

Calculating pull for non-singlet jets

Yunjia Bao and Andrew J. Larkoski

*Physics Department, Reed College,
Portland, OR 97202, U.S.A.*

E-mail: baoy@reed.edu, larkoski@reed.edu

ABSTRACT: The pull vector is a jet observable sensitive to the distribution of soft radiation controlled by the color flow in a collider event. We present calculations to leading order in the soft and collinear limits for the pull vector measured between pairs of jets that do not form a color-singlet dipole. Our calculations are presented within the context of $e^+e^- \rightarrow$ three jets events, on which pull is measured between the two subleading jets. A subset of these calculations can be re-interpreted as a bottom-anti-bottom quark jet pair in a color octet configuration, which can be a background to Higgs production at large boost. We also present a universal expression for the pull distribution in the high-boost and small jet radius limit. This distribution is controlled by color SU(3) quadratic Casimirs that arise from product representations of pairs of QCD jets.

KEYWORDS: Jets, QCD Phenomenology

ARXIV EPRINT: [1910.02085](https://arxiv.org/abs/1910.02085)

Contents

1	Introduction	1
2	Observable definitions	3
3	Pull in $e^+e^- \rightarrow 3$ jet events	4
3.1	Inclusive prediction for pull	7
4	$H \rightarrow b\bar{b}$ vs. $g \rightarrow b\bar{b}$	9
4.1	Discrimination power of pull	10
5	Other color representations	14
6	Conclusions	16
A	Non-connected soft function calculation	17
B	Gluon jet function	18

1 Introduction

In the search for new physics at the Large Hadron Collider (LHC), a central goal is to measure all quantum numbers and couplings of known Standard Model particles as well as to observe as-of-yet undiscovered particles. At a collider experiment, measuring the mass of a particle is straightforward because detectors measure nearly all of the energy produced in collision. Techniques exist for determining a particle’s electric charge by weighting hits in the tracking system by their energy [1, 2]. However, direct measurement of a particle’s charge under quantum chromodynamics (QCD) or color is subtle, as all of the particles that are actually detected by experiment are color-neutral. A particle’s color, or at least if it has non-zero color, is inferred from jet production and quantities measured on jets [3–7].

The observable pull [8] was introduced to be directly sensitive to the flow of color between pairs of jets. The distribution of soft radiation throughout a collision event is determined by the location and connections of color dipoles, the ends of which are the observed jets. Pull quantifies the location of the dominant soft radiation between two jets, thus providing a measure of their color connectedness or their color “pull” on one another. Two jets that form a color-singlet dipole, from the decay of a color-singlet resonance, for example, will dominantly emit soft radiation in the region between the jets, because gluons emitted at wide angles would only see the net zero color of the two jets. By contrast, two jets that are produced from standard QCD processes at the LHC would in general have color connections to the initial state, as well as to whatever other color objects were produced. Thus, radiation about these jets would have a much weaker correlation with the relative locations of the jets of interest.

While pull has a very nice physical interpretation, has been studied extensively in simulation, and has even been measured in experiment [9–11] and used in searches [12–16], there has been little theoretical analysis of the observable to honestly justify that it does what it is claimed to do. The first calculations of pull as measured on the two jets from color-singlet decay were presented in ref. [17], which illustrated the challenges of the calculation and demonstrated that at least in this restricted case, pull acts as expected. Further, pull had been measured in experiment on jets from W boson decay, which enabled direct comparison to data.

Nevertheless, to demonstrate that pull is indeed sensitive to the color connection between two jets, we should demonstrate that there is a significant difference between the calculated distribution of pull for pairs of jets that do and do not form a color-singlet dipole. This is our goal in this paper. The simplest collision event in which pairs of jets do not form color-singlets is $e^+e^- \rightarrow$ three jet events, which is what we consider first. For simplicity, we restrict our analysis to leading-order in the strong coupling and to leading power in the soft or collinear limits. We will measure the color-connectedness of the two jets closest in angle with pull. With three jets in the final state, there are therefore three distinct dipoles from which soft radiation can be emitted. We will show that the radiation from two of these dipoles can be accounted for from calculations presented in ref. [17], with some re-interpretation. We present a new calculation for the distribution of radiation from the dipole which does not include the jet on which pull is directly measured.

Our calculations for pull in three-jet final states can then be leveraged to theoretically understand one of the original motivations for the observable. Identifying the decay of the Higgs boson to bottom quarks at high significance is a challenge at the LHC. Once jets have been tagged as containing bottom quarks, a major background to $H \rightarrow b\bar{b}$ decay is the gluon splitting process $g \rightarrow b\bar{b}$. Because this splitting lacks a soft singularity, once the invariant mass of the bottom quarks is selected for, the kinematics of the bottom quarks from Higgs decay and gluon splitting are nearly identical. However, because the Higgs boson is a color singlet, radiation from the bottom quarks is confined to lie between them, distinct from the gluon splitting case. This suggests that pull may provide discrimination power between these two processes, though this has not been observed in simulation [18–20].

With an explicit calculation, we are able to understand the discrimination properties of pull in a controlled, well-defined context. Working in the limits in which both the Higgs is highly boosted and the radii of the bottom quark subjects is small, we are able to explicitly calculate the discrimination power of pull for this problem, quantified in the signal vs. background efficiency curve. We conjecture that in these boosted and collinear limits the pull distribution exhibits a universality, exclusively depending on the color configuration of the two nearby jets. In these limits, we are able to write down a master formula for the distribution of pull, for any two jets in QCD on which it might be measured. We enumerate all possible irreps of color $SU(3)$ that arise in the product representation of the color of two QCD jets and how that affects the corresponding pull distribution. This explicitly demonstrates that pull is indeed sensitive to the color flow between a pair of jets in a simplified limit. We leave validation of this observation in simulation to future work.

This paper is organized as follows. In section 2, we first define the pull observable, presenting a slightly modified definition from that originally proposed that is more natural

in the e^+e^- collision case. In section 3, we present the calculation of the distribution of the pull observable measured on the closest two jets in angle produced in $e^+e^- \rightarrow$ three jets events. Section 4 expands on these results, and applies them to the problem of discrimination of $H \rightarrow b\bar{b}$ and $g \rightarrow b\bar{b}$ in the highly boosted limit. We explore pull for all possible representations of SU(3) color that can arise in the product of the color of two QCD jets in section 5, and conclude and discuss future directions in section 6. appendices contain details of calculations quoted in the body of the paper.

2 Observable definitions

For application to jets produced in hadron collisions, ref. [8] introduced the pull vector \vec{t} as:

$$\vec{t}_{\text{original}} = \sum_{i \in J} \frac{p_{\perp i} |\vec{r}_i|}{p_{\perp J}} \vec{r}_i. \tag{2.1}$$

Here, the sum runs over the particles i in a jet J of interest, p_{\perp} is the momentum transverse to the collision beam, and the vector \vec{r}_i is

$$\vec{r}_i = (y_i - y_J, \phi_i - \phi_J). \tag{2.2}$$

The jet center is located at rapidity-azimuth of (y_J, ϕ_J) and particle i is located at (y_i, ϕ_i) . The jet center is just defined as the vector sum of the momenta of all particles that compose the jet. Pull is therefore defined as a two-dimensional vector in the plane of the cylindrical detector. As used to probe color connections, pull can be measured on two jets and the directions of their vectors compared. Jets with a strong color connection (i.e., jets that form a color-singlet dipole), will have pull vectors that point toward one another, while weakly color-connected jets will have pull vectors with a random relative orientation.

For the calculations presented in this paper, however, we use a slightly different definition of the pull vector motivated both by our study of jets in e^+e^- collisions as well as simplifying analytical calculations. This modified definition of the pull vector was introduced in ref. [17] and is

$$\vec{t}_{\text{modified}} = \sum_{i \in J} \frac{E_i \sin^2 \theta_i}{E_J} (\cos \phi_i, \sin \phi_i). \tag{2.3}$$

Now, E is the energy, θ_i is the angle from particle i to the jet center, and ϕ_i is the azimuthal angle of the particle about the jet center. In the collinear limit, these two definitions are identical, but in general differ at finite angle. As our investigation mainly concerns with the limit where the jet radius $R \ll 1$, this modified definition of the pull vector simplifies the calculation without losing the limit behavior of interest. For all results presented in this paper, the azimuthal angle ϕ_i will be defined with respect to the location of a neighboring jet's center. Specifically, if a particle i lies on the line between the jet J and the reference jet, $\phi_i = 0$, while if it is on the other side of jet J , $\phi_i = \pi$. This is what we will mean by measuring pull on a pair of jets: we calculate the pull vector of one jet whose components are defined with respect to the location of the second jet.

