios.

DOI: 10.1103/PhysRevLett.111.102002 PACS numbers: 12.39.St, 13.87.Ce, 13.87.Fh

Jets are collimated sprays of particles that arise from the fragmentation of energetic quarks and gluons. Nearly every measurement at the Large Hadron Collider (LHC) involves jets in some way, either directly as probes of physics in and beyond the standard model, or indirectly as a source of backgrounds and systematic uncertainties. In order to predict jet-based observables using quantum chromodynamics (QCD), one typically performs infrared- and collinear-safe (IRC safe) jet measurements which involve only the kinematics of the jet constituents [1]. In particular, IRC safe jet measurements do not distinguish between charged and neutral particles, despite the fact that, for example, charged pions ($\pi^-$) are measured using both tracking and calorimetry whereas neutral pions ($\pi^0$) are measured using calorimetry alone.

In this Letter, we develop the theoretical formalism to calculate track-based observables, which depend on the kinematics of charged particles alone but not on their individual properties or multiplicities. The experimental motivation for track-based measurements is that tracking detectors offer better pointing and angular resolution than calorimetry. By only using tracks, one can substantially mitigate the effects of pileup (multiple collision events in a single bunch crossing) which is becoming more relevant as the LHC achieves higher luminosity (see, e.g., [2–6] for alternative approaches). In addition, tracks can aid in jet substructure studies where the angular energy distribution in the jet discriminates between different jet types [7,8]. While we focus on charged particles, this formalism applies to any (otherwise) IRC safe measurement performed only on a subset of particles.

To describe track-based observables in QCD, we introduce the track function $T_i(x, \mu)$. A parton (quark or gluon) labeled by $i$ with four-momentum $p_i^{\mu}$ hadronizes into charged particles with total four-momentum $\bar{p}_i^\mu \equiv x p_i^\mu + \mathcal{O}(\Lambda_{\text{QCD}})$. The distribution in the energy fraction $0 \leq x \leq 1$ is the track function and is by definition normalized

$$\int_0^1 dx T_i(x, \mu) = 1. \quad (1)$$

The track function is similar to a fragmentation function (FF) or a parton distribution function (PDF) in that it is a fundamentally nonperturbative object that absorbs infrared (IR) divergences in partonic calculations. Like FFs and PDFs, the track function has a well-defined dependence on the renormalization group (RG) scale $\mu$ through a DGLAP-type evolution [9–13], though the specific evolution is more reminiscent of the jet charge distribution [14,15]. For the observables we consider, each parton has its own independent track function. Hadronization correlations are captured by power corrections (beyond the scope of this Letter).

Consider the cross section for an IRC safe observable $e$ measured using partons

$$\frac{d\sigma}{de} = \sum_N d\Pi_N \frac{d\sigma_N}{d\Pi_N} \delta[e - \hat{e}(\bar{p}_i^\mu)]. \quad (2)$$

where we drop possible convolutions with PDFs to keep the notation simple. Here, $\Pi_N$ denotes $N$-body phase space, $d\sigma_N/d\Pi_N$ is the corresponding partonic cross section, and $\hat{e}(\bar{p}_i^\mu)$ implements the measurement on the partonic four-momenta $p_i^\mu$. Since $e$ is an IRC safe observable, the KLN theorem [16,17] guarantees a cancellation of final state IR divergences between real and virtual diagrams. The cross section for the same observable measured using only tracks is
\[
\frac{d\sigma}{d\vec{e}} = \sum_N \int d\Pi_N \frac{d\sigma_N}{d\Pi_N} \int d\bar{\sigma} \{x_i T_i(x_i) \delta[\vec{e} - \hat{e}(x_i p_i^\mu)]\},
\]

(3)

where \(T_i(x_i)\) is the track function for parton \(i\). This equation defines a matching onto track functions where \(d\bar{\sigma}/d\Pi_N\) represents the short distance matching coefficient, which is calculable in perturbation theory. In the absence of track functions, \(d\bar{\sigma}/d\Pi_N\) would exhibit a mismatch between real and virtual diagrams in the form of uncombined IR divergences in the partonic computation. The track functions absorb these IR divergences, and the partonic cross section \(\bar{\sigma}_N\) is correspondingly modified with respect to \(\sigma_N\). We will show below for the example of \(e^+e^- \rightarrow q\bar{q}g\) how the mismatch in the absence of \(T_i(x_i)\) occurs. Figure 1 shows schematically how we determine the IR-finite matching coefficient \(\bar{\sigma}_3\) for this case, by using that Eq. (3) is valid both at the hadronic and partonic level. The fact that we consider factorizable (otherwise) IRC-safe observables modified to include only charged particles and that collinear divergences are known to be universal in QCD [18–20] guarantees a valid matching to all orders in the strong coupling constant \(\alpha_s\).

