Jet Substructure and the Search for Neutral Spin-One Resonances in Electroweak Boson Channels

Andrey Katz,[1] Minho Son,[2] and Brock Tweedie[3,4]

[1]*Department of Physics, University of Maryland, College Park, MD 20742*
[2]*Department of Physics, Yale University, New Haven, CT 06511*
[3]*Department of Physics and Astronomy, Johns Hopkins University, Baltimore, MD 21218*
[4]*Physics Department, Boston University, Boston, MA 02215*

Strongly coupled models at the TeV scale often predict one or more neutral spin-one resonances (Z') which have appreciable branching fractions to electroweak bosons, namely the Higgs and longitudinal W and Z. These resonances are usually believed to have multi-TeV mass due to electroweak precision constraints, placing them on the edge of LHC discovery reach. Searching for them is made particularly challenging because hadronically decaying electroweak bosons produced at such high energy will appear very similar to QCD jets. In this work we revisit the possibility of discovering these resonances at the LHC, taking advantage of recently developed jet substructure techniques. We make a systematic investigation of substructure performance for the identification of highly Lorentz-boosted electroweak bosons, which should also be applicable to more general new physics searches. We then estimate the model-independent Z' discovery reach for the most promising final-state channels, and find significant improvements compared to previous analyses. For modes involving the Higgs, we focus on a light Higgs decaying to $b\bar{b}$. We further highlight several other novelties of these searches. In the case that vertex-based b-tagging becomes inefficient at high p_T, we explore the utility of a muon-based b-tag, or no b-tag at all. We also introduce the mode $Z' \to Zh \to (\nu\bar{\nu})(b\bar{b})$ as a competitive discovery channel.

I. INTRODUCTION

In the next few years, expectations are high that the LHC will discover the Higgs boson, finally verifying all of the basic building blocks of the Standard Model. There is also much anticipation that this discovery will not signal the end of novel physics near the electroweak scale. However, given the lack of clear anomalies in existing flavor and electroweak precision measurements, a very real possibility is that the Standard Model with a light elementary Higgs will turn out to be a reasonably good description of LHC phenomenology at energies of less than about a TeV. Should this prove to be the case, after a few 10's of fb^{-1} of data attention will become highly focused on carefully testing the properties of the Higgs through a combination of coupling measurements [1] and investigations of W, Z, and h production at progressively higher energies [2, 3]. One of the most important goals is to determine to what extent the Higgs is truly an elementary particle, versus a light state arising from a new composite sector.

Like most scenarios for physics beyond the Standard Model, specific composite Higgs models are under a certain amount of tension between, on the one hand, satisfying the flavor and electroweak precision constraints and, on the other hand, reintroducing the fine-tuning which the models were invented to solve, though still at a much reduced level (see for reviews e.g. [4, 5]). The usual compromise is to choose a compositeness scale of several TeV. If this kind of scenario proves to be true, then one of the main hints that there is something unusual about the Higgs will be the production of multi-TeV resonances with large couplings to Higgs bosons.

In this paper, we aim to make a detailed investigation of the LHC's sensitivity to such scenarios via the production of a neutral spin-1 resonance in $q\bar{q}$ annihilation. The existence of one or more Z' particles with these quantum numbers is rather ubiquitous in composite Higgs models, for example serving as the analogs of the ρ mesons of QCD or as higher states of an enlarged broken gauge sector. Owing to the large coupling with the Higgs, as well as with the (longitudinal) W and Z bosons, the decays of these Z' resonances will typically be dominated by the modes Zh and W^+W^-. This is in stark contrast to more standard $U(1)'$ models, which usually predict much larger relative rates into leptons and quarks. Observation of a Z' with large branching fraction to either of these electroweak boson modes would therefore serve as highly suggestive evidence for a composite Higgs

sector, whereas lack of an observation would place important constraints.

The two intermediate states Zh and W^+W^- can lead to a variety of final-state configurations. Despite the additional structure compared to, say, a simple dileptonic final state, these can be some of the most challenging search channels. Indeed, the flavor and electroweak data hint at the most difficult possible situation: the Higgs may be light and the Z' may be heavy. The former presents a difficulty since the Higgs decays then become dominated by $b\bar{b}$, versus the more distinctive $WW^{(*)}/ZZ^{(*)}$ modes of heavier Higgses. The latter suggests that Z' production rates will be small (e.g., $O(1 \text{ fb})$ for $m_{Z'} = 3$ TeV [6, 7]), and also that the secondary decay products of the highly energetic electroweak bosons will be Lorentz-boosted into tight, jet-like configurations in the detector. For leptonic decays of the W and Z bosons, this last observation does not necessarily pose a problem, beyond the usual loss of efficiency due to leptonic branching fraction. But for any boson decays involving hadrons, the distinction with a QCD-induced jet can become uncomfortably blurred.