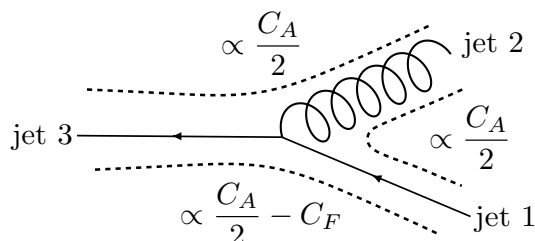


Figure 1. Illustration of color flow between the three jets produced in $e^+e^- \rightarrow q\bar{q}g$ collisions. The dashed lines represent the flow of color in dipoles that stretch between pairs of jets with the corresponding absolute value of the color charge carried by each dipole indicated.

Rather than the Cartesian components of the pull vector, we will typically express it as its magnitude t and azimuthal angle ϕ_p . We call ϕ_p the pull angle and it is defined as

$$\phi_p = \cos^{-1} \frac{t_x}{t} = \cos^{-1} \frac{\sum_{i \in J} E_i \sin^2 \theta_i \cos \phi_i}{\left| \sum_{i \in J} E_i \sin^2 \theta_i (\cos \phi_i, \sin \phi_i) \right|}. \quad (2.4)$$

The pull vector is infrared and collinear (IRC) safe, and so its distribution can be calculated order-by-order in perturbation theory. However, the pull angle ϕ_p alone is not IRC safe. We will have to deal with this later as the pull angle is the aspect of the pull vector that is most sensitive to color connections between jets.

3 Pull in $e^+e^- \rightarrow 3$ jet events

To illustrate the form of the pull distribution for pairs of jets that do not form a color-singlet dipole, we will study pull as measured on pairs of jets in $e^+e^- \rightarrow$ three jets events. The setup of how we measure pull on this final state is illustrated in figure 1. For a three-jet final state in the center-of-mass frame, those three jets lie in a plane and the most energetic will be isolated in a hemisphere about the collision point. We refer to this most energetic jet as jet 3. The two lower-energy jets are the pair closest in angle. Of these two jets, jet 1 is the most energetic and the jet on which we measure pull. Jet 2, the lowest-energy jet, defines the axis along which the pull angle is defined to be 0. Figure 1 shows that the quark is the most energetic jet, and the anti-quark is the second most-energetic jet. Thus, in this configuration, we measure pull on the anti-quark jet with respect to the gluon jet's direction.

Figure 1 also shows the strength of color correlation between the pairs of jets, as measured by the product of each particle's color matrix. Because the total color of the final state is 0, the sum of the color of the quark, anti-quark, and gluon is 0:

$$\mathbf{T}_q + \mathbf{T}_{\bar{q}} + \mathbf{T}_g = 0. \quad (3.1)$$

The square of any of the color matrices is just the quadratic Casimir for that particular color representation. In QCD, we have

$$\mathbf{T}_q^2 = \mathbf{T}_{\bar{q}}^2 = C_F = \frac{4}{3}, \quad \mathbf{T}_g^2 = C_A = 3. \quad (3.2)$$

By dotting the individual color matrices with eq. (3.1), we can solve for the values of the dot products of pairs of color matrices, that correspondingly determine the strength of color connectedness of two jets. We have

$$\mathbf{T}_q \cdot \mathbf{T}_{\bar{q}} = \frac{C_A}{2} - C_F, \quad \mathbf{T}_q \cdot \mathbf{T}_g = \mathbf{T}_{\bar{q}} \cdot \mathbf{T}_g = -\frac{C_A}{2}. \quad (3.3)$$

With this setup, we would like to calculate the differential cross section of the pull vector, or equivalently, the double differential cross section of the pull magnitude t and the pull angle ϕ_p . To do this calculation, we will work to lowest order in the strong coupling α_s and in the leading soft or collinear limits, where the pull magnitude is small, $t \ll 1$. With these approximations, the double differential cross section decomposes into a sum of contributions from soft and collinear emissions:

$$\frac{d^2\sigma^{t \ll 1}}{dt d\phi_p} = S_{q\bar{q}g}(t, \phi_p) + J(t, \phi_p). \quad (3.4)$$

Here, we refer to $S_{q\bar{q}g}(t, \phi_p)$ as the soft function for pull and $J(t, \phi_p)$ as the jet function for pull which encode the contribution to pull from soft and collinear emissions, respectively. Note that no rigorous, all-orders factorization of the pull cross section is assumed or implied here; this decomposition simply follows from the factorization of QCD matrix elements into these components.

The separation of the emission phase space into soft and collinear regions is arbitrary, so one needs to use some regularization scheme to do it. For pull, dimensional regularization is sufficient to uniquely define the soft and jet functions individually. Thus, calculation of the soft function, for example, proceeds by calculating the distribution of pull on the dimensionally-regulated phase space with the single-emission, eikonal matrix element. There are three possible dipoles off of which a soft particle can be emitted, and we need to sum together each of their contributions. We won't present explicit calculations in the text here, but will discuss how the pieces fit together for the complete soft function.

Following the identification of jets in figure 1, one of the dipoles that can emit a soft gluon is the dipole formed from jets 1 and 2. The calculation of this contribution to the soft function was done in ref. [17], with the only necessary change to the case at hand to replace the Casimir C_F in that calculation with $(-\mathbf{T}_1 \cdot \mathbf{T}_2)$ for the three-jet case. Next, the soft gluon could be emitted off of the dipole formed by jets 1 and 3. Note that, relative to jet 2, jet 3 has an azimuthal angle π about jet 1, by momentum conservation. Also, this dipole still contains the jet 1, on which pull is measured, and so we can use the results of ref. [17] again, with two modifications. First, the Casimir C_F in the soft function result of ref. [17] should be replaced by $(-\mathbf{T}_1 \cdot \mathbf{T}_3)$ for the three-jet case. Second, the pull angle ϕ_p should be rotated by π to represent the orientation of jet 3 with respect to jet 2. The contribution of soft gluons emitted off of the dipole formed from jets 2 and 3 is novel, and requires a new calculation. For this case, the dipole that emits the soft gluon does not contain the jet on which pull is measured. As such, this contribution to the soft function lacks a collinear singularity. This calculation is presented in appendix A.

Adding together all three possible sources of soft radiation, the leading-order soft function for pull when $t > 0$ measured on $e^+e^- \rightarrow$ three jet events is:

$$S_{q\bar{q}g}(t, \phi_p) = \frac{\alpha_s}{\pi^2} \frac{1}{t} \left[\mathbf{T}_1^2 \log \frac{\mu^2 \tan^2 \frac{R}{2}}{t^2 E_J^2 \sin^2 \phi_p} \right. \\ \left. + (-\mathbf{T}_1 \cdot \mathbf{T}_2) f(\phi_p, \theta_{12}) + (-\mathbf{T}_1 \cdot \mathbf{T}_3) f(\pi + \phi_p, \theta_{13}) + (-\mathbf{T}_2 \cdot \mathbf{T}_3) g(\phi_p) \right]. \quad (3.5)$$

Here, μ is the dimensional regularization scale and the function $f(\phi_p, \theta)$ was calculated in ref. [17] and is

$$f(\phi_p, \theta) = 2 \cot \phi_p \tan^{-1} \frac{\frac{\tan \frac{R}{2}}{\tan \frac{\theta}{2}} \sin \phi_p}{1 - \frac{\tan \frac{R}{2}}{\tan \frac{\theta}{2}} \cos \phi_p} - \log \left(1 + \frac{\tan^2 \frac{R}{2}}{\tan^2 \frac{\theta}{2}} - 2 \frac{\tan \frac{R}{2}}{\tan \frac{\theta}{2}} \cos \phi_p \right). \quad (3.6)$$

The function $g(\phi_p)$ is calculated in appendix A and is

$$g(\phi_p) = \frac{(\tan \frac{\theta_{12}}{2} + \tan \frac{\theta_{13}}{2})^2}{\tan^2 \frac{\theta_{12}}{2} + \tan^2 \frac{\theta_{13}}{2} + 2 \tan \frac{\theta_{12}}{2} \tan \frac{\theta_{13}}{2} \cos(2\phi_p)} \\ \times \left[\frac{\sin \frac{\theta_{12}-\theta_{13}}{2}}{\sin \frac{\theta_{12}+\theta_{13}}{2}} \log \left(\frac{\frac{\tan^2 \frac{R}{2}}{\tan^2 \frac{\theta_{13}}{2}} + 1 + 2 \frac{\tan \frac{R}{2}}{\tan \frac{\theta_{13}}{2}} \cos \phi_p}{\frac{\tan^2 \frac{R}{2}}{\tan^2 \frac{\theta_{12}}{2}} + 1 - 2 \frac{\tan \frac{R}{2}}{\tan \frac{\theta_{12}}{2}} \cos \phi_p} \right) \right. \\ \left. + 2 \cot \phi_p \tan^{-1} \left(\frac{\left(\frac{\tan \frac{R}{2}}{\tan \frac{\theta_{12}}{2}} - \frac{\tan \frac{R}{2}}{\tan \frac{\theta_{13}}{2}} + 2 \frac{\tan \frac{R}{2}}{\tan \frac{\theta_{12}}{2}} \frac{\tan \frac{R}{2}}{\tan \frac{\theta_{13}}{2}} \cos \phi_p \right) \sin \phi_p}{1 - \left(\frac{\tan \frac{R}{2}}{\tan \frac{\theta_{12}}{2}} - \frac{\tan \frac{R}{2}}{\tan \frac{\theta_{13}}{2}} \right) \cos \phi_p - \frac{\tan \frac{R}{2}}{\tan \frac{\theta_{12}}{2}} \frac{\tan \frac{R}{2}}{\tan \frac{\theta_{13}}{2}} \cos(2\phi_p)} \right) \right]. \quad (3.7)$$

