Measuring collinear $W$ emissions inside jets

Frank Krauss,$^1$ Petar Petrov,$^1$ Marek Schönherr,$^1$ and Michael Spannowsky$^{1,}$

$^1$Institute for Particle Physics Phenomenology, Department of Physics, Durham University, DH1 3LE, United Kingdom

Single and multiple emission of electroweak gauge bosons and in particular of $W$ bosons is discussed in the parton shower language. Algorithms and observables for the reconstruction of both leptonically and hadronically decaying $W$ bosons inside light quark jets are compared, and they are applied to a study of how emission rates of $W$ bosons in light-jet events at the LHC could be measured.

I. INTRODUCTION

After the recent discovery of a Standard Model-like Higgs boson by ATLAS and CMS\cite{1,2}, the main focus of the LHC’s upcoming 14 TeV run will be on further precision studies of this newly found particle to fully establish all of its predicted properties. In addition, of course, the hunt for new physics phenomena and, possibly, heavy new particles will continue. Many models for such new physics predict the existence of short-lived TeV-scale resonances, which predominantly decay to those particles in the Standard Model that are thought to be most sensitive to electroweak-scale particles, e.g. the top quark, and the $W$, $Z$, or Higgs bosons. The anticipated large masses of the primary resonances imply that their decay products are highly boosted; that is, they tend to have relatively large momenta and therefore quite often momenta transverse to the beam axes which are significantly larger than their mass. As a consequence, many search strategies for such new particles by now concentrate on signatures including a number of highly boosted weak gauge bosons accompanied by jets. However, the findings up to now, or, better put, their absence has placed a number of stringent lower mass bounds on such new heavy particles, in general of $\mathcal{O}$(TeV). This in turn pushes their anticipated masses even higher and it also translates into even larger boosts. This poses an interesting challenge, since obviously increasing the boost of a heavy decaying particle directly translates into decreasing distances of its decay products, and traditional strategies for their isolation start to become obsolete\cite{3-10}.

In this situation, precise background predictions are of utmost importance to maximize the sensitivity in searches with leptons, missing transverse energy and jets at large center-of-mass energies. Especially when electroweak–scale resonances are highly boosted theory uncertainties become large: firstly, higher–order QCD corrections may open up hitherto unconsidered channels or allow for more extreme kinematics\cite{11,12}; in addition, virtual electroweak corrections will become increasingly important with increasing scales tested in a process. This can be seen from a rough estimate, approximating such corrections through their leading logarithmic behavior as $\alpha_W \ln^2(\hat{s}/m_W^2)$, where $\hat{s}$ is the typical scale of the process\cite{13-23}. The large logarithms typically discussed in this first round of publications emerge due to a non-cancellation of real and virtual contributions. Neglecting, physically justified, the quite distinct emission of real heavy gauge bosons leads to large Sudakov–type logarithms of the form above, which can be resummed through exponentiation, similar to the QCD case. And, in quite good analogy with the better known case of QCD, real emissions tend to cancel these logarithms. In contrast to QCD, however, the large masses of the heavy gauge bosons prevent a complete cancellation due to the phase space constraints imposed by them on the real radiation pattern. This has been studied in some detail in\cite{24}, showing that more careful consideration has to be placed on how inclusive processes are being studied. This first study has been supplemented by a similar, but more precise consideration of such partial cancellation effects, including terms of higher logarithmic accuracy in\cite{25}. Of course, with the advent of improved calculational technologies, especially for the virtual (loop) contributions, such effects can be studied at complete next-to leading order accuracy in a very coherent fashion, for some recent work, cf.\cite{26-28}.

However, in searches for new physics, the emission of multiple gauge bosons and the interplay of real and virtual corrections to such processes may become an important aspect, which clearly stretches beyond the next-to leading order accuracy, including on one gauge boson only, marking the current level of typical precision calculations in the electroweak sector. And while fully automated tree-level matrix element generators such as ALPGEN\cite{29},

\begin{itemize}
\item Electronic address: frank.krauss@durham.ac.uk
\item Electronic address: petar.petrov@durham.ac.uk
\item Electronic address: marel.schoenherr@durham.ac.uk
\item Electronic address: michael.spannowsky@durham.ac.uk
\end{itemize}
**II. ELECTROWEAK SHOWER AND EVENT GENERATION**

To evaluate how well resummed electroweak corrections can be measured at the LHC we study the production of \( W \) bosons in the parton shower on top of QCD dijet events, \( pp \to jj \), at \( \sqrt{s} = 14 \) TeV. We use the SHERPA [37] event generator with the modifications detailed below implemented on top of version 2.1.0. The hard scattering matrix elements are computed by COMIX [32] and showered using the modified CSSHOWER [38, 39]. Hadronization [40] and the underlying event [41] are taken into account, including hadron decays. Higher order QED corrections are accounted for both in hadron and leptonic gauge boson decays [42]. The CT10 parton distribution functions [43] have been used.

The following describes the implementation of the radiation of electroweak gauge bosons in the parton shower formalism as included in SHERPA. Some critical observables are defined which allow the reconstruction of the emitted gauge bosons in Sec. III, where we focus both on leptonically or hadronically decaying \( W \) bosons. In Sec. IV we estimate how well LHC’s multi-purpose experiments can measure emissions of collinear electroweak bosons in run 2. We summarize and add a concluding discussion in Sec. V.

\[
\frac{d\sigma_{n+V}}{dt\,dz} = \frac{1}{n_{\text{spec}}} \sum_{f} \sum_{s} \int\! \frac{d\phi}{2\pi} \frac{1}{n_{\text{spec}}} \frac{d\phi}{2\pi} \frac{1}{n_{\text{spec}}} \int\! J(t, z) K_{f(s)\to f(s)\to V(s)}(t, z),
\]

wherein the labels \( f \) and \( s \) run over all fermions of the \( n \)-particle configuration and signify the emitter and spectator fermions, with \( n_{\text{spec}} \) the number of spectators. In the collinear limit any number of spectators, and in particular the choice of one spectator only would be a valid choice. However, in order to maintain a dipole–like formulation all particles with electroweak charges present acceptable choices as spectators. The emission phase space is parametrized in terms of the evolution variable \( t \), the splitting variable \( z \) and the azimuthal angle \( \phi \). Their precise definition, depending on whether \( f \) and/or \( s \) are in the initial or final state, as well as the values of the Jacobean factors \( J(t, z) \) can be found in [39]. The latter also contain a ratio of parton distribution functions to account for the possible change in initial state parton flavors and/or momenta.