In the past several years, various techniques have been developed for dealing with highly boosted Ws, Zs, Higgses, and tops (e.g., [3, 8–20]. For our present study with boosted electroweak bosons, we will utilize the procedure of [8]. Originally, this was developed to aid in the identification of semi-boosted light Higgses in the high-p_T tail of Standard Model Higgsstrahlung production. The procedure adapts on an event-by-event basis to pick out hard substructures within a region of interest in the detector, while actively attempting to remove uncorrelated radiation which can degrade mass resolution. We will find that the ideas of [8] easily extrapolate into the truly high-boost regime for the Higgs, as well as to hadronic decays of boosted Ws and Zs. Along the way, we will find that the scale-invariant nature of the procedure offers a number of nontrivial advantages.

Beyond the simple $b\bar{b}$ decays of light Higgses, as well as $q\bar{q}$ decays of electroweak gauge bosons, substructure-type techniques can also be applied to the other possible decay modes of a boosted Higgs, such as $\tau^+\tau^-$ and $WW^{(*)}/ZZ^{(*)}$. Including these modes in a $Z' \to Zh$ search offers several advantages. First, a discovery in multiple modes would bolster confidence that we are seeing a real signal. Second, broadening the set of Higgs decay modes also broadens the sensitivity over Higgs masses, and allows for increased sensitivity in cases with nonstandard Higgs couplings (e.g., gaugephobic [21–23] or fermiophobic [24]). Third, a measurement of Higgs branching fraction ratios gives us further clues regarding its identity (especially whether it is the same "Higgs" as the one discovered at lower energies)

and, in the Standard Model limit, can give us a secondary handle on the mass. In this paper, we focus on the specific case of a 120 GeV Higgs decaying to $b\bar{b}$. We separately explore the decays $h \to \tau^+\tau^-$ and $h \to WW^{(*)}/ZZ^{(*)}$ in two companion papers [25, 26]. (See also [27] for an analysis of $Z' \to Zh \to Z(ZZ)$ in the all-leptonic channel.)

Given this restriction, we find that the most promising Z' discovery channels are

- $WW \to (l\nu)(q\bar{q}')$

- $Zh \to (l^+l^-)(b\bar{b})$

- $Zh \to (\nu\bar{\nu})(b\bar{b})$

The first two of these have been explored before in [6, 28, 29]. (The first channel was also studied in [3, 30] in the context of weak boson fusion. We also point out a study in [31], closely related to our second channel, of $h \to ZZ \to (l^+l^-)(q\bar{q})$ at $m_h = 300\text{-}600$ GeV using substructure methods.) By applying jet substructure techniques, we obtain significant improvements, for example achieving sensitivity to fb-scale cross sections for $m_{Z'} = 3$ TeV in W^+W^- with only 100 fb^{-1} of data from a 14 TeV LHC. The last mode, utilizing the invisible Z decay in Zh, appears to have been largely underappreciated for resonance searches, but we find that it is quite competitive with the leptonic Z. We further explore the utility of b-tagging the Higgs-jet in the two Zh modes, using a muon-based technique that should extrapolate robustly to high-p_T. Application of such a tag appears to be beneficial for finding Z' bosons in roughly the 1-2 TeV mass range for the leptonic Z mode, and up to about 3 TeV for the invisible Z. Beyond these masses, the background becomes so small that an untagged analysis is more sensitive.

The paper is organized as follows. In section II we take a more detailed look at the theoretical motivation for our study and classify the relevant Z' final states. In section III we review jet substructure techniques, characterize their performance for tagging boosted electroweak bosons, and make a comparison with a more traditional jet-based approach. We present our model-independent discovery reach analyses in section IV. Section V contains conclusions. We include discussions of pileup and detector modeling in appendices A and B, respectively.

II. Z' THEORY AND FINAL-STATE CHANNELS

A. Theoretical motivation

For our purposes, the relevant couplings of a lone generic Z' can be parametrized as[1]

$$\mathcal{L}_{\text{int}} = Z'_{\mu} j'^{\mu} \quad \text{with} \quad j'^{\mu} = \sum_i g'_i \psi_i^{\dagger} \bar{\sigma}^{\mu} \psi_i + g'_H \left(H^{\dagger} i \partial^{\mu} H + h.c. \right). \tag{1}$$

Here, ψ_i stands for the chiral Standard Model (SM) fermions (in complete $SU(2)_L$ multiplets) and H for the Higgs doublet (neutral Higgs will be denoted by h). g'_i and g'_H are the respective coupling constants. The coupling g_H controls the decay rate of the Z' into Zh and W^+W^- by virtue of the Goldstone boson equivalence theorem. (We exclusively work in the limit $m_{Z'} \gg m_Z, m_W$.) Immediate consequences of this are that the rates into Zh and W^+W^- are nearly equal, and the Zs and Ws produced in the decays are highly longitudinally polarized. More general models (as we will see below) can have more than one Z' which can mix with each other and with the Z, leading to $SU(2)_L$-nonuniversalities in the couplings of the individual Z' mass eigenstates. However, the approximate equality of total Zh and W^+W^- rates is usually maintained, as electroweak symmetry breaking (EWSB) is necessarily a small effect in the heavy Z' sector.