θ_{ij} is the angle between jets i and j . For the coefficient of the logarithmic term in eq. (3.5) above, we have used the conservation of color to express

$$-\mathbf{T}_1 \cdot \mathbf{T}_2 + -\mathbf{T}_1 \cdot \mathbf{T}_3 = \mathbf{T}_1 \cdot (-\mathbf{T}_2 - \mathbf{T}_3) = \mathbf{T}_1^2. \quad (3.8)$$

For a general configuration of the three final state jets, jet 1, on which pull is measured, can be any of the quark, anti-quark, or gluon jet. For the collinear emission contribution to the cross section, we then need to calculate pull as measured on either collinear emissions from quark jets or from gluon jets. The calculation of the quark jet function for pull was presented in ref. [17], while the gluon jet function is novel, and its calculation is presented in appendix B. We can express the jet function for either quarks or gluons as

$$J(t, \phi_p) = \frac{\alpha_s \mathbf{T}_1^2}{\pi^2} \frac{1}{t} \left[\log \frac{4t E_J^2 \sin^2 \phi_p}{\mu^2} - B_1 \right], \quad (3.9)$$

where B_1 comes from hard collinear splittings and is

$$B_q = \frac{3}{4}, \quad B_g = \frac{11}{12} - \frac{n_f}{6C_A}, \quad (3.10)$$

for quark and gluon jets, respectively. Other than the (trivial) $\sin^2 \phi_p$ dependence in the logarithm, collinear emissions are flat in ϕ_p : they are at too small of an angle to know the specific direction of jet 2 and are uncorrelated with any other jets in the event.

Adding the soft and jet functions together, we find the cross section for pull at leading order to be

$$\frac{d^2 \sigma^{t \ll 1}}{dt d\phi_p} = \frac{\alpha_s}{\pi^2} \frac{1}{t} \left[\mathbf{T}_1^2 \log \frac{4 \tan^2 \frac{R}{2}}{t} - \mathbf{T}_1^2 B_1 \right. \\ \left. + (-\mathbf{T}_1 \cdot \mathbf{T}_2) f(\phi_p, \theta_{12}) + (-\mathbf{T}_1 \cdot \mathbf{T}_3) f(\pi + \phi_p, \theta_{13}) + (-\mathbf{T}_2 \cdot \mathbf{T}_3) g(\phi_p) \right]. \quad (3.11)$$

Note that the dimensional regularization scale μ has dropped out; the physical cross section is independent of this unphysical scale.

3.1 Inclusive prediction for pull

With the differential cross section of the pull vector, we can assume that the soft and collinear contribution of the pull vector factorizes from an $e^+e^- \rightarrow q\bar{q}g$ event. Then, the soft and jet functions are regarded as a conditional probability density of the pull vector for observing an additional soft or collinear emission off from one of the three particles in the final state. Thus, the inclusive differential cross section of the pull vector for an $e^+e^- \rightarrow q\bar{q}g + X$ event can be found by integrating over the phase space of the final state particles

$$\frac{d^2 \sigma_{q\bar{q}g}}{dt d\phi_p} = \int_0^1 dx_q \int_0^1 dx_{\bar{q}} \Theta(x_q + x_{\bar{q}} - 1) \frac{1}{\sigma_0} \frac{d^2 \sigma(e^+e^- \rightarrow q\bar{q}g)}{dx_q dx_{\bar{q}}} \sum_{\text{jet orderings}} \frac{d^2 \sigma^{t \ll 1}}{dt d\phi_p} \Theta_{\text{pull}}, \quad (3.12)$$

in which $\frac{d^2 \sigma(e^+e^- \rightarrow q\bar{q}g)}{dx_q dx_{\bar{q}}}$ denotes the cross section for $e^+e^- \rightarrow q\bar{q}g$ in terms of energy fractions x_q and $x_{\bar{q}}$, and Θ_{pull} is the phase space constraints for identifying the two jets on which pull is measured. The differential cross section for $e^+e^- \rightarrow q\bar{q}g$ is

$$\frac{1}{\sigma_0} \frac{d^2 \sigma(e^+e^- \rightarrow q\bar{q}g)}{dx_q dx_{\bar{q}}} = \frac{\alpha_s C_F}{2\pi} \frac{x_q^2 + x_{\bar{q}}^2}{(1-x_q)(1-x_{\bar{q}})}, \quad (3.13)$$

in which the energy fraction x_i of particle i is defined as

$$x_i = \frac{2p_i \cdot Q}{Q^2}, \quad Q = p_1 + p_2 + p_3. \quad (3.14)$$

Note that the angular dependence in the jet and soft function frequently appears as $\tan(\theta_{ij}/2)$. A straightforward algebraic manipulation with dot products shows that

$$\tan\left(\frac{\theta_{ij}}{2}\right) = \sqrt{\frac{1 - \cos \theta_{ij}}{1 + \cos \theta_{ij}}} = \sqrt{\frac{x_i + x_j - 1}{(1-x_i)(1-x_j)}}. \quad (3.15)$$

where $\theta_{ij} \in [0, \pi]$.

With Θ_{pull} , we aim to describe the following algorithms for the pull measurement: (1) identify the two lower energy jets, (2) identify the one of them with a higher energy, and

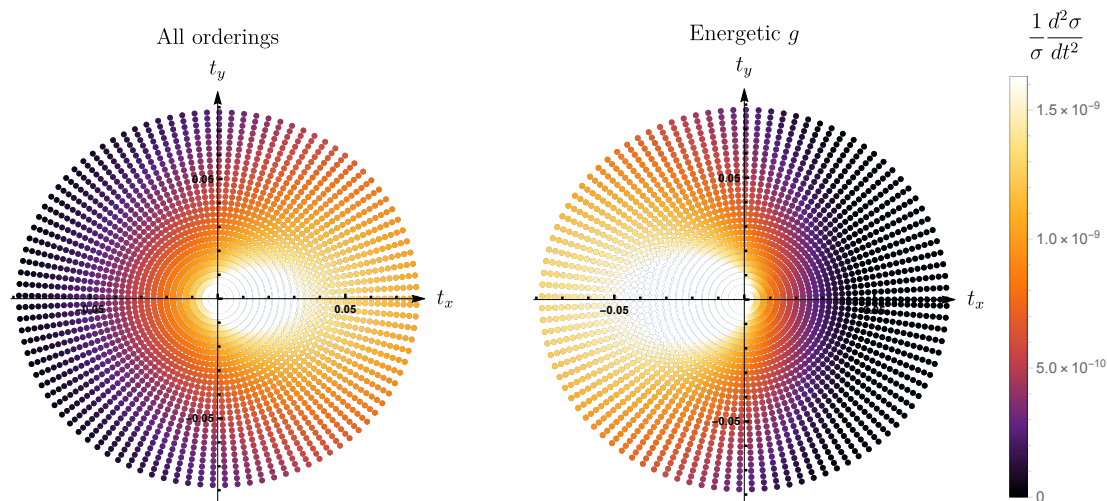


Figure 2. Plots of the relative double differential cross section of the pull vector measured on $e^+e^- \rightarrow q\bar{q}g$ events in the $(t_x, t_y) = (t \cos \phi_p, t \sin \phi_p)$ plane. The colors range over a linear scale with white (black) corresponding to the largest (smallest) cross section. To make these plots, we have set the jet radius $R = 0.5$, the jet energy cut $x_{\text{cut}} = 0.2$ and the number of active fermions $n_f = 5$. (left) Pull vector summed over all six jet orderings. (right) Pull vector when gluon is the most energetic jet.

(3) ensure that those two jets are separated by at least $2R$ such that they are recognized as distinct jets. Therefore, we demand that the following three inequalities hold:

$$E_1 > E_2, \quad E_3 > E_1, \quad \theta_{12} > 2R. \quad (3.16)$$

These can correspondingly be expressed as in terms of the three-body final state energy fractions x_i :

$$\Theta_{\text{pull}} = \Theta(x_3 - x_1)\Theta(x_1 - x_2)\Theta\left(\frac{x_1 + x_2 - 1}{(1 - x_1)(1 - x_2)} - \tan^2 R\right)\Theta(x_2 - x_{\text{cut}}). \quad (3.17)$$

The rightmost Θ -function enforces $x_2 > x_{\text{cut}}$ and ensures that the cross section is IRC safe.