The splitting functions \( K \) are an adaptation of the expressions calculated in [48], cast in a form suitable for SHERPA’s CSSHOWER. For the collinear emission of an electroweak gauge boson, \( W^\pm \) or \( Z \), off a fermion or anti-fermion \( f \) in

---

* Further implementations of this or similar algorithms for QCD multi–particle final state have been described in [51,57].
the presence of a spectator $s$ in the high energy limit $E \gg m_V$ they read

\[ K_{f(s)\to f'W}(t,z) = \frac{\alpha}{2\pi} \left[ f_W c_\perp V_{\text{CDST}}^{f(s)\to f'b(s)}(t,z) + f_h c_\perp^W \frac{1}{2} (1 - z) \right] \]

\[ K_{f(s)\to fZ}(t,z) = \frac{\alpha}{2\pi} \left[ f_Z c_\perp V_{\text{CDST}}^{f(s)\to f'b(s)}(t,z) + f_h c_\perp^Z \frac{1}{2} (1 - z) \right]. \tag{2} \]

Therein, $V_{\text{CDST}}^{f(s)\to f'b(s)}$ are the spin–averaged Catani–Dittmaier–Seymour–Trocsanyi (massive) dipole splitting kernels, neglecting their color factors $K_{f(s)\to f'b(s)}$ and $\alpha$ is the QED coupling constant. $c_\perp$ and $c_L$ are the coupling factors of the transverse and longitudinal gauge boson polarizations. They are given by

\[
c_\perp^W = s_{\text{eff}} \frac{1}{2s_W} |V_{f'f}|^2, \quad c_\perp^Z = s_{\text{eff}} \frac{s_W^2}{c_W^2} Q_2^2 + (1 - s_{\text{eff}}) \frac{(t_{z}' - s_W^2 Q_0)^2}{s_W c_W^2}, \quad c_L^W = \frac{t_z^2}{s_W} \frac{m_q^2}{m_W}, \quad c_L^Z = \frac{t_z^2}{s_W} \frac{m_q^2}{m_W}, \tag{3}\]

with $Q_f$, $t_z$ and $m_f$ the charge, the three–component of the weak isospin and the mass of the respective fermion. $s_W^2$ and $c_W^2$ are the squared sine and cosine of the Weinberg angle, and $m_W$ is the mass of the $W$ boson. $V_{f'f}$ is the CKM matrix element in case of emissions off quarks and the unit matrix in case of emissions off leptons. $s_{\text{eff}}$ is the fraction of left–handed fermions in the spin–averaged fermion line. As the coupling to $W^\pm$ and $Z$ bosons differs for left– and right–handed fermions but the shower operates in a spin-averaged approximation, this factor $s_{\text{eff}}$ is essential for a correct description of the splitting probabilities. It is important to note that a global definition of $s_{\text{eff}}$ is only sensible in a limited set of processes: i) pure QCD/QED reactions which are left-right symmetric ($s_{\text{eff}} = \frac{1}{2}$), and ii) electroweak processes where all fermion lines are connected to the same type of electroweak gauge bosons, e.g. $qq' \to e\nu e\nu q'$, a global $s_{\text{eff}}$ cannot be defined. This, of course, directly implies that in the present case of dijet production, after the first EW gauge boson is radiated the global $s_{\text{eff}}$ is only correct for all fermion lines not connected to the one that radiated the first EW boson. Radiation off this quark line will be underestimated as its now polarized state is not accounted for. This is a general problem when embedding electroweak splittings into spin–averaged parton showers, and we do not suggest a solution here. In the present case, however, it is only a minor effect due to the small radiation probabilities and does not affect the outcome of this analysis.

As can be seen, the couplings of the longitudinal modes are derived through Goldstone boson equivalence, therefore, they only couple to the fermion mass. In the case of light jet production studied here they can therefore be neglected to a very good approximation. The $W^\pm$ and $Z$ boson masses enter via the splitting kinematics, using the recoil scheme the construction of the splitting kinematics detailed in [??]. All $W^\pm$ and $Z$ bosons produced are decayed immediately and all decay channels are considered. Although this neglects additional radiative branchings of the type $W^\pm \to W^\pm q$, $W^\pm \to W^\pm Z$ or $Z \to W^\pm W^\mp$, it ensures that the color singlet $q\bar{q}^{(i)}$ pair produced in the hadronic decay modes starts its evolution at the correct scale of $m_W$ or $m_Z$, respectively. Higgs radiation, can also be neglected to a very good approximation, similar to the longitudinal boson polarization modes.

The factors $f_W$, $f_Z$ and $f_h$, which have been added by hand in Eq. (2), are used in the analysis to modify the coupling strength of the different electroweak bosons to fermions. Their Standard Model values of course are all unity, $f_W = f_Z = f_h = 1$. For the present study it has been found that contributions from the radiation of $Z$ or Higgs bosons to all observables investigated are negligible. We therefore set $f = f_W$ and $f_Z = f_h = 0$ in the analysis of Secs. III and IV.