Roughly speaking, Z' models can be categorized into two classes. In the first class, the Z' is realized as a gauge boson of a new gauge group broken near the TeV scale. (See [32] for a review.) In the second class, the Z' emerges from a new strongly-coupled sector which also generates the Higgs boson or directly breaks electroweak symmetry. One simple example of the former class is an admixture of $U(1)_Y$ and $U(1)_{B-L}$. Another typical example is an admixture of the $U(1)_{\chi}$ and $U(1)_{\psi}$ subgroups of E_6.[2] The class of composite Z' includes various models of Technicolor [33–41], composite Higgs [42–46], partial compositeness [47], and warped 5D models [48–51]. The Little Higgs [4, 52] effectively lives in between these two classes.

---

[1] We neglect four-point couplings with the SM Higgs, as well as terms induced by replacing $\partial^{\mu} \to D^{\mu}$. (These are both important for a more complete, gauge-invariant treatment of Z' couplings to the SM gauge bosons.) The possibility of kinetic mixing with hypercharge is already incorporated in our parametrization. The field redefinitions necessary to undo such a mixing will simply shift $j'^{\mu} \to j'^{\mu} + \kappa\, j_Y^{\mu}$, where κ is a (usually small) number proportional to the coefficient of the mixing term.

[2] The first example is automatically anomaly free without extra exotic fermions other than right-handed neutrinos. The second example requires new exotic fermions near the TeV scale.

The simple gauge Z' couples fairly democratically between fermions and Higgs, usually leading to much higher branching fractions into fermions simply due to multiplicity (as well as a factor of two enhancement for decays into fermions versus decays into scalars). In fact, when the lepton Yukawa couplings have SM-like minimal structure, and respect the $U(1)'$ symmetry, we can make a sharper statement:

$$BR(Z' \rightarrow l^+l^-) \geq BR(Z' \rightarrow Zh \text{ or } W^+W^-), \tag{2}$$

where l is any individual species of lepton. If this relation is violated, we can infer that either the lepton Yukawa sector has non-minimal structure (such as multiple Higgs doublets or $U(1)'$-breaking spurions) or the couplings of the Z' do not furnish a gauge symmetry. Note that for a hypercharge-sequential Z', the branching fraction to each species of lepton is 12%, whereas the branching fraction to Zh or W^+W^- is 1.2%. For a $B - L$ boson, the branching fraction to each lepton is 15%, and into electroweak bosons it is zero. It is important to realize that even if a resonance is first discovered in the l^+l^- modes (which are very straightforward experimentally), concurrent searches in Zh and W^+W^- can yield very valuable information about the nature of the resonance, even in the case that nothing is found.

The situation drastically changes if the both the Z' and the Higgs doublet are composite particles generated by a new strongly-coupled sector. In this case, the coupling to the Higgs doublet can be hierarchically larger than the couplings to the fermions, and the decays of the Z' become dominated by the modes Zh and W^+W^- (and perhaps $t\bar{t}$, as we will see).[3] On the other hand, the typically small couplings between the composite sector and the fundamental SM fermions can significantly reduce the production cross section via $q\bar{q}$ annihilation. Indeed, weak boson fusion is often taken as the most robust way to produce composite resonances, but a very high price in effective luminosity (see, e.g., [3]). Nonetheless, realistic and well-motivated models have been identified where the competition between small production rate via $q\bar{q}$ and large decay rate into electroweak bosons can roughly compensate. A very

---

[3] It is possible in such composite scenarios that the Higgs boson proper, the neutral CP-even scalar component of the doublet, is not a well-defined particle. This is the case in Technicolor and other Higgsless theories. The decays of the Z' may then become completely dominated by W^+W^- (in close analogy with $\rho^0 \rightarrow \pi^+\pi^-$ in QCD), or by other light composites. As we ultimately perform a model-independent analysis in this paper, in particular an independent study of W^+W^-, our results can also be applied to these cases if $m_{Z'} \gtrsim 1$ TeV.

well-studied class of such models are the warped 5D constructions inspired by the Randall-Sundrum (RS) model [48], with the SM fields propagating in the bulk [49, 50]. According to AdS/CFT duality [53], these can be interpreted as weakly-coupled holographic duals of large-N_c 4D field theories at very strong coupling and which exhibit near-conformality above the TeV scale [54, 55].

As an illustrative example, we briefly review the minimal custodial RS model with bulk SM fields [51]. (A more detailed review of the Z' phenomenology of this model appears in [6].) Implementing a gauged custodial symmetry in the bulk (the holographic dual of a global custodial symmetry of the strongly-coupled sector) prevents large shifts of the electroweak T parameter, and can be modified to better accommodate measurements of $Z \to b\bar{b}$ [56]. The complete bulk gauge symmetry of the electroweak sector is $SU(2)_L \times SU(2)_R \times U(1)_{B-L}$, and boundary conditions are chosen so that only the $SU(2)_L \times U(1)_Y$ subgroup has zero modes (which subsequently acquire mass through the usual Higgs mechanism). The higher KK excitations sample from the complete gauge group, with each KK level containing two W' bosons and three Z' bosons. All of the bosons have nearly degenerate masses set by the 5D geometry, and can be highly mixed after EWSB.