With these phase space restrictions and matrix elements identified, we can then perform the integrals in eq. (3.12), summing over all possible orderings of the q , \bar{q} , and g jets. The result of this calculation is shown on the left in figure 2. In that figure, we have plotted the relative differential cross section, in terms of the components of the pull vector

$$t_x = t \cos \phi_p, \quad t_y = t \sin \phi_p. \quad (3.18)$$

Recall that $\phi_p = 0$ is in the direction of the nearby reference jet. To make this plot, we have set the jet radius $R = 0.5$, the jet energy cut value $x_{\text{cut}} = 0.2$, and the number of active fermions to $n_f = 5$. The cross section is highly peaked about the positive t_x axis, demonstrating that most radiation is present in the region between the two jets. This is to be expected: because of the soft and collinear singularities of the gluon in the $e^+e^- \rightarrow q\bar{q}g$ matrix element, the dominant configuration of the jets is with the gluon as the lowest

energy jet. Therefore, pull is measured about the quark or anti-quark jet, with respect to the direction of the gluon jet, and this pair of jets lives in the $\mathbf{3}$ or $\bar{\mathbf{3}}$ representation of SU(3) color. The product of color factors in this configuration is $-\mathbf{T}_1 \cdot \mathbf{T}_2 = C_A/2 > 0$, and so the cross section peaks around $t_y = 0$ and $t_x > 0$ ($\phi_p = 0$) because the functions $f(\phi_p, \theta)$ and $g(\phi_p)$ have maxima at $\phi_p = 0$.

It is also interesting to restrict to studying the non-dominant jet configuration, by forcing the gluon jet to have the largest energy of all three final state jets. In this configuration, the jets on which pull is measured, the quark and anti-quark, are in the $\mathbf{8}$ representation of SU(3) color. As such, the product of their color matrices is negative, $-\mathbf{T}_q \cdot \mathbf{T}_{\bar{q}} = C_F - C_A/2 = -1/6$, and so the radiation is dominantly outside of the two jets on which pull is measured. This is illustrated in the plot on the right of figure 2. The cross section in this configuration is peaked about the negative t_x axis, as expected.

4 $H \rightarrow b\bar{b}$ vs. $g \rightarrow b\bar{b}$

The results presented in the previous section, along with prior calculations [17], can inform the use of pull for identification of $H \rightarrow b\bar{b}$ decays. One of the original motivations for pull presented in ref. [8] was that it could be used to identify Higgs decays to bottom quarks from the dominant background of gluon splitting to bottom quarks. Because there is no soft singularity for $g \rightarrow b\bar{b}$ splitting, the kinematics of $H \rightarrow b\bar{b}$ and $g \rightarrow b\bar{b}$ are nearly identical, once the mass of the pair of bottom quarks is fixed. Thus, observables sensitive to jet kinematics, such as subjet energy fractions, are not useful for this problem. However, the Higgs boson is a color singlet, while the gluon is a color octet, and this distinction is imprinted on the distribution of soft radiation within and about the pair of bottom quarks. Pull is explicitly sensitive to the orientation of soft radiation, and so can be used to improve identification of Higgs decays.

In this section, we will study the discrimination power of the pull angle ϕ_p for Higgs decays to bottom quarks. We will work in the highly-boosted limit in which the energy or transverse momentum of the Higgs boson is much larger than its mass m_H , so that the bottom quark jets are relatively collimated. Further, we assume that the radii of the individual bottom quark jets R is significantly smaller than the angular separation of the bottom quarks. These limits are relevant and can easily be borne out in practice. For example, for a Higgs boson with $p_\perp = 250$ GeV, the angular separation of the bottom quarks θ_{12} is approximately

$$\theta_{12} \simeq \frac{2m_H}{p_\perp} \simeq 1. \tag{4.1}$$

Correspondingly, subjet radii for individual bottom quarks of $R \lesssim 0.4$ are reasonable as now even down to $R = 0.2$ is used in experiment [21]. Importantly, for the subjets to be well-defined and non-overlapping, their jet radius R should be less than half of the angular separation of the bottom quarks, θ_{12} . Then, taking the $R \ll \theta_{12} \ll 1$ limits of the expression for the pull distribution for a color singlet from ref. [17] we find

$$\frac{d^2\sigma^{H \rightarrow b\bar{b}, R \ll \theta_{12} \ll 1}}{dt d\phi_p} = \frac{\alpha_s C_F}{\pi^2} \frac{1}{t} \left[\log \frac{R^2}{t} - \frac{3}{4} + 4 \frac{R}{\theta_{12}} \cos \phi_p \right] + \mathcal{O}(R^2), \tag{4.2}$$

where $\phi_p \in [0, \pi]$.

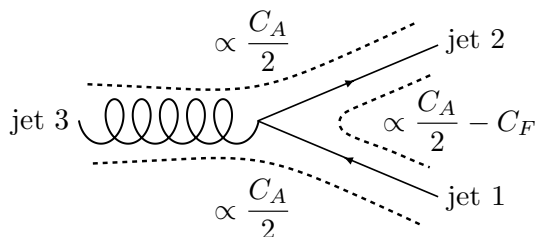


Figure 3. Color flow in the $e^+e^- \rightarrow$ three-jet configuration relevant for comparison to boosted $H \rightarrow b\bar{b}$ in which the two quarks form a color octet. Pull is measured about jet 1 with respect to jet 2.

For the color octet configuration, we use the more general expression presented in eq. (3.11). The configuration of jets we consider produced from e^+e^- collisions is illustrated in figure 3 where we restrict to the configuration in which the quark and anti-quark jets are closest in angle. For expanding to linear order in the jet radius R , we first note that the expansion of the function $g(\phi_p)$ is proportional to R^2 . This contribution comes from a dipole that does not include the jet of interest, so for radiation from this dipole to land in the jet, it must hit an uncorrelated region of area R^2 . Continuing, the contribution proportional to $\mathbf{T}_1 \cdot \mathbf{T}_3$ expands to linear order in R as

$$f(\pi + \phi_p, \theta_{13})|_{R \ll \theta_{13}} = 2R \cot \frac{\theta_{13}}{2} \cos(\pi + \phi_p) + \mathcal{O}(R^2) = -2R \cot \frac{\theta_{13}}{2} \cos(\phi_p) + \mathcal{O}(R^2). \quad (4.3)$$

In the high-boost limit of the two bottom quark jets, $\theta_{13} \rightarrow \pi$, and so $\cot \frac{\theta_{13}}{2} \rightarrow 0$. Therefore, this term is also ignorable to linear order in the jet radius R . The only relevant term in the cross section in these limits is proportional to $\mathbf{T}_1 \cdot \mathbf{T}_2$, which is just the product of the color matrices of the bottom quarks. The relevant color factor is

$$\mathbf{T}_q \cdot \mathbf{T}_{\bar{q}} = \frac{C_A}{2} - C_F. \quad (4.4)$$

Therefore, the pull distribution for the color octet configuration in the small R limit is

$$\frac{d^2\sigma^{g \rightarrow b\bar{b}, R \ll \theta_{12} \ll 1}}{dt d\phi_p} = \frac{\alpha_s C_F}{\pi^2} \frac{1}{t} \left[\log \frac{R^2}{t} - \frac{3}{4} - 4 \left(\frac{C_A}{2C_F} - 1 \right) \frac{R}{\theta_{12}} \cos \phi_p \right] + \mathcal{O}(R^2). \quad (4.5)$$

This has been written with an explicit negative sign in front of the linear in R term because $C_A/(2C_F) - 1 = 1/8 > 0$. Going forward, we will drop the remainder $\mathcal{O}(R^2)$ as it will be implicit in the following.

4.1 Discrimination power of pull

These distributions are already informative, but we would like to determine the distribution of the pull angle ϕ_p alone to identify its power as a discrimination observable. Because the pull angle is not IRC safe, we cannot determine the distribution of ϕ_p by just integrating these distributions over t . However, ϕ_p is Sudakov safe [17, 22, 23], and so we can calculate its distribution by marginalizing against the probability distribution of the pull magnitude,

t . That is, for probability distribution $p(t)$ and conditional probability distribution $p(\phi_p|t)$, the distribution of the pull angle is

$$p(\phi_p) = \int dt p(t) p(\phi_p|t). \quad (4.6)$$

As long as $p(t)$ has no support around $t = 0$, this integral is finite. This is indeed the case for the physical (or resummed) distribution, so eq. (4.6) provides a way to define the distribution of ϕ_p . Here, we will just calculate the conditional probability distribution $p(\phi_p|t)$ to lowest order in the limits we have discussed.