At the leading order (LO) in \(\alpha_s\), the cross section depends on a single partonic multiplicity \(N\) and there are no IR divergences implying \(\bar{\sigma}_N = \sigma_N\). The LO \(T_i^{(0)}(x_i)\) is simply a finite distribution which can be obtained directly from the energy fraction of charged particles in a jet initiated by a parton \(i\). Ideally, we would extract this information from data, but just for illustrative purposes, we can determine it from (tuned) Monte Carlo event generators. We stress that our formalism does not rely on the use of these programs nor on their built-in hadronization models. In Fig. 2, we show the track functions obtained from pure quark and gluon jet samples produced by PYTHIA 8.150 [21,22] and clustered using the anti-\(k_T\) algorithm [23] in FASTJET 2.4.4 [24]. (To extract the track function at next-to-leading order (NLO) we use Eq. (11): the jet radius \(R\) is correlated with the RG scale \(\mu\).) As expected, the up- and down-quark track functions are very similar, with a peak at \(x = 0.6\). This means that on average 60% of the energy of the initial quark is contained in charged hadrons, in agreement with a recent CMS study [25]. The small difference between up and down is due to strangeness, since \(u\) and \(d\) mesons are charged whereas \(s\) mesons are neutral. Because gluons have a larger color factor than quarks, they yield a higher track multiplicity, and the corresponding track functions are narrower, as expected from the central limit theorem.

Formally, the (bare) track function is defined in QCD in a fashion analogous to the unpolarized FF (cf. [26,27]). Expressed in terms of light-cone components,

\[
T_q(x) = \int dy^+ d^2y \gamma^+ y^+/2 \frac{1}{2N_i \sum \delta(x - \frac{p_C}{k^+})} \times \text{tr} \left[ \frac{\gamma^+}{2} (0|\psi(0,x^+,0,y^+)CN)(CN|\psi(0)|0) \right],
\]

(4)

where \(\psi\) is the quark field, \(C (N)\) denote charged (neutral) hadrons, and \(p_C\) is the large momentum component of all charged particles. Whereas the FF describes the energy fraction carried by an individual hadron, the track function describes the energy fraction carried by all charged particles. As for the FF, gauge invariance requires the addition of eikonal Wilson lines. The gluon track function is defined analogously [28].

Treating the intermediate states in Eq. (4) partonically, we obtain the bare track functions \(T_{i,bare}^{(1)}\) at NLO in pure dimensional regularization with \(d = 4 - 2\epsilon\),

\[
T_{i,bare}^{(1)}(x) = \frac{1}{2} \sum_{j,k} \int dx \left[ \frac{\alpha_s(\mu)}{2\pi} \left( \frac{1}{\epsilon_{UV}} - \frac{1}{\epsilon_{IR}} \right) P_{i-jk}(z) \right] \times \delta(x - z x_i - (1 - z)x_j),
\]

(5)

which arise from collinear splittings, controlled by the timelike Altarelli-Parisi splitting functions \(P_{i-jk}(x)\) [12].
In contrast with the analogous partonic FF calculation, track functions involve contributions from both branches of the splitting. Renormalizing the ultraviolet divergences in Eq. (5) in \( \overline{\text{MS}} \) leads to the evolution equation for the track function

\[
\mu \frac{d}{d\mu} T_f(x, \mu) = \frac{1}{2} \sum_{k} \int dz dx_1 dx_2 \frac{\alpha_s(\mu)}{\pi} P_{\gamma \rightarrow k}(z) T_f(x_1, \mu)
\]

\[
\times T_k(x_2, \mu) \delta[x - z x_1 - (1 - z) x_2].
\]  

(6)

Like for a PDF, the track function can be extracted at one scale and RG evolved to another scale, and the evolution preserves the normalization in Eq. (1). Unlike a PDF, Eq. (6) involves a convolution of two track functions at NLO (and more convolutions at higher orders corresponding to multiple branchings), so it is numerically more involved to perform the \( \mu \) evolution. At leading logarithmic (LL) order, the RG evolution in Eq. (6) is equivalent to a parton shower, and Fig. 3 demonstrates excellent agreement between our numerical evolution and the parton shower in PYTHIA.