### III. W RECONSTRUCTION IN DIJET EVENTS

To select events of interest, we group all visible final state particles with $p_T > 0.5$ GeV and $|\eta| < 5.0$ into cells of size $\Delta \eta \times \Delta \phi = 0.1 \times 0.1$ to account for the granularity of the detector. We identify an isolated electron or muon with $p_T > 25$ GeV and $|\eta| < 2.5$ if the hadronic energy deposit within a cone of radius $R = 0.2$ is less than 10% of the lepton candidate’s transverse energy. After removing the isolated leptons from the calorimeter cells we use the remaining visible final state of the above selection to construct jets. For triggering on jets we use the anti-$k_T$ algorithm [??]. We then apply a pre-selection trigger requiring one fat jet with $R = 1.5$ and $p_T > 200$ GeV. At this point we consider two cases: if there are no isolated leptons we perform a hadronic $W$ reconstruction and if there is

\[^{\dagger}\text{A solution for this problem may follow some or all of the ideas presented in [61, 62].}\]
exactly one lepton we perform a leptonic W reconstruction detailed below. In both analyses additional conditions on the fat jets are applied, requiring at least two such jets with $p_{Tj} > 500, 750$ or $1000 \text{ GeV}$. These additional cuts will force the radiated Ws successively further into the collinear region where the approximations of Sec. II are valid. Finally, we infer the missing transverse momentum vector from the sum of all transverse momenta of visible final state particles with $|\eta| < 5.0$. We cluster jets using FASTJET \cite{fastjet} and analyze the hadronic final state using RIVET \cite{rivet}.

A. Hadronic Analysis

To reconstruct a hadronically decaying W in dijet events, we study methods which use either the mass of the W boson or the distribution of the radiation emitted off the W decay products, so-called jet shape observables. In general we find that subjet-based reconstruction methods perform better than jet shape observables and are more efficient in separating the W decay products from other hadronic activity in the event. However, applying a jet shape observable in combination with a subjet-based mass reconstruction technique gives the best performance.

We aim to reconstruct the W boson in the three different kinematic regimes $p_{T_W} \gg m_W$, $p_{T_W} > m_W$ and $p_{T_W} \simeq m_W$. The first two methods employ jet sub-structure techniques aimed at boosted objects, and the last is an event-wide W search:

(A) To select a boosted W boson, we recluster the fat jet constituents with a Cambridge/Aachen jet algorithm (C/A) \cite{C/A} with $R=0.5$ and require $p_{Tj} > 200 \text{ GeV}$ for every subjet $\iota$. Then we apply the BDRS algorithm \cite{BDRS} to the second hardest of these subjets. We accept the W candidate if the filtered mass $m_{W_{BDRS}} \in [74, 90] \text{ GeV}$.

(B) Just like in method A we recluster the fat jet constituents into C/A subjets but now using $R=0.3$ and $p_{Tj} > 20 \text{ GeV}$. We call this set of subjets microjets. Since the quark generated in the hard process carries the bulk of the energy, we discard the hardest microjet. In case a W boson is emitted off a quark, the W decay products are likely to have larger transverse momentum than gluon emissions \cite{gluon}. Therefore, if the decay is symmetric and boosted enough, most of the time the second and third microjets are due to the W decay products. We then choose the combined four-momentum of the second and third microjets as a W candidate and their invariant mass $m_{23}$ as a discriminating variable. We find that the mass window which leaves the best signal to background ratio, while taking into account detector resolution effects, is $m_{23} \in [70, 86] \text{ GeV}$. If $m_{23}$ is within this mass window, we accept the W candidate.

(C) Since the W boson is likely to be emitted at a relatively large angle with respect to the initial high-$p_T$ quark, we recluster the full event into small anti-$k_T$ jets of radius $R=0.4$ with $p_{Tj} > 40 \text{ GeV}$. We require at least five jets to accept the event. The hardest two jets in the event originate from partons produced in the hard interaction, so we discard them. We use the remaining jets in the event to form pairs, which combine to an invariant mass $m_{kl}^2 = (p_{jk} + p_{j})^2$. We apply additional restrictions on the jet pairing to avoid biasing pure QCD events. At LHC energies even QCD radiation can occur at high virtuality, producing enough mass to match the W. If we use all jets within the event, the probability to find a pair with invariant mass within 10 GeV around $m_W$ in a generic event and miss-tag a W boson is non-negligible. To avoid this bias, only jets $j_k$, $k \in [3, 6]$, participate in the pairing algorithm. Additionally, we do not include $m_{34}$ because $j_3$ or $j_4$ is likely to be induced by QCD radiation. This leaves $m_{34}$ and $m_{4l}$ as W mass candidates, where $l \in [5, 6]$ depending on the event multiplicity.

We tag a W boson if the candidate pair mass $m_{kl} \in [70, 86] \text{ GeV}$ cut. If several combinations of jet pairings are within this window, we choose the pair with smallest $\Delta m = |m_{kl} - m_W|_{\text{min}}$ and label the pair mass $m_{\text{min}}$.

Method A is most sensitive if the W boson is highly boosted. Because the angular separation of the double decay products is $\Delta R_{ab} \simeq 2m_W/p_{T_W}$ a subjet radius of $R=0.5$ implies $p_{T_W} \gtrsim 300 \text{ GeV}$. In Fig. 4 we show the results of method A for three different fat jet $p_T$ selections. For the two free parameters ($\mu, y_{\text{cut}}$) of the BDRS method we follow the suggestion of \cite{BDRS} using (0.54, 0.13) for $200 < p_{Tj} < 500$ and (0.72, 0.09) for $500 < p_{Tj}$. The plots show an excess around $m_{W_{BDRS}} = 80 \text{ GeV}$. Higher multiplicative factors $f$ increase the EW radiation rate, resulting in larger excesses. Moreover, the W reconstruction is more successful as the fat jets have larger transverse momenta. This is not surprising because more energetic quarks emit boosted W bosons more frequently. This allows the BDRS method to tag them more efficiently.