To lowest order, the conditional distribution $p(\phi_p|t)$ is just the ratio of the double differential cross section of t and ϕ_p to the cross section for t alone:

$$p(\phi_p|t) = \frac{\frac{d^2\sigma}{dt d\phi_p}}{\frac{d\sigma}{dt}}. \quad (4.7)$$

Above, we had calculated the distributions for both the $H \rightarrow b\bar{b}$ and $g \rightarrow b\bar{b}$ configurations. To determine the distribution for t exclusively, we can just integrate over ϕ_p . For either the singlet or octet color configurations, the result is the same in the limits in which we work:

$$\frac{d\sigma}{dt} = \int_0^\pi d\phi_p \frac{d^2\sigma}{dt d\phi_p} = \frac{\alpha_s C_F}{\pi} \frac{1}{t} \left[\log \frac{R^2}{t} - \frac{3}{4} \right]. \quad (4.8)$$

Note crucially that the term linear in R integrates to 0. From the expressions for the double differential distributions above, the conditional distributions for the Higgs decay and gluon splitting are:

$$p_{H \rightarrow b\bar{b}}(\phi_p|t) = \frac{\frac{d^2\sigma^{H \rightarrow b\bar{b}}}{dt d\phi_p}}{\frac{d\sigma}{dt}} = \frac{1}{\pi} + \frac{4}{\pi} \frac{1}{\log \frac{R^2}{t} - \frac{3}{4}} \frac{R}{\theta_{12}} \cos \phi_p, \quad (4.9)$$

$$p_{g \rightarrow b\bar{b}}(\phi_p|t) = \frac{\frac{d^2\sigma^{g \rightarrow b\bar{b}}}{dt d\phi_p}}{\frac{d\sigma}{dt}} = \frac{1}{\pi} - \frac{4}{\pi} \frac{\frac{C_A}{2C_F} - 1}{\log \frac{R^2}{t} - \frac{3}{4}} \frac{R}{\theta_{12}} \cos \phi_p. \quad (4.10)$$

We can then determine the distribution of the pull angle. For the Higgs decay, for example, we have

$$p_{H \rightarrow b\bar{b}}(\phi_p) = \int dt p_{H \rightarrow b\bar{b}}(t) p_{H \rightarrow b\bar{b}}(\phi_p|t). \quad (4.11)$$

We are restricting our analysis to linear order in the jet radius R , which will dramatically simplify what follows. As shown above, the pull magnitude distribution $p(t)$ actually has no contribution to it that is linear in R . Thus, in this integral, we only need to keep the terms in $p(t)$ at leading order in the $R \rightarrow 0$ limit. This is correspondingly the collinear limit in which the only thing that the pull magnitude depends on is the flavor of the jet of interest. For both $H \rightarrow b\bar{b}$ and $g \rightarrow b\bar{b}$, the jet of interest is always a quark, and so the distribution of t is identical for the two processes, up to corrections of order R^2 :

$$p_{H \rightarrow b\bar{b}}(t) = p_{g \rightarrow b\bar{b}}(t) + \mathcal{O}(R^2). \quad (4.12)$$

To the order we work, we can then safely set $p_{H \rightarrow b\bar{b}}(t) = p_{g \rightarrow b\bar{b}}(t) \equiv p(t)$, independent of production process.

With this simplification, it follows that the pull angle distribution for Higgs decay is

$$\begin{aligned}
 p_{H \rightarrow b\bar{b}}(\phi_p) &= \int dt p(t) p_{H \rightarrow b\bar{b}}(\phi_p|t) = \int dt p(t) \left(\frac{1}{\pi} + \frac{4}{\pi} \frac{1}{\log \frac{R^2}{t} - \frac{3}{4}} \frac{R}{\theta_{12}} \cos \phi_p \right) \\
 &= \frac{1}{\pi} + d_0 \cos \phi_p.
 \end{aligned} \tag{4.13}$$

Note that the distribution $p(t)$ is normalized and integrates to 1, by definition. We define d_0 as the corresponding moment of the pull magnitude distribution:

$$d_0 = \frac{4}{\pi} \frac{R}{\theta_{12}} \int dt \frac{p(t)}{\log \frac{R^2}{t} - \frac{3}{4}}. \tag{4.14}$$

With this notation, it then follows that the gluon splitting distribution is

$$p_{g \rightarrow b\bar{b}}(\phi_p) = \frac{1}{\pi} - d_0 \left(\frac{C_A}{2C_F} - 1 \right) \cos \phi_p. \tag{4.15}$$

We can estimate the value of d_0 by determining the mean value of the pull magnitude, $\langle t \rangle$. In the collinear limit for a quark jet, this is at lowest order

$$\begin{aligned}
 \langle t \rangle &= \frac{\alpha_s C_F}{2\pi} \int_0^1 dz \int_0^{R^2} \frac{d\theta^2}{\theta^2} \frac{1 + (1-z)^2}{z} z(1-z) |1 - 2z| \theta^2 \\
 &= \frac{13}{64} \frac{\alpha_s C_F}{\pi} R^2.
 \end{aligned} \tag{4.16}$$

Further, the angle between the two quark jets θ_{12} can be approximated from the mass and energy of the singlet resonance. In the collinear or high-boost limit, we have

$$\theta_{12} \simeq \frac{2m_H}{p_\perp}, \tag{4.17}$$

where m_H is the mass of the Higgs and p_\perp is its transverse momentum, assuming it is central in the detector.

Using these results and assuming that the distribution $p(t)$ is highly peaked around its mean, we then have that

$$\begin{aligned}
 d_0 &= \frac{4}{\pi} \frac{R}{\theta_{12}} \int dt p(t) \frac{1}{\log \frac{R^2}{t} - \frac{3}{4}} \simeq \frac{4}{\pi} \frac{1}{\log \frac{R^2}{\langle t \rangle} - \frac{3}{4}} \frac{R}{\theta_{12}} \\
 &= \frac{4}{\pi} \frac{1}{\log \left(\frac{64}{13} \frac{\pi}{\alpha_s C_F} \right) - \frac{3}{4}} \frac{R p_\perp}{2m}.
 \end{aligned} \tag{4.18}$$

Evaluating everything except for the jet radius, transverse momentum, and mass, this is approximately

$$d_0 \simeq 0.2 \frac{R p_\perp}{m_H}, \tag{4.19}$$

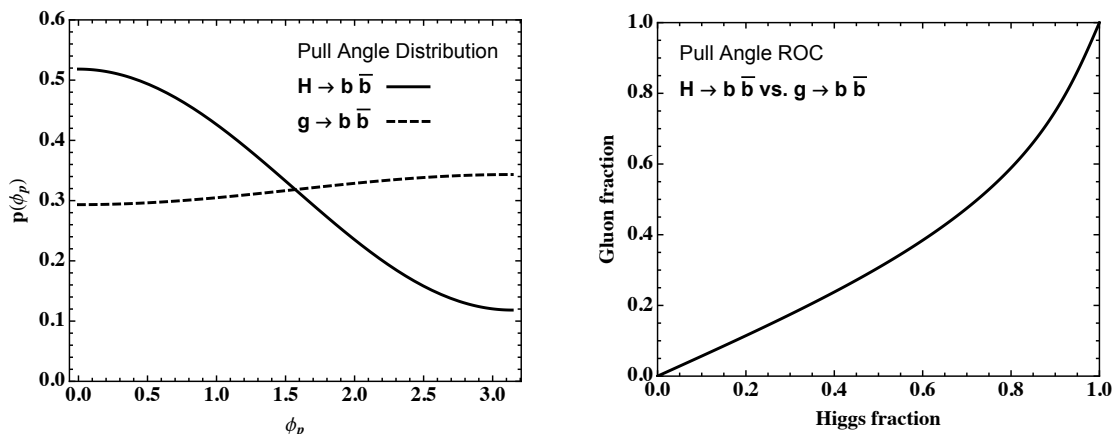


Figure 4. (Left) Distributions of the pull angle from Higgs decay and gluon splitting to bottom quarks in the high-boost and narrow jet limits. We use $d_0 = 0.2$ to make the plots. (Right) Corresponding signal versus background efficiency curve for Higgs decay and gluon splitting to bottom quarks.

where we have used $\alpha_s = 0.1$. Recall that for the two bottom quark jets to be non-overlapping, we require that $R \lesssim m_H/p_\perp$, less than approximately half of the angle between the bottom quark jets. So, d_0 is bounded from above by about 0.2.

From these distributions, we can then quantify the discrimination power of the pull angle by making a sliding cut on the value of ϕ_p . The distributions of the pull angle from Higgs decay and gluon splitting are plotted in figure 4, where we use $d_0 = 0.2$. On the right we plot the signal versus background efficiency curve or receiver operating characteristic (ROC) curve found from keeping those events that have pull angle below a sliding cut. Because the pull angle peaks at small values for $H \rightarrow b\bar{b}$, this procedure amplifies the signal over the background, as manifest by the ROC curve lying below the diagonal.