In this scenario, the couplings of the Z' bosons to the SM fermions and the Higgs doublet have a natural geometrical interpretation. The couplings are determined by the overlaps among the wavefunctions along the fifth dimension. In order to solve the hierarchy problem as in the original RS, the Higgs profile is localized near (or confined to) the IR brane. Light fermions are localized near the UV brane, leading to small overlap with the Higgs, and therefore small 4D Yukawa couplings. Heavy fermions such as the top quark are localized toward the IR brane. The gauge zero modes are uniformly spread throughout the space, but their KK excitations are localized near the IR brane. Consequently, the couplings of Z' to light fermions are suppressed, whereas the couplings to the Higgs doublet (and top quarks) are enhanced. As the region near the UV brane represents non-composite physics in the dual description, whereas the bulk and the IR brane represent composite physics, the tail of the KK wavefunctions near the UV brane indicates the kind of non-composite/composite mixing that occurs between the photon and ρ^0 in QCD. This mixing can be significant enough to allow reasonable production rates in $q\bar{q}$ annihilation.

However, mixing between the composite sector and light SM sector also leads to tension, as there will be associated shifts in the values of electroweak and flavor observables. In order

to satisfy electroweak constraints, coming mainly from the S parameter, the mass of the first gauge KK level should be at least $O(3 \text{ TeV})$ [51, 57, 58]. Incorporating constraints from flavor and CP can push the mass even higher [59]. These observations tend to generalize beyond this specific model, especially regarding electroweak corrections. The implication is that if the Higgs is composite, then the physics associated with that compositeness may be hiding at the multi-TeV scale. For this particular model, a 2(3) TeV Z' can be produced in $q\bar{q}$ annihilation with $\sigma \times BR(Zh \text{ or } W^+W^-)$ of roughly 15(1.5) fb at a 14 TeV LHC [6]. (This is, in fact, the dominant production mechanism.)

While we have so far discussed two extreme cases, namely simple gauge $U(1)'$ models and models where the Higgs and Z' are both composite particles, it is also possible to interpolate between the behavior of these two. A well-known example is the Little Higgs scenario [52] (reviewed in [4]). Here, the Higgs is a composite pseudo-Goldstone boson arising from the breakdown of a partially-gauged global symmetry, and inherits special protection from one-loop mass corrections. The gauge sector contains two copies of $SU(2) \times U(1)$ as a subgroup, of which our own electroweak group is a mixture. The orthogonal gauge group contains two Z' bosons. Unlike the Higgs, they are now elementary, but they becomes massive due to the dynamics of the strongly-coupled sector that generates the Higgs. The relative couplings of these Z' bosons with the Higgs and with the SM fermions depends on the mixing angles between the two copies of the gauge group, and can be continuously dialed.[4] Similar to the warped 5D models, electroweak constraints favor TeV-scale masses or higher for the Z' bosons of the Little Higgs (see, e.g., [60, 61]). To get a sense for what cross sections are possible, we cite the numbers for the "Littlest Higgs" [62] (with $\cot\theta = 0.5$) studied in [28]: 10(1.1) fb for $m_{Z'} = 2(3)$ TeV at a 14 TeV LHC.

In summary, we emphasize that many well-motivated models contain one or more Z' with appreciable coupling to the Higgs doublet, and therefore appreciable branching fraction into electroweak bosons. Such Z' are constrained by electroweak precision tests (as well as flavor measurements) to live in the multi-TeV mass range. Through the analog of γ-ρ^0 mixing, it is possible to produce these resonances at the LHC through $q\bar{q}$ annihilation, even if the Z' is a composite or belongs (mostly) to a gauge group that is decoupled from the light SM

---

[4] While the Z' bosons here represent simple broken gauge symmetries, the branching fractions into fermions and Zh/W^+W^- can escape the constraint of equation (2) since the Yukawa couplings in these models contain spurion fields to balance the charges.

to satisfy electroweak constraints, coming mainly from the S parameter, the mass of the first gauge KK level should be at least $O(3 \text{ TeV})$ [51, 57, 58]. Incorporating constraints from flavor and CP can push the mass even higher [59]. These observations tend to generalize beyond this specific model, especially regarding electroweak corrections. The implication is that if the Higgs is composite, then the physics associated with that compositeness may be hiding at the multi-TeV scale. For this particular model, a 2(3) TeV Z' can be produced in $q\bar{q}$ annihilation with $\sigma \times BR(Zh \text{ or } W^+W^-)$ of roughly 15(1.5) fb at a 14 TeV LHC [6]. (This is, in fact, the dominant production mechanism.)