For a calculation at NLO, both the partonic cross section and the track functions have IR divergences which cancel in Eq. (3). To demonstrate this in a simple example, consider the process \( e^+ e^- \rightarrow \text{hadrons} \) at a center-of-mass energy \( Q \) where one measures the total energy fraction \( w \) of charged particles. At NLO the partonic process is \( e^+ e^- \rightarrow q\bar{q}g \), whose kinematics are described by the energy fractions \( y_1 \) and \( y_2 \) of the quark and antiquark. Applying Eq. (3) we find

\[
\frac{d\sigma}{dw} = \int dy_1 dy_2 \frac{d\tilde{\sigma}}{dy_1 dy_2} \int dx_1 dx_2 dx_3 T_q(x_1) T_{\bar{q}}(x_2) T_g(x_3)
\]

\[
\times \delta(w - [y_1 x_1 + y_2 x_2 + (2 - y_1 - y_2) x_3])/2),
\]  

(7)

since \( T_q = T_{\bar{q}} \). The matching coefficient \( d\tilde{\sigma} \) is extracted by evaluating this equation at the partonic level (see Figure 1).

We then use it in Eq. (7) together with nonperturbative hadronic track functions to obtain the physical cross section \( d\sigma/dw \). At the LO partonic level

\[
\frac{d\sigma^{(0)}}{dy_1 dy_2} = \sigma^{(0)} \delta(1 - y_1)\delta(1 - y_2),
\]  

(8)

where \( \sigma^{(0)} \) is the total Born cross section. At NLO, the cross section can be expressed using plus functions as

\[
\frac{d\sigma^{(1)}}{dy_1 dy_2} = \sigma^{(0)} \frac{\alpha_s(\mu) C_F}{2 \pi} \left( \frac{\pi^2}{2} - 4 \right) \delta(1 - y_1) \delta(1 - y_2)
\]

\[
+ \frac{\theta(y_1 + y_2 - 1)(y_1^2 + y_2^2)}{2(1 - y_1)(1 - y_2)} + \delta(1 - y_1)
\]

\[
\times \left[ - \frac{1}{\epsilon_{IR}} \frac{P_{g\rightarrow q\bar{q}}(y_2)}{C_F} + (1 + y_2^2) \left[ \frac{\ln(1 - y_2)}{1 - y_2} \right]_+ 
\]

\[
+ \frac{P_{g\rightarrow q\bar{q}}(y_2)}{C_F} \ln y_2 Q^2 \frac{1}{\mu^2} + 1 - y_2 \right] + (y_1 \leftrightarrow y_2),
\]  

(9)

where \( C_F = 4/3 \) and both real and virtual contributions are included. The \( 1/\epsilon_{IR} \) divergences in \( \sigma^{(1)} \) are canceled by the ones in \( T^{(1)} \) from Eq. (5), and the finite remainder defines \( \tilde{\sigma}^{(1)} \) in \( \overline{\text{MS}} \). (For different IR regulators, the track function may also contribute finite terms to the matching.) By performing this kind of matching calculation, one can determine \( \tilde{\sigma}_N \) for any process to any order in \( \alpha_s \). For the case of Eq. (9) which involves a single scale \( Q \), we choose \( \mu \approx Q \) to minimize the logarithms in \( \tilde{\sigma} \). In Fig. 4 we show the LO and NLO distributions for the energy fraction \( w \), using the track functions extracted from PYTHIA by means of Eq. (11). There is good convergence from LO to NLO and our fixed-order calculation agrees well with PYTHIA.

Ultimately, we are interested in applying the track function formalism to jet-based measurements at the LHC, which typically involve multiple scales. For narrow well-separated jets, though, the contributions from soft radiation are power suppressed, so the energy fraction \( x \) of the charged particles within a single jet depends on the jet scale

\[
\mu_f \approx p_T R,
\]  

(10)

where \( p_T \) is the transverse momentum of the jet and \( R \) denotes its azimuthal-rapidity size as defined by a jet algorithm (anti-\( k_T \) in this Letter). Indeed, by varying \( R \) (and trying different jet algorithms) while keeping
\[ \mu_J = p_T R \text{ fixed, we find nearly identical } x \text{ distributions in } \text{PYTHIA. At LO, the normalized distribution in } x \text{ defines } T_i^{(0)}(x) \text{ itself, shown in Fig. 2. At NLO, the distribution of the energy fraction } x \text{ within a jet initiated by parton } i \text{ has the same form as for the jet charge distribution [15]} \]

\[ \frac{1}{\sigma_i} \frac{d\sigma}{dx} = \frac{1}{2} \sum_{j,k} \int dx_1 dx_2 dz \frac{J_{ij}(p_T R, z, \mu_J)}{(2\pi)^3 J_i(p_T R, \mu_J)} T_j(x_1, \mu_J) \times T_k(x_2, \mu_J) \delta(x - x_1 - (1 - x_2)). \quad (11) \]