Fig. 2 shows the W candidate mass distribution for different emission rate hypotheses using method B. We observe the same tendency here as with method A. Higher quark energy results in more frequent emissions of W bosons and their decay products form microjets two and three within the fat jet. Note the different scales and lowest points on the y-axes of Fig. 1 and 2. Although the W mass peak seem more pronounced in Fig. 2 closer inspection reveals that $S/B$ is similar as for method 1.

Fig. 5 shows the mass distribution resulting from method C. In contrast to methods 1 and 2 the signal to background ratio within the excess region does not change drastically with an increasing fat jet $p_T$ selection cut.
We find that combining methods \textbf{A} and \textbf{B} with jet shape observables, i.e. \textit{n}subjettiness $\tau_n$ \cite{7} and ellipticity $\hat{t}$ (Appendix \textbf{A}), can improve on the $W$ boson identification.

We measure these observables using the constituents of the successfully reconstructed $W$ with methods \textbf{A} and \textbf{B}. Ellipticity and $\tau_{21} = \hat{t}_2 / \hat{t}_1$ achieve the best results when applied on the second hardest boosted subjet of radius $R = 0.5$ and mass $m_{\text{BDRS}} \in [74, 90]$ GeV. In Fig. 4 we show the two distributions $\hat{t}$ (top row) and $\tau_{21}$ (bottom row).

We find that the total cross section substantially increases with $f$. This reflects the fact that we only use subjets that pass selection method \textbf{A}. Just as importantly, the shape of the distributions also changes as $f$ is varied. The peak region of the distribution of both jet shapes shifts to smaller values.

We construct ellipticity in such a way that, if the bulk of the jet radiation in the transverse plane is along a single line, the value of the jet shape observable is small. In contrast, a more circular distribution of radiation results in a large $\hat{t}$ peak region of the distribution of both jet shapes shifts to smaller values.

However, if the radiation forms two well separated clusters, other.

FIG. 1: $W$ candidate mass distribution using method \textbf{A} for $p_T \tau_J > 500$ (left), 750 (center) and 1000 (right) GeV.

FIG. 2: $W$ candidate mass distribution based on microjets $\tau_2$ and $\tau_3$ as described in method \textbf{B} for $p_T \tau_J > 500$ (left), 750 (center) and 1000 (right) GeV.

We assume at this stage that the event has already passed the tagging criteria of a single isolated lepton with transverse momentum $p_T > 25$ GeV and $|\eta| < 2.5$. A leptonically decaying $W$ gives rise to a substantial amount of missing transverse energy. We therefore require $\not{E}_T > 50$ GeV.

To reconstruct the leptonic $W$ we define its transverse mass as

$$m_T = \sqrt{2E_T \not{E}_T (1 - \cos \theta)},$$

where $\theta$ is the angle between the missing energy vector and the isolated lepton. Fig. 5 shows a pronounced peak for $m_T$ as defined in Eq. 4 in the mass window $[60, 100]$ GeV. The $W$ candidate is accepted if the transverse mass of

B. Leptonic Analysis

FIG. 2

FIG. 3
the missing momentum and lepton momentum system falls in the aforementioned bin. See Table III for the cross section of accepted events. We see there is virtually no tagged events when \( f = 0 \). Therefore, this method provides a perfect QCD rejection. The sharpness of the leptonic \( W \) peak slightly broadens as the fat jet selection cut becomes more restrictive. The lepton isolation criteria might restrict lepton tagging as the \( W \) is more likely to be emitted at a small angle from the parent quark, thereby getting more radiation within the \( R = 0.2 \) isolation radius, described in the beginning of this section. A more flexible mini-isolation criterion \[67\] might help recover sensitivity in the highly boosted regime.

IV. MEASURING W BOSON EMISSION RATES

We display the cross sections for a \( W \) boson emitted in light dijet production at various stages of the analysis in Tabs. I, II and III. The first column always refers to the modification factor \( f = f_W \) of the Sudakov factor as defined in Eq. (2). The first table, Tab. I, shows the cross sections after the different event selection criteria for the hadronic and the leptonic analyses. The first column for each analysis states the cross section after requiring exactly zero or one isolated lepton in the event, labeled \( n_l = 0 \) and \( n_l = 1 \), respectively. At this stage only the presence of one fat jet with \( p_{TJ} > 200 \) GeV is required. The three remaining columns then detail the effects of additionally applying minimum fat jet transverse momentum requirements of \( p_{TJ} > 500, 750 \) and \( 1000 \) GeV, as described in Sec. III. For \( f = 0 \), assuming pure QCD evolution, no \( W \) bosons are emitted by the jets. Then, only rarely leptons from meson or baryon decays are accepted as isolated, resulting in cross sections of \( \mathcal{O}(1) \) fb or less for the leptonic analysis.

Tab. II then further details the cross sections remaining after additionally applying mass window cuts on each mass-
The three minimum fat jet transverse momenta involved in the sample preparation and the requirements on their subjets affected during their subsequent reclustering strongly affect the cross sections left after applying methods A, B and C. As the subjets related variable for the three different methods of the hadronic analysis as described in Sec. [11]A The three minimum fat jet transverse momenta involved in the sample preparation and the requirements on their subjets strongly affect the cross sections left after applying methods A, B and C. As the subjets used in method A have $p_T > 200 $ GeV, followed by method B with $p_T > 40 $ GeV and method C with $p_T > 20 $ GeV, we observe the cross sections ascending in the same order. While the cross section of methods B and C maintain a constant difference (roughly of factor 3) as the fat jet selection cut increases, method A has a slower drop in cross section with increasing $p_T$. We attribute this behavior to the fact that emissions of subjets with $p_T > 200$ GeV and separation $\Delta R > 0.5$ are relatively rare if $p_T > 500$ GeV. For example, Tab. [11] shows that roughly 50% of all tagged W candidates with fat jet selection $p_T > 500$ GeV are actually emitted when the fat jet has $p_T > 750$ GeV. In comparison, almost 90% of all tagged W candidates in method C with selection cut $p_T > 500$ GeV stem from a fat jet with $p_T < 750$ GeV. Tab. [11] is dedicated to the leptonic analysis. It shows the cross sections after successively applying the minimum missing transverse energy cut in the left column and the transverse mass requirement in the right column, as outlined in Sec. [11]B.