While we have so far discussed two extreme cases, namely simple gauge $U(1)'$ models and models where the Higgs and Z' are both composite particles, it is also possible to interpolate between the behavior of these two. A well-known example is the Little Higgs scenario [52] (reviewed in [4]). Here, the Higgs is a composite pseudo-Goldstone boson arising from the breakdown of a partially-gauged global symmetry, and inherits special protection from one-loop mass corrections. The gauge sector contains two copies of $SU(2) \times U(1)$ as a subgroup, of which our own electroweak group is a mixture. The orthogonal gauge group contains two Z' bosons. Unlike the Higgs, they are now elementary, but they becomes massive due to the dynamics of the strongly-coupled sector that generates the Higgs. The relative couplings of these Z' bosons with the Higgs and with the SM fermions depends on the mixing angles between the two copies of the gauge group, and can be continuously dialed.[4] Similar to the warped 5D models, electroweak constraints favor TeV-scale masses or higher for the Z' bosons of the Little Higgs (see, e.g., [60, 61]). To get a sense for what cross sections are possible, we cite the numbers for the "Littlest Higgs" [62] (with $\cot\theta = 0.5$) studied in [28]: 10(1.1) fb for $m_{Z'} = 2(3)$ TeV at a 14 TeV LHC.

In summary, we emphasize that many well-motivated models contain one or more Z' with appreciable coupling to the Higgs doublet, and therefore appreciable branching fraction into electroweak bosons. Such Z' are constrained by electroweak precision tests (as well as flavor measurements) to live in the multi-TeV mass range. Through the analog of γ-ρ^0 mixing, it is possible to produce these resonances at the LHC through $q\bar{q}$ annihilation, even if the Z' is a composite or belongs (mostly) to a gauge group that is decoupled from the light SM

---

[4] While the Z' bosons here represent simple broken gauge symmetries, the branching fractions into fermions and Zh/W^+W^- can escape the constraint of equation (2) since the Yukawa couplings in these models contain spurion fields to balance the charges.

fermions. In the case that the Z' is a simple elementary gauge boson, we usually expect the branching fractions into electroweak bosons to be smaller than the branching fractions into fermions. However, even if this turns out to be the case, verifying that the rate into electroweak bosons is either absent or subleading would be very useful to test the gauge-like nature of the Z'.

B. Possible final-state decay channels

While we have been discussing only two options for the initial decay stage of the Z' (roughly equal rates into Zh and W^+W^-), the secondary decays of the electroweak bosons lead to a proliferation of possible final states. Matters are particularly complicated by the fact that we do not yet know the mass of the Higgs boson, nor even if its branching fractions for a given mass strictly follow those predicted by the SM.

We can start by categorizing the final states from the well-understood W^+W^- case. The available topologies (identical those well-known from $t\bar{t}$) are dileptonic, l+jets, and all-hadronic. (Dileptonic and l+jets include cases with taus, though we do not explicitly investigate these here.) The dileptonic channel (5% BR) is in principle the cleanest, but suffers from low branching fraction and ambiguities in kinematic reconstruction. It was analyzed in [6] and found to be less sensitive than l+jets. Given this, and given that the current paper is focused on jet substructure applications, we do not re-analyze this channel. The l+jets channel (30% BR) was also studied in [6] (as well as in [29]), and was shown to be the most powerful. We investigate the search reach in this channel in subsection IV A. Finally, the all-hadronic channel (45% BR) has high rate but faces overwhelming dijet background. Even applying two hadronic W-tags to the event (using the substructure techniques of the next section) is not enough to combat this to the point where all-hadronic achieves a reasonable S/B ratio.[5] Therefore, we do not investigate it.

The Zh case offers significantly more options. We can globally classify these using the h

---

[5] Since the l+jets and all-hadronic branching fractions are comparable, all-hadronic can be competitive only if the single-tagged W+jet $\rightarrow$ $(l\nu)$+jet and double-tagged dijet background rates are also comparable. Equivalently, the ratio between the untagged background cross sections (W+jet/dijet) should be of order the W mistag rate or larger. We estimate this ratio of cross sections to be only 0.03% in our kinematic range of interest, more than two orders of magnitude too small, given the mistag rates we estimate in subsection III C.

decay mode, and more locally by also specifying the Z decay mode.

The simplest option, and the one favored by a light SM-like Higgs (70% BR for $m_h = 120$ GeV), is $h \to b\bar{b}$. This two-body decay was explored in the highly boosted case in [6, 28], and in the intermediate-boost case in [8]. The $b\bar{b}$ system will be accompanied by the decaying Z, which can decay to l^+l^- ($l = e, \mu$, 7% BR), neutrinos (20% BR), $\tau^+\tau^-$ (3% BR), or jets (70% BR). The usual logic is to focus on the cleanest possible final-state: $(l^+l^-)(b\bar{b})$, usually with one or two b-tags. We will analyze this mode anew in subsection IV B, applying jet substructure and replacing vertex-based b-tagging, which performs uncertainly at high-p_T, with μ-based tagging or no tagging at all.[6] The invisible Z should be similarly clean, but has not been studied in detail before at high-p_T. (It is mentioned in [6], and studied explicitly at intermediate-boost in [8].) While complete kinematic reconstruction is impossible, the Z' mass can be approximated event-by-event by using $2p_T(h)$. This has a Jacobian peak at $m_{Z'}$, which is further reinforced by spin terms in the decay matrix element. While the resonance peak here is inevitably broader than in the fully reconstructed leptonic case, and is therefore subject to higher background, this interplays nontrivially with the three times higher statistics. We study this mode in subsection IV C. Taus we will not consider, as they are challenging and have very small rate. (However, we return to them in the context of Higgs decays below.) Finally, the case with a jetty Z faces the same kind of overwhelming background problem as all-hadronic W^+W^-, and is at best extremely challenging.[7] We do not pursue it here.