The jet functions \( J_i(p_T R, \mu) \) [29] arise from the \( 1/\sigma_i \) normalization factor and describe jet production without any additional measurement. The matching coefficients are \( J_{ij} \) [15,30]. At NLO there is at most a \( 1 \rightarrow 2 \) splitting which completely fixes the kinematics, so the same \( J_{ij} \) appear in the jet charge [15] and in fragmentation inside an identified jet [31,32]. The coefficients will not be the same at higher orders. We can invert Eq. (11) to determine the NLO track functions in Fig. 2.

As a track-based measurement relevant for the LHC, consider the track-only jet mass spectrum in \( pp \) to Higgs plus one jet. This example is more complicated than the \( e^+ e^- \) example above since it involves several scales: the hard scale set by the transverse momentum of the jet \( p_T^f \), the jet mass scale \( m_J \), and the scale associated with soft radiation \( m_J^2/p_T^f \). In a resummed jet mass calculation, one would need to treat each emission in the exponentiated soft function as hadronizing independently, i.e., convolved with its own track function [28]. At LL order, however, we can use the fact that the parton shower already describes the numerous parton emissions that build up the jet mass distribution. Thus, a correct and instructive use of our formalism at LL is to run the perturbative PYTHIA parton shower down to a low scale \( \mu \approx \Lambda_{QCD} \) and match each final state parton onto a track function (also evolved to \( \mu \approx \Lambda_{QCD} \)).

In Fig. 5, we show the track mass \( m_f \) spectrum using the \( E \) scheme [33] for the anti-\( k_T \) jet algorithm with \( R = 1.0 \). We impose realistic cuts on the jet rapidity \( \eta_J \) and \( p_T^f \) at the calorimeter level (using all the particles), to select the same events for both curves. There is excellent agreement between the track function and the Lund hadronization model [34] in PYTHIA, which is a nontrivial check considering that the track functions were originally extracted from PYTHIA using jet energies (and not jet mass) at \( \mu = 100 \text{ GeV} \). Since the track function does not include correlations between the hadronization of different partons, the agreement in Fig. 5 shows that independent fragmentation is an excellent approximation for these observables. The small difference between the two distributions in the peak region is due to nonperturbative power corrections, which to first approximation can be described by a shift in \( m_f^2 \) (see, e.g., [33,35–37]).

Despite the advantages of using track-based observables for pileup suppression, it should be noted that the track functions have a rather large width (see Fig. 2). Fluctuations in the charged energy fraction produce an effective energy smearing for a track-based measurement relative to a calorimetric one. To partially address this issue, one can focus on observables for which the track function would only have a modest effect, such as for dimensionless ratios of observables. A particularly useful example are \( N \)-subjettiness ratios [38], which are relevant for jet substructure studies. Though a calculation of such ratios is beyond the scope of this Letter, we plot in Fig. 6 the ratio of the jet mass to jet \( p_T \) in PYTHIA, measured using either tracks alone or all particles (calorimeter). As expected, the smearing effect is reduced since hadronization fluctuations are correlated between the numerator and denominator.

Given the potential experimental benefits from track-only measurements at high luminosities, we expect that track functions will offer an important theoretical handle for future precision jet studies at the LHC. Beyond the tests of our formalism against PYTHIA performed here, we stress that it is possible to systematically improve the accuracy of track-based predictions. At fixed order in \( \alpha_s \), one can calculate higher-order corrections for the
FIG. 6 (color online). Comparison between track and calorimeter measurements of the ratio of jet mass and jet $p_T$. Here the cuts are on $p_T^J$ and $\eta_J$ from the tracks in the jet.

(process-dependent) matching onto track functions. More ambitiously, one could match track functions onto automated NLO calculations, perhaps using a more convenient IR regulator (such as dipole subtraction [39]) rather than the $\overline{MS}$ scheme used here. To improve the accuracy of resummation, one needs to determine the higher-order RG evolution of the track functions, and precision track-based studies would also require power corrections. Ultimately, one would want to follow the example of PDFs and extract track functions for the LHC from global fits to data.

We thank A. Manohar and I. Stewart for discussions and feedback on this Letter. We also thank M. Pierini for discussions. H.-M. C. and W. J. W. are supported by DOE Grant No. DE-FG02-05ER-41360 and No. DE-FG02-11ER-41741.