The three mass reconstruction observables $m_{min}$, $m_{23}$, and $m_{BDRS}$ in the hadronic analysis of Sec. [11]A do not suffer from low statistics even in the most boosted kinematic region where both fat jets have $p_T > 1$ TeV, although the cross sections after cuts are in the sub-picobarn range. This is well within the Run 2 capabilities of the LHC, expected to reach an integrated luminosity of $\mathcal{L} \approx 100$ fb$^{-1}$. Therefore, the sensitivity in measuring electroweak emissions in the collinear approximation is mostly limited by the signal vs background ratio achieved with the reconstruction methods and their intrinsic experimental uncertainties. The observables we use induce a peak structure which allows to perform side-band analyses, reducing sensitivity on theoretical uncertainties.

In order to estimate how well the electroweak emissions can be measured at the LHC, we use a binned log-likelihood hypothesis test when varying the emission-probability modification factor $f$, as documented in [65] and implemented in the ROOT [66] package TLIMIT. Throughout we include a flat systematic error added linearly to the statistical error for each bin, $\sigma_{syst} + \sigma_{stat} = \sigma_{tot}$, when performing the likelihood test. In general the statistical error is dominant at low integrated luminosity. However, as the number of events increases, the relative statistical error decreases and the fixed statistical error can become dominant, i.e. increasing the integrated luminosity will not improve the exclusion limit. We evaluate the LHC’s potential to exclude modified W emission rates ($f \neq 1$) in favor of the Standard Model ($f = 1$).

First, we test whether we can observe hadronically decaying W bosons over the large QCD background. Here we use $1 - \text{CL}_{95}$ confidence level (as defined in [65] with $f = 0$ as background hypothesis. The results in Fig. 6 indicate what systematic error $\sigma_{syst}$ on each mass distribution would allow the exclusion of the QCD-only ($f = 0$) from the combined electroweak and QCD parton shower ($f = 1$) at 95% and 99.9% CL. Only here we set $f = 1$ to be our $S+B$ hypothesis. We plot $1 - \text{CL}_{95}$ as a function of the integrated luminosity in Fig. 6. In each row we show the background

TABLE I: Cross sections of the hadronic and leptonic analyses in pb. Where applicable a column has three numbers to account for different fat jet $p_T$ cuts: $p_T > 500$ (left), 750 (middle) and 1000 (right) GeV.

| $f$ | $n_l = 0$ | $n_l = 1$ |
|-----|-----------|-----------|
| $p_T$ | $m_T$ [GeV] | $d\sigma/dm_T$ [pb/$3$ GeV] |
| $p_T$ | $m_T$ [GeV] | $d\sigma/dm_T$ [pb/$3$ GeV] |
| $p_T$ | $m_T$ [GeV] | $d\sigma/dm_T$ [pb/$3$ GeV] |
| 0.0 | 2116 | 551.2 | 59.53 | 10.24 | 0.001 | 0.002 | 0.0002 | $3 \times 10^{-5}$ |
| 1.0 | 2092 | 539.1 | 57.74 | 9.856 | 23.37 | 3.663 | 0.5795 | 0.1286 |
| 1.1 | 2090 | 537.9 | 57.57 | 9.826 | 25.73 | 4.056 | 0.6341 | 0.1389 |
| 2.0 | 2070 | 527.5 | 56.00 | 9.481 | 45.71 | 7.081 | 1.117 | 0.2439 |

FIG. 5: Transverse mass $m_T$ for $p_T > 500$ (left), 750 (center) and 1000 (right).
rejection by one of the hadronic mass reconstruction methods from Sec. III A. From left to right the fat jet selection cut becomes more stringent. Despite the large statistics for $p_{T,J} > 500$ GeV, the variables result in a better $S/B$ and a better sensitivity in excluding $f = 1$ in the regimes were the fat jet is harder. We find that all mass reconstruction methods exclude the QCD-only hypothesis in favor of the Standard Model at 95% CL with $\sigma_{\text{syst}} = 5\%$, while using the BDTRS method allows for a 99.9% exclusion with $\sigma_{\text{syst}} = 5\%$. Moreover, $m_{\text{BDTRS}}$ gives a far more successful than 99.9% exclusion (almost 5$\sigma$). Hence, all methods designed for hadronic $W$ reconstruction will suffice to disprove a QCD-only hypothesis during the second LHC run.

Second, to evaluate how well one can measure different $W$-emission rates, we now use the CL$_{S}$ method and take the Standard Model value $f = 1$ as background $B$, aiming to exclude $S + B$ hypotheses with $f > 1$. In Figs. 7 we show the signal confidence levels CL$_{S}$ as a function of the integrated luminosity. In an ideal scenario, without systematic uncertainty, data from the LHC will provide sufficient statistics in all methods to exclude $f = 1$ at 99.9% CL. However, more realistically, after adding a systematic error $\sigma_{\text{syst}} = 1\%$ $f = 1$ cannot be excluded anymore with methods A, C alone. This is not surprising because the $W$ emission rate is only increased by 10% with respect to the Standard Model rate, resulting in $S/B \sim O(1\%)$.

Apparently a more feasible target is $f = 2$ for which we show results in Fig. 8. We see that $W$ mass reconstruction methods B and C allow the exclusion at 95% CL with $\sigma_{\text{syst}} \gtrsim 2.5\%$ and 2% respectively. Again, the reconstruction involving mass drop and filtering fares better. Using $m_{\text{BDTRS}}$, it is possible to achieve $S/B \approx 20\%$ with the $p_{T,J} > 1$ TeV selection. This allows to exclude the signal to 95% with 5% systematic uncertainty and to 99.9% with 3.5% systematic uncertainty.