This rounds out the set of modes which we explore in this paper: $(l\nu)(q\bar{q}')$, $(l^+l^-)(b\bar{b})$, $(\nu\bar{\nu})(b\bar{b})$. We will pursue two other important options for Higgs decays, $h \to \tau^+\tau^-$ and $h \to WW^{(*)}/ZZ^{(*)}$, in a set of companion papers [25, 26].

III.   JET SUBSTRUCTURE FOR BOOSTED ELECTROWEAK BOSONS

The electroweak bosons produced in the decay of a multi-TeV Z' will be highly Lorentz-boosted. Consequently, their own decay products tend to become highly collimated into

---

[6] While we do not explicitly exploit this, our no-tagging analysis will also capture the signal from $h \to gg$.

[7] We might ask whether it could be recovered by demanding two hadronic electroweak boson tags and two b-tags on the Higgs side. If the b-tag operates perfectly (100% tag, 0% mistag), this may not be so terrible. However, it seems unlikely that this mode can be made competitive with more realistic b-tag numbers.

single "jets" — W-jets, Z-jets, and h-jets. On the one hand, this collimation can complicate searches, as these boson-jets can easily be confused with QCD-induced jets. E.g., the mode $Z' \to WW \to (l\nu)(q\bar{q}')$ will suffer from large, quasi-two-body W+jets backgrounds Wq and Wg. Even modes with leptons within a jet (such as from boosted $h \to \tau^+\tau^- \to lX$), will face backgrounds from ordinary jets containing bottom/charm production, lepton fakes, or W/Z-strahlung. On the other hand, the internal kinematic configurations of boson-jets are still rather distinctive from QCD, and we can capitalize on this to improve discrimination. Indeed, a variety of strategies have been developed to discriminate hadronic (or semi-hadronic) decays of boosted heavy particles from ordinary QCD jets utilizing jet substructure, i.e. the detailed distribution of particles and/or energy within a jet. In this section, we detail in what sense substructure techniques are beneficial for discriminating boosted $W/Z/h$ decaying to two quarks, and how robust these techniques are to contamination and detector effects. (We discuss new techniques which we develop for the even more distinctive case of boosted $h \to \tau^+\tau^-$ in [25], and boosted $h \to WW^{(*)}/ZZ^{(*)}$ in [26].)

The section is organized as follows. We first describe the basic utility of substructure methods and describe our nominal choice of algorithm. We then make a comparison with a more traditional jet-based analysis to further clarify the advantages of substructure. Finally, we estimate the performance of boosted electroweak boson tagging, and discuss the robustness of our estimates against theory and detector modeling.

A.   Utility of substructure and the BDRS algorithm

The simplest jet observable sensitive to substructure is the jet's mass. While the decay products of the boosted electroweak boson may end up within a single jet, this jet will still have a mass close to the electroweak boson mass, and the Z' resonance will show up as a localized bump in the joint distribution of event mass and jet mass. For backgrounds with QCD jets, the event mass and the jet mass will both tend to be peaked toward zero.

Taking an even closer look at the energy distribution within a jet offers two immediate improvements. First, ordinary QCD jets with mass near, say, m_W, will preferentially acquire their mass via relatively soft emissions near the edge of the jet area (depending in detail on the jet definition and p_T scale). This situation is in contrast to genuine electroweak bosons, which usually consist of two well-localized subjets with roughly democratic energy-

sharing. Substructure techniques [3, 8, 12–15] can identify the more QCD-like kinematic configurations and remove them from analysis, thereby improving signal-to-background. Second, by identifying the subjets corresponding to the two decay partons, we can restrict our boson reconstruction to include only the particles (tracks and calorimeter cells) near their cores, thereby ignoring wider-angle radiation in the jet. This tends to be dominated by uncorrelated sources, namely ISR, underlying event, and pileup. By refining the set of particles to those in the immediate vicinity of the subjets, the jet mass resolution can be optimized.[8]

To take advantage of both of these improvements, we focus here on the subjet-finding technique developed by Butterworth, Davison, Rubin, and Salam (BDRS) [8] and further investigated by ATLAS [64]. Our modes $Zh \to (l^+l^-)(b\bar{b})$ and $Zh \to (\nu\bar{\nu})(b\bar{b})$ were also studied in those papers (as well as the closely related $Wh \to (l\nu)(b\bar{b})$), at lower boosts and in the context of a Standard Model Higgsstrahlung search. The BDRS technique trivially extrapolates to boosted hadronic Ws and Zs, as well as to arbitrarily high energies, limited only by the detector's spatial resolution (a topic which we address in appendix B).