To quantify the absolute power of the pull angle to discriminate Higgs from gluon splitting, we can integrate under the ROC curve. This area-under-the-curve (AUC) vanishes for perfect discrimination and takes value $1/2$ for identical distributions. The AUC can be calculated as an ordered integral over the two distributions and we find

$$\begin{aligned} \text{AUC}_{(H \text{ vs. } g)} &= \int d\phi_s \int d\phi_b p_{H \rightarrow b\bar{b}}(\phi_s) p_{g \rightarrow b\bar{b}}(\phi_b) \Theta(\phi_s - \phi_b) = \frac{1}{2} - \frac{d_0}{\pi} \frac{C_A}{C_F} \\ &\simeq 0.35. \end{aligned} \tag{4.20}$$

The numerical value on the second line was found from setting $R = m_H/p_T$ and using the value of d_0 identified earlier. For comparison, this value of the AUC is comparable to the value of the AUC for other jet discrimination problems, such as discriminating quark-from gluon-initiated jets. For that problem, the discrimination power of the jet mass as quantified by the AUC at leading logarithmic accuracy is [24]

$$\text{AUC}_{(q \text{ vs. } g)} = \frac{1}{1 + \frac{C_A}{C_F}} \simeq 0.31. \tag{4.21}$$

We also note, however, that these theoretical prediction of metrics may not be borne out in simulation or experiment, but are at least representative of the possible information available in the pull distribution for discrimination.

This simple calculation of course ignores many relevant physical effects that would exist in a real jet and would affect discrimination power. The largest such effect would likely be from soft radiation uncorrelated or only weakly correlated with the direction of the jet. Radiation that lands in the jet from color dipoles that are not color connected to the jet of interest would be approximately uniformly distributed over the area of the jet, with no preferred azimuthal direction. The profile of such radiation in the pull angle ϕ_p would therefore be flat, and work to wash out the distinction between color singlet and octet dipoles studied in this section. To mitigate this effect, one might groom the jet, removing soft, wide-angle radiation in the jet, but preserving radiation collinear to the jet axis. However, collinear radiation is also flat in pull angle ϕ_p to lowest order in the collinear angular size, so this is also likely to wash out these subtle differences. These considerations demonstrate the fragility of color correlations and possibly explain why pull has not been observed to be a useful discriminant in simulation [18–20].

Beyond these practical considerations, one would like to have a formal understanding of the accuracy of the distributions and discrimination metrics derived in this section. While the pull angle ϕ_p is not an IRC safe observable, its Sudakov safety means that one can vary renormalization scales in the integrand of eq. (4.6) to have some estimate of theoretical uncertainties. Such a procedure was also used in in ref. [25] to estimate theoretical uncertainties for the groomed energy fraction z_g [23]. As observed in that case as well, we expect that this scale variation underestimates theoretical uncertainties on the calculated pull angle distribution. Moving away from the small jet radius R limit, we expect that increasing R will likely improve discrimination power to a point. More soft radiation that is sensitive to the dipole configuration will be included in the jets, but so too will more uniform contamination radiation, as mentioned earlier. The effect of contamination radiation scales like the area of the jet, R^2 , while the leading color-correlations between pairs of jets scales like R , so we expect there is some range of R where contamination is small but color correlations are relatively large. Further, higher-order effects like non-global logarithms [26] will likely increase these color correlations present in the pull angle distribution. Non-global effects will pull radiation in the jet toward ends of the dipoles that lie outside of the jet, in principle enhancing differences between the color singlet and octet configurations.

5 Other color representations

The analysis of the previous section suggests a more general result for the pull distribution, appropriate for any combination of a pair of jets on which pull is measured. We denote the jet 1 as the jet on which pull is measured and the jet 2 as the reference jet that defines the origin of the pull angle ϕ_p . In the limits in which the boost of the jet pair is large ($\theta_{12} \ll 1$) and jet radius is small ($R \ll \theta_{12}$), the leading-order expression for the pull distribution is

$$\frac{d^2\sigma^{R \ll \theta_{12} \ll 1}}{dt d\phi_p} = \frac{\alpha_s}{\pi^2} \frac{1}{t} \left[\mathbf{T}_1^2 \log \frac{R^2}{t} - \mathbf{T}_1^2 B_1 + 2 [\mathbf{T}_1^2 + \mathbf{T}_2^2 - (\mathbf{T}_1 + \mathbf{T}_2)^2] \frac{R}{\theta_{12}} \cos \phi_p \right]. \quad (5.1)$$

The coefficient of the term proportional to the jet radius R is just another way to express the product of color matrices:

$$\mathbf{T}_1^2 + \mathbf{T}_2^2 - (\mathbf{T}_1 + \mathbf{T}_2)^2 = -2\mathbf{T}_1 \cdot \mathbf{T}_2. \quad (5.2)$$

We conjecture that this distribution is universal, in the limits $R \ll \theta_{12} \ll 1$ described above. A unique aspect of this distribution is that there is a non-trivial term linear in the jet radius, R . For many (if not nearly all) other observables, the first corrections to the distribution due to a finite jet radius are quadratic in R . This is true of the jet mass, for example, and the magnitude of the pull vector, t . The term proportional to R integrates to 0 on $\phi_p \in [0, \pi]$, and so does not contribute to the pull magnitude's distribution. The universality of this distribution along with the simplicity of color representations of pairs of jets in QCD enables us to explicitly enumerate all possible values for the quadratic Casimir difference, the coefficient of the R term. This is a concrete manifestation and justification for the name “pull”: the difference of quadratic Casimirs explicitly corresponds to how soft radiation is pulled around the jets. If the product representation is smaller than the sum of jets' Casimirs, then radiation is pulled between the pair of jets. By contrast, if the product representation is larger than the pair of jets individually, radiation is pushed out of the pair.

A general irreducible representation of $SU(3)$ can be represented with two non-negative integers m_1 and m_2 and denoted as $D(m_1, m_2)$. The dimension for such a representation is

$$\dim(m_1, m_2) = (1 + m_1)(1 + m_2) \left[1 + \frac{m_1 + m_2}{2} \right]. \quad (5.3)$$

The quadratic Casimir of this representation is

$$C_{(m_1, m_2)} = \mathbf{T}_{(m_1, m_2)}^2 = \frac{1}{3} (m_1^2 + m_2^2 + 3m_1 + 3m_2 + m_1 m_2) \quad (5.4)$$

The fundamental and adjoint representations are $\mathbf{3} = D(1, 0)$ and $\mathbf{8} = D(1, 1)$, respectively, and these formulae give the correct values for the dimension and Casimir of these representations. Exhaustive information about the representation theory of $SU(3)$ can be found in ref. [27]. The jets that form the pair on which pull is measured can only be some combination of quarks and gluons in QCD, so enumerating the possible product representations of color $SU(3)$ that can appear is a simple exercise with $SU(3)$ Clebsch-Gordan coefficients.

In table 1, we list all possible QCD jet pairs, the irreps of $SU(3)$ color that appear in their corresponding product representation, and then the value of the difference of quadratic Casimirs, using the formula presented earlier. For most of the representations in the table, the difference of Casimirs is positive, indicating that the pull angle distribution peaks at $\phi_p = 0$; that is, most radiation lies between the jet pair. Only the highest dimension product representations produce negative Casimir differences, indicating that most radiation in this case is emitted outside of the region between the jet pair. Intriguingly, the $\mathbf{10}$ representation of the color of a pair of gluons exhibits a perfectly flat pull angle distribution in this limit. Apparently this representation corresponds to exactly the same amount of radiation between as outside of the pair of gluon jets.

Jet Pair	Product Irrep	$\mathbf{T}_1^2 + \mathbf{T}_2^2 - (\mathbf{T}_1 + \mathbf{T}_2)^2$
$q\bar{q}$	1	$\frac{8}{3}$
	8	$-\frac{1}{3}$
qq	$\bar{\mathbf{3}}$	$\frac{4}{3}$
	6 = $D(2, 0)$	$-\frac{2}{3}$
qg	3	3
	$\bar{\mathbf{6}}$	1
	15 = $D(2, 1)$	-1
gg	1	6
	8	3
	10 = $D(3, 0)$	0
	27 = $D(2, 2)$	-2

Table 1. Table of the possible jet pairs in QCD on which pull can be measured. The irreps of color SU(3) that appear in the corresponding product representation are presented in the middle column. We have only listed those irreps that correspond to unique values of the quadratic Casimir, e.g., **10** and **$\bar{\mathbf{10}}$** have the same quadratic Casimir. In the final column, we calculate the “pull” of the product representation; the difference between the individual quadratic Casimirs of the two jets and their product representation.

6 Conclusions

The pull observable was designed to be sensitive to the flow of color between pairs of jets and thus sensitive to their product representation of SU(3) color. This has been studied in simulation extensively and motivated measurements, but had not been justified theoretically. In this paper, we demonstrated that pull, especially the pull angle, takes on a different distribution for pairs of jets in distinct product representations of color. We performed explicit calculations at leading order for pull measured on $e^+e^- \rightarrow$ three jets events, studied the discrimination power of pull for identification of $H \rightarrow b\bar{b}$ decays, and presented a conjecture for the pull distribution in the high-boost, small jet radius limit.

The results presented in this paper suggest a number of extensions. Observables that are more sensitive to color flow between jets can be designed, motivated by recent work in machine learning for particle physics [28–30]. In particular, soft, wide-angle radiation in a jet is most sensitive to the colors of the other jets in an event, and to leading order, the distribution of this radiation is described by eikonal matrix elements. With these eikonal matrix elements, one can construct the theoretically-optimal observable for discrimination of, say, a pair of jets that form a color singlet from a pair of jets that do not. This optimal observable is the likelihood ratio by the Neyman-Pearson lemma [31], and in general is not the pull angle. Designing such observables may also resolve issues regarding residual color flow information in machine learning studies, even for jets on which pull is measured.