Consequently, to be able to exclude $f = 1.1$ we need to increase $S/B$ by improving on the $W$ reconstruction. The combination of the jet shapes discussed in Sec. III A, i.e. $t$ and $\tau_{21}$, with method A allows to boost the sensitivity of the likelihood ratio. Fig. 8 shows the signal confidence level using $t$ (left) and $\tau_{21}$ (right) in combination with method A. A fat jet transverse momentum cut $p_{T,J} > 750$ yields the best result for two reasons: the $W$s are boosted enough for the jet shapes to perform well and the cross section is large enough to accumulate sufficient statistics during Run 2 of the LHC. The combination of method A with either of the jet shapes can exclude the signal to 95% CL with 2.5% systematic uncertainty, while $\sigma_{\text{syst}} = 1\%$ would allow to exclude it in favor of the Standard Model to 3$\sigma$. What allows the discrimination is the strong rejection of QCD emissions with low jet shape values. Still, this additional step is not enough to exclude $f = 1.1$ at 95% CL given a 5% systematic uncertainty.

Therefore, in order to test small deviations from the Standard Model emission rate, a $W$ reconstruction with small statistical and systematic uncertainties is needed. The leptonic analysis outlined in Sec. III B can be of use. Even though the cross sections for $f = 1$ and $f = 1.1$ differ only by a few femtobarns, i.e. for $p_{T,J} > 1$ TeV, the high luminosity or run 2 will provide sufficient statistics. A clear advantage of the leptonic reconstruction is the improved $S/B \approx 10\%$ compared to $S/B \approx 1\%$ for the hadronic mass reconstruction. This quantitative difference is due to

| $f$ | $p_{T,J}$ | $p_{T,J}$ | $p_{T,J}$ |
|-----|-----------|-----------|-----------|
|     | 500 GeV   | 750 GeV   | 1000 GeV  |
| 0   | 0.9939    | 0.4906    | 0.1447    |
| 1.0 | 1.219     | 0.6202    | 0.1923    |
| 1.1 | 1.251     | 0.6386    | 0.1977    |
| 2.0 | 1.422     | 0.7312    | 0.2286    |

TABLE II: Cross sections after the three mass reconstruction cuts in the three different methods for the hadronic analysis in pb. Each column contains three numbers to account for different fat jet cuts: $p_{T,J} > 500$ (left), 750 (middle) and 1000 (right) GeV.

| $f$ | $\mathcal{E}_{T} > 50$ GeV | $m_{T} \in [60,100]$ GeV |
|-----|-----------------------------|---------------------------|
|     | $p_{T,J}$ | $p_{T,J}$ | $p_{T,J}$ | $p_{T,J}$ | $p_{T,J}$ |
| 0   | 0.001 | $1 \times 10^{-5}$ | $4 \times 10^{-7}$ | $6 \times 10^{-5}$ | $5 \times 10^{-6}$ | $1 \times 10^{-7}$ |
| 1.0 | 2.062 | 0.3481 | 0.07988 | 0.5769 | 0.09271 | 0.02156 |
| 1.1 | 2.280 | 0.3795 | 0.08654 | 0.6402 | 0.1046 | 0.02323 |
| 2.0 | 4.000 | 0.6765 | 0.1531 | 1.108 | 0.1830 | 0.04099 |

TABLE III: Cross sections after the $\mathcal{E}_{T} > 50$ GeV cut and the $m_{T}$ cut in the leptonic analysis in pb. Each column contains three numbers to account for different fat jet cuts: $p_{T,J} > 500$ (left), 750 (middle) and 1000 (right) GeV.
We attribute it to the fact that here the S/B ratio does not improve with more boosted jets and to the reduced isolated tagging efficiency explained in Sec. III B. Even though the hadronic decay comes with a larger cross section, the improved background rejection in the leptonic analysis allows for a better discrimination of non-Standard Model emission rates.

V. CONCLUSION

We propose leptonic and hadronic W-tagging strategies for bosons produced in association with a hard jet. Hadronic W reconstruction relies on jet substructure techniques and a subsequent use of jet shapes to separate W jets from the overwhelming QCD background. We compare the Standard Model with increased W emission rates. Using a binned log-likelihood approach while varying "systematic" uncertainties we find that electroweak emission rates can be measured at the LHC.

Hadronic W mass reconstruction methods are capable of distinguishing differences in the cross section comparable
FIG. 7: CLs obtained from the W mass reconstruction through method A using $m_{\text{BDRS}}$ (top row), method B using $m_{23}$ (center row), and method C using $m_{\text{min}}$ (bottom row) of the hadronic analysis for the three different minimum jet transverse momenta: $p_{TJ} > 500$ (left column), 750 (center column) and 1000 (right column) GeV. The background corresponds to the Standard Model emission rate ($f = 1$) and signal + background to $f = 2$.

with the SM value, i.e. $|\sigma_{f, \text{SM}} - \sigma_{\text{SM}}| / \sigma_{\text{SM}} \approx 1$. In other words, hadronic mass reconstructions can exclude the $f = 0$ and $f = 2$ hypotheses in favor of the SM ($f = 1$) with a 5% systematic uncertainty up to 95% CL. The best subjet-based reconstruction method we tested, $m_{\text{BDRS}}$, can exclude $f = 0$ to 99.9% with $\sigma_{\text{syst}} = 5\%$ and $f = 2$ with $\sigma_{\text{syst}} = 3.5\%$.