The algorithm runs as follows. An event is first iteratively clustered into large-radius "fat-jets" with the Cambridge/Aachen (C/A) jet algorithm [65, 66]. After a fat-jet of interest is identified, it is iteratively *declustered* by working backwards through the clustering stages. At a given stage, a protojet j is split into progenitors j_1 and j_2. The splitting is categorized by a fractional mass-drop measure $\max(m_1, m_2)/m_{12}$ and a symmetry measure $\min(p_{T1}^2, p_{T2}^2)\Delta R_{12}^2/m_{12}^2$. The algorithm stops, identifying two subjets, when the former is less than a threshold $\mu = 0.67$, and the latter is greater than a threshold $y_{\text{cut}} = 0.09$. If these criteria are not met, the more massive of j_1 and j_2 is used as the j for the next iteration, and the less massive is thrown away. The algorithm continues searching down to arbitrarily small ΔRs and arbitrarily small protojet masses, until either suitable substructure is found or there are no more particles left to operate on.

While we will be applying this algorithm in a new kinematic regime, we do not modify its original declustering parameters. Partially this is to maintain contact with the earlier literature, but more practically we do not find dramatic gains in signal/background discrim-

---

[8] A number of promising techniques have also been developed to specifically clean a jet of uncorrelated radiation, independent of the identification of localized subjets [12, 13, 63].

ination by changing parameters. Indeed, the fairly scale-free nature of the procedure leads to similar behaviors at all p_T scales. We do note that after completing our studies with the original BDRS algorithm, we found that the mass-drop criterion appears to play almost no role. For all of our analysis samples, it is almost always strictly weaker than the symmetry criterion.[9] Of course, a more realistic study should more closely explore optimization of the procedure and its parameters.

For our clustering radius, we use $R = 1.4$, corresponding to quasi-hemispheric fat-jets. This may seem unnecessarily, even counterproductively, large for a kinematic region where the interesting structure becomes smaller than a normal jet. However, we have found that this very large jet radius offers a number of advantages. First, choosing a jet radius based on the known ratio m_W/p_T or the expected ratio m_h/p_T would introduce a hidden dimensionful scale into the analysis, manifesting itself as a turn-off feature beyond m_W or m_h in the background's jet-mass spectrum. Choosing a large R pushes this feature away from the physical boson mass, flattening out the sidebands. Second, using a large R broadens the range of Z' masses over which our search could operate efficiently, down to $m_{Z'} = 300 \sim 400$ GeV. Third, as discussed in subsection III C, using such a large fat-jet radius picks up on more global event information, which appears to be useful for discriminating electroweak boson jets from QCD jets.

After finding subjets, BDRS further refine them using a technique called filtering. The set of particles constituting the two subjets found by the mass-drop/asymmetry procedure are reclustered with C/A, using a refined jet-radius parameter $R_{\text{filt}} = \min(0.3, \Delta R_{\text{subjets}}/2)$. The hardest three of these refined subjets is kept and used to reconstruct the boson. Filtering has been shown to be important to achieve optimal mass resolution at moderate boost, where the initial two subjets can be quite large and quite contaminated by the underlying event and pileup (see [67] for a detailed study). As we transition to high boost, these subjets become smaller, and the effects of contamination quickly become much less dramatic. For simplicity, we do not use filtering in the present study, nor do we introduce pileup into

---

[9] Analyses run with the mass-drop criterion turned off yield kinematic distributions (both shape and normalization) for signal and background subjets that are nearly identical to analyses run with the mass-drop turned on. However, running with the mass-drop but without the symmetry criterion leads to significant reshaping of background, and in particular enhances the number of background events in the electroweak boson mass windows by a factor of $2 \sim 3$.

our simulation samples. We separately address the utility of filtering for removing pileup (mainly relevant for the 1 TeV samples) in appendix A.

Before proceeding, we also point out the possible utility of jet "superstructure" [68] for discriminating boosted electroweak bosons from QCD jets. The shower products of a boosted electroweak boson tend to be very well-collimated, at the ΔR scale of the two daughter quarks, as there is no color connection with the rest of the event. QCD jets, on the other hand, can shower at all angles, often with bias towards one of the beams. This can lead to differences in the large-angle energy patterns. While simple measures of this difference lead to only modest discrimination [68, 69], it is possible that a more sophisticated analysis could yield a more powerful discriminator [70]. We will see in subsection III C that our choice of quasi-hemispheric jet clustering already incorporates some aspect of this more global discrimination.