Prospects for observation of other hadronic decays of the Higgs boson could potentially be improved by using pull, or related color flow observables. The Standard Model Higgs boson decays to pairs of gluons nearly 10% of the time, and yet the $H \rightarrow gg$ decay

mode is extremely challenging to observe. Because gluons carry more color individually than quarks, the strength of color connection between the gluons in $H \rightarrow gg$ decays is substantially larger than between the bottom quarks in $H \rightarrow b\bar{b}$ decays, as shown in table 1. This may suggest that it is easier to discriminate the gg color singlet representation from non-singlet color representations; however, this may also mean that identification of the $H \rightarrow gg$ decay at high boost is more challenging to identify because the two hard prongs in the jet are less well-defined.

Finally, as the pull angle is not IRC safe, its calculation relies on resummed multi-differential cross sections to be well-defined. Thus, ideally one would like the two-dimensional resummed cross section for the pull vector, from which the pull angle can be defined by marginalization. How this resummation would proceed for different color configurations of jets would be interesting to determine. Further, measurements of pull on $g \rightarrow b\bar{b}$ decay, for example, would test the extent to which the results derived in this paper were accurate at all at describing reality. This then may point to a whole new class of observables that can be used to study global correlations in particle collision events.

Acknowledgments

A. L. thanks Simone Marzani and Chang Wu for collaboration on related projects and extensive discussions and Ben Nachman for comments on the draft. Y.B. was supported by the physics research award from the Department of Physics at Reed College.

A Non-connected soft function calculation

The double differential soft function for emissions from a dipole, neither of whose ends include the jet on which we measure the pull, can be calculated from

$$S_{23}(t, \phi_p) = (-\mathbf{T}_2 \cdot \mathbf{T}_3) g^2 \mu^{2\epsilon} \int [d^d k]_+ \frac{2n_2 \cdot n_3}{(k \cdot n_2)(k \cdot n_3)} \Theta_{\text{jet},R} \delta_t \delta_{\phi_p}. \quad (\text{A.1})$$

Here, we call the two ends of the dipole 2 and 3 and the phase space constraints $\Theta_{\text{jet},R}$, δ_t , and δ_{ϕ_p} are the jet radius R constraint, the measurement of the pull vector magnitude, and the measurement of the pull angle, respectively. The light-like vectors n_2 and n_3 have unit 0th component and point along the direction of particles 2 and 3. Three-jet production in e^+e^- collisions in the center-of-mass frame is restricted to a plane, which simplifies the expression of the integrand. With $\overline{\text{MS}}$ dimensional regularization, the soft function can be written as

$$S_{23}(t, \phi_p) = (-\mathbf{T}_2 \cdot \mathbf{T}_3) \frac{\alpha_s}{\pi^{3/2} \Gamma(1/2 - \epsilon)} \mu^{2\epsilon} \int_0^\infty dk_\perp k_\perp^{-1-2\epsilon} \int_{-\infty}^\infty d\eta \int_0^\pi d\phi \sin^{-2\epsilon} \phi \quad (\text{A.2})$$

$$\times \frac{1 - \cos \theta_{23}}{(\cosh \eta - \cos \phi \sin \theta_{12} - \sinh \eta \cos \theta_{12})(\cosh \eta + \cos \phi \sin \theta_{13} - \sinh \eta \cos \theta_{13})}$$

$$\times \Theta \left(\tan \frac{R}{2} - e^{-\eta} \right) \delta(\phi_p - \phi) \delta \left(t - \frac{k_\perp}{E_J \cosh \eta} \right).$$

Here, E_J is the energy of the jet of interest, θ_{12} is the angle between the jet on which pull is measured and jet 2 (and correspondingly for θ_{13} and θ_{23}). Note the difference in sign in the $\cos \phi$ term in the two factors in the denominator of the matrix element term: jet 2 has an azimuthal angle of $\phi_2 = 0$ about jet 1 while jet 3 has an azimuthal angle of $\phi_3 = \pi$ about jet 1. This assignment follows from the fact that jet 2 defines the location of the origin of the pull angle and that the collision occurs in the center-of-mass frame and the final state is confined to a plane.

The integrals that remain are finite for $\epsilon \rightarrow 0$, so we can just set $\epsilon = 0$ to calculate the corresponding pull distribution for $t > 0$. We find

$$\begin{aligned}
 S_{23}(t, \phi_p) &= (-\mathbf{T}_2 \cdot \mathbf{T}_3) \frac{\alpha_s}{\pi^2} \frac{1}{t} \frac{(\tan \frac{\theta_{12}}{2} + \tan \frac{\theta_{13}}{2})^2}{\tan^2 \frac{\theta_{12}}{2} + \tan^2 \frac{\theta_{13}}{2} + 2 \tan \frac{\theta_{12}}{2} \tan \frac{\theta_{13}}{2} \cos(2\phi_p)} \quad (\text{A.3}) \\
 &\times \left[\frac{\sin \frac{\theta_{12} - \theta_{13}}{2}}{\sin \frac{\theta_{12} + \theta_{13}}{2}} \log \left(\frac{\tan^2 \frac{R}{2} + 1 + 2 \frac{\tan \frac{R}{2}}{\tan \frac{\theta_{13}}{2}} \cos \phi_p}{\tan^2 \frac{R}{2} + 1 - 2 \frac{\tan \frac{R}{2}}{\tan \frac{\theta_{12}}{2}} \cos \phi_p} \right) \right. \\
 &\left. + 2 \cot \phi_p \tan^{-1} \left(\frac{\left(\frac{\tan \frac{R}{2}}{\tan \frac{\theta_{12}}{2}} - \frac{\tan \frac{R}{2}}{\tan \frac{\theta_{13}}{2}} + 2 \frac{\tan \frac{R}{2}}{\tan \frac{\theta_{12}}{2}} \frac{\tan \frac{R}{2}}{\tan \frac{\theta_{13}}{2}} \cos \phi_p \right) \sin \phi_p}{1 - \left(\frac{\tan \frac{R}{2}}{\tan \frac{\theta_{12}}{2}} - \frac{\tan \frac{R}{2}}{\tan \frac{\theta_{13}}{2}} \right) \cos \phi_p - \frac{\tan \frac{R}{2}}{\tan \frac{\theta_{12}}{2}} \frac{\tan \frac{R}{2}}{\tan \frac{\theta_{13}}{2}} \cos(2\phi_p)} \right) \right] \\
 &= (-\mathbf{T}_2 \cdot \mathbf{T}_3) \frac{\alpha_s}{\pi^2} \frac{1}{t} \left(\frac{R^2}{4} \left(\cot \frac{\theta_{12}}{2} + \cot \frac{\theta_{13}}{2} \right)^2 + \mathcal{O}(R^3) \right).
 \end{aligned}$$

In the final line, we have Taylor expanded the expression in powers of the jet radius R . As expected, because there is no collinear singularity in this dipole configuration, the soft function is proportional to the area of the jet, R^2 . One can also find a closed-form analytic expression for the more general case in which the jets 1, 2, and 3 do not lie in a plane, but we will not present it here.

B Gluon jet function

To calculate the general distribution of pull measured on any jet in $e^+e^- \rightarrow$ three jets events, we must include the possibility of the jet being a gluon. So, for the complete calculation, we also need to calculate the distribution of pull from collinear emissions in a gluon jet. The $d = 4 - 2\epsilon$, dimensionally-regulated, $\overline{\text{MS}}$ gluon jet function on which we measure pull is

$$\begin{aligned}
 J_g(t, \phi_p) &= \frac{\alpha_s}{2\pi} \frac{1}{\pi^{1/2} \Gamma(1/2 - \epsilon)} \left(\frac{\mu^2}{E_J^2} \right)^\epsilon \int_0^1 dz \int_0^\infty d\theta^2 (\theta^2)^{-1-\epsilon} \int_0^\pi d\phi \sin^{-2\epsilon} \phi \quad (\text{B.1}) \\
 &\times z^{-2\epsilon} (1-z)^{-2\epsilon} \left[C_A \left(\frac{1}{z} + \frac{1}{1-z} + z(1-z) - 2 \right) + \frac{n_f}{2} \left(1 - \frac{2}{1-\epsilon} z(1-z) \right) \right] \\
 &\times \delta(t - z(1-z)|1 - 2z|\theta^2) \delta(\phi_p - \phi).
 \end{aligned}$$

n_f is the number of active fermions. Performing the integrals over the δ -functions, we then have

$$J_g(t, \phi_p) = \frac{\alpha_s}{2\pi} \frac{1}{\pi^{1/2}\Gamma(1/2 - \epsilon)} \left(\frac{\mu^2}{E_J^2 \sin^2 \phi_p} \right)^\epsilon \frac{1}{t^{1+\epsilon}} \int_0^1 dz z^{-1-\epsilon} \quad (\text{B.2})$$

$$\times (1-z)^{-\epsilon} |1-2z|^\epsilon \left[C_A (2 + z^2(1-z) - 2z) + \frac{n_f}{2} \left(z - \frac{2}{1-\epsilon} z^2(1-z) \right) \right].$$