However, if $|\sigma_{f, \text{SM}} - \sigma_{\text{SM}}| / \sigma_{\text{SM}} \approx 0.1$, i.e. $f = 1.1$, no hadronic mass reconstruction can discriminate signal from background given a non-negligible systematic uncertainty. To this end, we calculate jet shapes of the subjets which pass the mass cuts given in Sec. III. The jet shape distributions we find most useful are ellipticity $t$ and $\tau_2$ applied on boosted subjets with $m_{\text{BDRS}} \in [74, 90]$ GeV. Binned log-likelihood analyses of these distributions allow a 95% CL exclusion of the $f = 1.1$ hypothesis with $\sigma_{\text{syst}} = 2.5\%$. With a smaller error of 1.5% we can exclude $f = 1.1$ with 99.9% CL. The statistical error at this stage amounts to $O(1)\%$. This suggests the possibility that in the high luminosity run, we would be able to exclude the $f = 1.1$ hypothesis with an even larger systematic uncertainty, or if the systematics are well understood, probe smaller values of $|\sigma_{f, \text{SM}} - \sigma_{\text{SM}}| / \sigma_{\text{SM}}$. While this seems a futile exercise – we essentially know the coupling of $W$ bosons to quarks – this kind of measurement will offer an opportunity for the determination of electroweak Sudakov effects.

The simple leptonic analysis rejects the background more successfully. It benefits from the large cross section of the low jet $p_T$ cut. Given a systematic uncertainty $\sigma_{\text{syst}} \leq 2.5\%$, a cut on the transverse mass $m_T$ of a leptonic $W$ candidate will exclude $f = 1.1$ up to 99.9%. Moreover, it can probe $|\sigma_{f, \text{SM}} - \sigma_{\text{SM}}| / \sigma_{\text{SM}} < 0.1$. 
FIG. 8: CLs obtained from the ellipticity $\hat{t}$ (left) and $\tau_{21}$ (right) distributions calculated from the constituents of the W candidates that pass the BDRS cut on the second boosted subjet. $p_{TJ} > 750$ GeV. The background is the SM emission rate ($f = 1$), signal + background sample is $f = 1.1$.

FIG. 9: CLs obtained from the W transverse mass $m_T$ reconstruction in the leptonic analysis. The background sample is the SM emission rate ($f = 1$). The signal plus background sample is $f = 1.1$.

Acknowledgments

MS acknowledges support by the Research Executive Agency (REA) of the European Union under the Grant Agreement number PITN-GA-2010-264564 (LHCPenoNet) and PITN-GA-2012-315877 (MCnet). MS further gratefully acknowledges fruitful discussion with A. Denner, kindly supplying the work with F. Hebenstreit, A. Huss for supplying the results of [26] to cross-check our implementation. FK and MS are grateful for useful discussions with S. Höche and other members of the SHERPA team. PP would like to acknowledge the support provided by the Science and Technology Facilities Council (STFC).

Appendix A: Ellipticity

To calculate the ellipticity $\hat{t}$ of a jet we define the particles’ three-momentum components $k_{Tj}$ transverse to the jet it is part of. Thus, it is defined in the plane transverse to the momentum $p_J = \sum_i p_i$, where $p_i$ are the three-momenta of the jet constituents, as

$$k_{Ti} = p_i - (p_J \cdot p_i) \frac{p_J}{|p_J|}.$$

While we take $p_J$ to be the thrust axis, we calculate thrust major $T_{maj}$ and and thrust minor $T_{min}$ using the $k_{Ti}$ as input

$$T_{maj} = \max_{n_{maj}} \frac{\sum_i |k_{Ti} \cdot n_{maj}|}{\sum_i |p_{Ti}|} \quad \text{and} \quad T_{min} = \frac{\sum_i |k_{Ti} \cdot n_{min}|}{\sum_i |p_{Ti}|},$$

where $n_{maj}^2 = n_{min}^2 = 1$, $n_{min} \cdot n_{maj} = 0$ and $n_{min} \cdot p_J = 0$. We then define the ellipticity to be the ratio

$$\hat{t} = \frac{T_{min}}{T_{maj}}.$$
The thereby defined ellipticity makes a distinction between different jet topologies. The two limiting cases are either when the radiation of the jet is distributed homogeneously within the jet cone (the energy profile in the jet transverse plane is a circle), leading to $T_{\text{maj}} = T_{\text{min}}$ and the ellipticity $\hat{e} = 1$, or when the radiation is two-dimensional (the energy profile in the jet transverse plane is one-dimensional), leading to $T_{\text{min}} = 0$ and $T_{\text{maj}} > 0$ and the ellipticity $\hat{e} = 0$.