In this expression, we have also symmetrized the first two terms of the splitting function, to isolate the divergence at $z = 0$. To integrate over z , we can expand the first factor in $+$ -functions:

$$z^{-1-\epsilon} = -\frac{1}{\epsilon} \delta(z) + \left(\frac{1}{z} \right)_+ + \dots \quad (\text{B.3})$$

The integral with the $\delta(z)$ is just $2C_A$. For the integral with the $+$ -function, we can set $\epsilon = 0$ and we have

$$\int_0^1 dz \left(\frac{1}{z} \right)_+ \left[C_A (2 + z^2(1-z) - 2z) + \frac{n_f}{2} (z - 2z^2(1-z)) \right] \quad (\text{B.4})$$

$$= \int_0^1 dz \left[C_A (z(1-z) - 2) + \frac{n_f}{2} (1 - 2z(1-z)) \right]$$

$$= -\frac{11}{6} C_A + \frac{n_f}{3}.$$

With these results, the jet function is

$$J_g(t, \phi_p) = \frac{\alpha_s C_A}{\pi} \frac{1}{\pi^{1/2}\Gamma(1/2 - \epsilon)} \left(\frac{\mu^2}{E_J^2 \sin^2 \phi_p} \right)^\epsilon \frac{1}{t^{1+\epsilon}} \left[-\frac{1}{\epsilon} - \frac{11}{12} + \frac{n_f}{6C_A} \right]. \quad (\text{B.5})$$

Only keeping those terms that contribute for $t > 0$, the jet function is then

$$J_g(t, \phi_p) = \frac{\alpha_s C_A}{\pi^2} \frac{1}{t} \left[\log \frac{4t E_J^2 \sin^2 \phi_p}{\mu^2} - \frac{11}{12} + \frac{n_f}{6C_A} \right]. \quad (\text{B.6})$$

Open Access. This article is distributed under the terms of the Creative Commons Attribution License ([CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/)), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.

References

- [1] R.D. Field and R.P. Feynman, *A parametrization of the properties of quark jets*, *Nucl. Phys.* **B 136** (1978) 1 [[INSPIRE](#)].
- [2] D. Krohn, M.D. Schwartz, T. Lin and W.J. Waalewijn, *Jet charge at the LHC*, *Phys. Rev. Lett.* **110** (2013) 212001 [[arXiv:1209.2421](#)] [[INSPIRE](#)].
- [3] G. Hanson et al., *Evidence for jet structure in hadron production by e^+e^- annihilation*, *Phys. Rev. Lett.* **35** (1975) 1609 [[INSPIRE](#)].
- [4] TASSO collaboration, *Evidence for planar events in e^+e^- annihilation at high-energies*, *Phys. Lett.* **86B** (1979) 243 [[INSPIRE](#)].

- [5] D.P. Barber et al., *Discovery of three jet events and a test of quantum chromodynamics at PETRA energies*, *Phys. Rev. Lett.* **43** (1979) 830 [INSPIRE].
- [6] PLUTO collaboration, *Evidence for gluon bremsstrahlung in e^+e^- annihilations at high-energies*, *Phys. Lett.* **86B** (1979) 418 [INSPIRE].
- [7] JADE collaboration, *Observation of planar three jet events in e^+e^- annihilation and evidence for gluon Bremsstrahlung*, *Phys. Lett.* **B 91** (1980) 142.
- [8] J. Gallicchio and M.D. Schwartz, *Seeing in color: jet superstructure*, *Phys. Rev. Lett.* **105** (2010) 022001 [arXiv:1001.5027] [INSPIRE].
- [9] D0 collaboration, *Measurement of color flow in $t\bar{t}$ events from $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV*, *Phys. Rev. D* **83** (2011) 092002 [arXiv:1101.0648] [INSPIRE].
- [10] ATLAS collaboration, *Measurement of colour flow with the jet pull angle in $t\bar{t}$ events using the ATLAS detector at $\sqrt{s} = 8$ TeV*, *Phys. Lett.* **B 750** (2015) 475 [arXiv:1506.05629] [INSPIRE].
- [11] ATLAS collaboration, *Measurement of colour flow using jet-pull observables in $t\bar{t}$ events with the ATLAS experiment at $\sqrt{s} = 13$ TeV*, *Eur. Phys. J. C* **78** (2018) 847 [arXiv:1805.02935] [INSPIRE].
- [12] CMS collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, *Phys. Lett.* **B 716** (2012) 30 [arXiv:1207.7235] [INSPIRE].
- [13] CMS collaboration, *Search for the standard model Higgs boson produced in association with a W or a Z boson and decaying to bottom quarks*, *Phys. Rev. D* **89** (2014) 012003 [arXiv:1310.3687] [INSPIRE].
- [14] CMS collaboration, *Search for invisible decays of Higgs bosons in the vector boson fusion and associated ZH production modes*, *Eur. Phys. J. C* **74** (2014) 2980 [arXiv:1404.1344] [INSPIRE].
- [15] CMS collaboration, *Measurement of electroweak production of two jets in association with a Z boson in proton-proton collisions at $\sqrt{s} = 8$ TeV*, *Eur. Phys. J. C* **75** (2015) 66 [arXiv:1410.3153] [INSPIRE].
- [16] CMS collaboration, *Search for dark matter in proton-proton collisions at 8 TeV with missing transverse momentum and vector boson tagged jets*, *JHEP* **12** (2016) 083 [Erratum *ibid.* **08** (2017) 035] [arXiv:1607.05764] [INSPIRE].
- [17] A.J. Larkoski, S. Marzani and C. Wu, *Theory predictions for the pull angle*, *Phys. Rev. D* **99** (2019) 091502 [arXiv:1903.02275] [INSPIRE].
- [18] ATLAS collaboration, *Reconstruction and modelling of jet pull with the ATLAS detector*, ATLAS-CONF-2014-048 (2014).
- [19] L. de Oliveira et al., *Jet-images — deep learning edition*, *JHEP* **07** (2016) 069 [arXiv:1511.05190] [INSPIRE].
- [20] J. Lin, M. Freytsis, I. Moutl and B. Nachman, *Boosting $H \rightarrow b\bar{b}$ with machine learning*, *JHEP* **10** (2018) 101 [arXiv:1807.10768] [INSPIRE].
- [21] ATLAS collaboration, *Identification of boosted Higgs bosons decaying into b-quark pairs with the ATLAS detector at 13 TeV*, *Eur. Phys. J. C* **79** (2019) 836 [arXiv:1906.11005] [INSPIRE].

- [22] A.J. Larkoski and J. Thaler, *Unsafe but calculable: ratios of angularities in perturbative QCD*, *JHEP* **09** (2013) 137 [[arXiv:1307.1699](#)] [[INSPIRE](#)].
- [23] A.J. Larkoski, S. Marzani and J. Thaler, *Sudakov safety in perturbative QCD*, *Phys. Rev. D* **91** (2015) 111501 [[arXiv:1502.01719](#)] [[INSPIRE](#)].
- [24] A.J. Larkoski, G.P. Salam and J. Thaler, *Energy correlation functions for jet substructure*, *JHEP* **06** (2013) 108 [[arXiv:1305.0007](#)] [[INSPIRE](#)].
- [25] A. Tripathy et al., *Jet substructure studies with CMS open data*, *Phys. Rev. D* **96** (2017) 074003 [[arXiv:1704.05842](#)] [[INSPIRE](#)].
- [26] M. Dasgupta and G.P. Salam, *Resummation of nonglobal QCD observables*, *Phys. Lett. B* **512** (2001) 323 [[hep-ph/0104277](#)] [[INSPIRE](#)].
- [27] J.J. de Swart, *The octet model and its Clebsch-Gordan coefficients*, *Rev. Mod. Phys.* **35** (1963) 916 [*Erratum ibid.* **37** (1965) 326] [[INSPIRE](#)].
- [28] A.J. Larkoski, I. Moult and B. Nachman, *Jet substructure at the Large Hadron Collider: a review of recent advances in theory and machine learning*, [arXiv:1709.04464](#) [[INSPIRE](#)].
- [29] D. Guest, K. Cranmer and D. Whiteson, *Deep learning and its application to LHC physics*, *Ann. Rev. Nucl. Part. Sci.* **68** (2018) 161 [[arXiv:1806.11484](#)] [[INSPIRE](#)].
- [30] A. Radovic et al., *Machine learning at the energy and intensity frontiers of particle physics*, *Nature* **560** (2018) 41.
- [31] J. Neyman and E.S. Pearson, *On the problem of the most efficient tests of statistical hypotheses*, *Phil. Trans. Roy. Soc. Lond.* **A 231** (1933) 289.