---

[1] G. Aad et al. (ATLAS Collaboration), Phys.Lett. B716, 1 (2012), 1207.7214.
[2] S. Chatrchyan et al. (CMS Collaboration), Phys.Lett. B716, 30 (2012), 1207.7235.
[3] J. M. Butterworth et al., A. R. Davison, M. Rubin, and G. P. Salam, Phys.Rev.Lett. 100, 242001 (2008), 0802.2470.
[4] D. Krohn, J. Thaler, and L.-T. Wang, JHEP 1002, 084 (2010), 0912.1342.
[5] S. D. Ellis, C. K. Vermilion, and J. R. Walsh, Phys.Rev. D81, 094023 (2010), 0912.0033.
[6] J.-H. Kim, Phys.Rev. D83, 011502 (2011), 1011.1493.
[7] J. Thaler and K. Van Tilburg, JHEP 1103, 015 (2011), 1011.2265.
[8] D. E. Soper and M. Spannowsky, Phys.Rev. D84, 074002 (2011), 1102.3480.
[9] S. D. Ellis, A. Hornig, T. S. Roy, D. Krohn, and M. D. Schwartz, Phys.Rev.Lett. 108, 182003 (2012), 1201.1914.
[10] A. Denner and S. Pozzorini, Eur.Phys.J. C18, 461 (2001), hep-ph/0101201.
[11] E. Accomando, A. Denner, and S. Pozzorini, Phys.Rev. D65, 073003 (2002), hep-ph/0110114.
[12] W. Beenakker and A. Werthenbach, Nucl.Phys. B630, 3 (2002), hep-ph/0112030.
[13] M. Denner, M. Melles, and S. Pozzorini, Nucl.Phys. B662, 299 (2003), hep-ph/0301241.
[14] M. Denner, S. Pozzorini, Nucl.Phys. B717, 48 (2005), hep-ph/0408068.
[15] A. Denner, B. Jantzen, and S. Pozzorini, JHEP 0811, 062 (2008), 0809.0800.
[16] A. Denner, S. Rijmen, and J. Rittinger, Eur.Phys.J. C70, 659 (2010), 1004.4117.
[17] S. Dittmaier, A. Huss, and C. Speckenh, JHEP 1211, 095 (2012), 1210.0438.
[18] J. Kühn, A. Scharf, and P. Uwer (2013), 1305.5773.
[19] A. Bierweiler, T. Kasprzak, and J. H. Kühn, JHEP 1312, 071 (2013), 1305.5402.
[20] M. L. Mangano, M. Moretti, F. Piccinini, R. Pittau, and A. D. Polosa, JHEP 07, 001 (2003), hep-ph/0206293.
[21] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer, JHEP 1106, 128 (2011), 1106.0522.
[22] F. Krauss, R. Kuhn, and G. Soff, JHEP 02, 044 (2002), hep-ph/0109036.
[23] T. Gleisberg and S. Höche, JHEP 0812, 039 (2008), 0808.3674.
[24] W. Kilian, T. Ohl, and J. Reuter, Eur.Phys.J. C71, 1742 (2011), 0708.4233.
[25] L. Dai, K. Melnikov, and F. Caola, JHEP 1204, 095 (2012), 1210.1523.
[26] J. R. Christiansen and T. Sjöstrand (2014), 1401.5238.
[27] T. Sjöstrand, S. Mrenna, and P. Z. Skands, Comput.Phys.Commun. 178, 852 (2008), 0710.3820.
[28] T. Gleisberg, S. Höche, F. Krauss, M. Schönherr, S. Schumann, et al., JHEP 0902, 007 (2009), 0811.4622.
[29] Z. Nagy and D. E. Soper, JHEP 0510, 024 (2005), hep-ph/0503053.
[30] S. Schumann and F. Krauss, JHEP 0803, 038 (2008), 0709.1027.
[31] J.-C. Winter, F. Krauss, and G. Soff, Eur.Phys.J. C36, 381 (2004), hep-ph/0311085.
[32] S. Alekhin, G. Altarelli, N. Ampanfa, J. Andersen, V. Andreev, et al. (2005), hep-ph/0601012.
[33] M. Schönherr and F. Krauss, JHEP 0812, 018 (2008), 0810.5071.
[34] H.-L. Lai, M. Guzzi, J. Huston, Z. Li, P. M. Nadolsky, et al., Phys.Rev. D82, 074024 (2010), 1007.2241.
[35] M. H. Seymour, Z.Phys. C56, 161 (1992).
[36] T. Sjöstrand, S. Mrenna, and P. Z. Skands, Phys.Rev. D75, 030002 (2007), 0610.1897.
[37] M. Bahr et al., Eur.Phys.J. C58, 639 (2008), 0803.0883.
[38] S. Höche, S. Schumann, and F. Siegert, Phys.Rev. D81, 034026 (2010), 0912.3501.
[39] A. Denner and F. Hehenstreit, unpublished (2006).
[40] S. Catani, F. Krauss, R. Kuhn, and B. Webber, JHEP 0111, 063 (2001), hep-ph/0109231.
[41] S. Höche, F. Krauss, S. Schumann, and F. Siegert, JHEP 0905, 053 (2009), 0903.1219.
[42] S. Lounblad, JHEP 0205, 046 (2002), hep-ph/0111228.
[43] M. L. Mangano, M. Moretti, F. Piccinini, and M. Treccani, JHEP 0701, 013 (2007), hep-ph/0611129.
[44] J. Alwall et al., Eur.Phys.J. C53, 473 (2008), 0706.2569.
[45] K. Hamilton, P. Richardson, and J. Tully, JHEP 0911, 038 (2009), 0905.3072.
[55] L. Lönnblad and S. Prestel, JHEP **1203**, 019 (2012), 1109.4829.
[56] L. Lönnblad and S. Prestel, JHEP **1302**, 094 (2013), 1211.4827.
[57] S. Plätzer, JHEP **1308**, 114 (2013), 1211.5467.
[58] T. Carli, T. Gehrmann, and S. Höche, Eur.Phys.J. **C67**, 73 (2010), 0912.3715.
[59] S. Catani and M. Seymour, Nucl.Phys. **B485**, 291 (1997), hep-ph/9605323.
[60] S. Catani, S. Dittmaier, M. H. Seymour, and Z. Trocsanyi, Nucl.Phys. **B627**, 189 (2002), hep-ph/0201036.
[61] Z. Nagy and D. E. Soper, JHEP **0807**, 025 (2008), 0805.0216.
[62] S. Hoeche, F. Krauss, M. Schonherr, and F. Siegert, JHEP **1209**, 049 (2012), 1111.1220.
[63] M. Cacciari, G. P. Salam, and G. Soyez, JHEP **0804**, 063 (2008), 0802.1189.
[64] M. Cacciari, G. P. Salam, and G. Soyez, Eur.Phys.J. **C72**, 1896 (2012), 1111.6097.
[65] A. Buckley, J. Butterworth, L. Lönnblad, D. Grellscheid, H. Hoeth, et al., Comput.Phys.Commun. **184**, 2803 (2013), 1003.0694.
[66] Y. L. Dokshitzer, G. Leder, S. Moretti, and B. Webber, JHEP **9708**, 001 (1997), hep-ph/9707323.
[67] K. Rehermann and B. Tweedie, JHEP **1103**, 059 (2011), 1007.2221.
[68] T. Junk, Nucl.Instrum.Meth. **A434**, 435 (1999), hep-ex/9902006.
[69] R. Brun and F. Rademakers, Nucl.Instrum.Meth. **A389**, 81 (1997).