B. Comparison with traditional jet analysis

We attempt to further clarify the utility of the BDRS procedure in heavy Z' searches by performing a simple particle-level comparison with a more traditional jet-based analysis. Analyses of this latter type have previously been investigated in [6, 28, 29]. We use for this comparison events from our $Z' \rightarrow WW \rightarrow (l\nu)(q\bar{q}')$ signal and combined $W(q/g)$ background.[10] Our BDRS-style analysis operates on the hardest $R = 1.4$ fat-jet found in each event. The traditional jet analysis clusters the event into $R = 0.4$ anti-k_T jets. To make the two analyses as similar as possible, in the traditional analysis we first find the hardest jet and then seek out the next-hardest jet within $R = 1.4$. If the pairing is not too asymmetric, in the sense of the BDRS y_{cut}, these two jets are kept as the "subjets" constituting the hadronic W candidate. (As in the substructure analysis, the mass-drop criterion is almost always satisfied.) Failing this, it is assumed that the W decay products have merged into one jet, and only the initial hardest jet is kept. In both analyses, we only consider relatively

---

[10] Details of the simulations and reconstruction can be found in section IV, though there are some important differences in the present analysis. No detector model is applied, and particle energies are not smeared. (Including these effects would not significantly alter our conclusions.) Also, unlike the full WW analysis in section IV, a jet veto is not applied here, as the sample of jets in the two analyses are very different, and we do not consider top backgrounds.
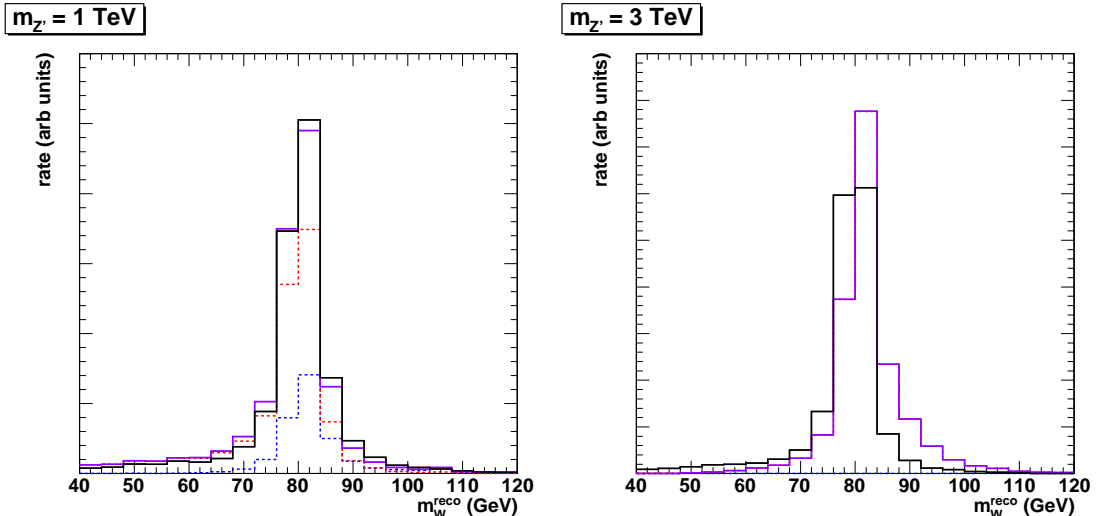
FIG. 1: *Distributions of the (particle-level) reconstructed hadronic W mass for 1 and 3 TeV $Z' \to WW \to (l\nu)(q\bar{q}')$. Displayed are the nominal BDRS procedure (solid black), traditional dijets (dashed blue), traditional monojets (dashed red), and combined dijet+monojet (solid purple).*

central hadronic W candidates, with $|\eta| < 1.5$.

The hadronic W mass reconstruction is shown in Fig. 1 for 1 and 3 TeV Z'. It is clear that the impact of switching to a traditional analysis is modest. Three features are worth noting. First, at 1 TeV we see a healthy admixture of traditional dijet and monojet reconstructions, whereas at 3 TeV monojet (unsurprisingly) dominates. Second, the traditional analysis is somewhat more efficient (by about 1.5% at 1 TeV and 7% at 3 TeV), due to the lack of internal phase space cuts in the monojet mode. Third, the 3 TeV monojet sample shows some broadening due to underlying event contamination, since the separation between W-subjets is becoming small compared to the jet radius.

The backgrounds are shown in Fig. 2, expressed as slices of the double-differential cross section in the $(m_W^{\mathrm{reco}}, m_{Z'}^{\mathrm{reco}})$ plane at $m_{Z'}^{\mathrm{reco}} = 1$ and 3 TeV. Again, 1 TeV events reconstructed near the W mass are a mixture of traditional dijet and monojet, and each has a very distinct shape. In the W mass region, the sum is nonetheless quite comparable to the BDRS result. At 3 TeV, the traditional analysis is again dominated by monojet. However, the effects of the internal kinematic cuts of the BDRS procedure are now evident, as the background processed through BDRS is roughly two times smaller, and significantly flatter.

While the advantage of BDRS over more traditional jet analysis is very clear in the 3 TeV

14
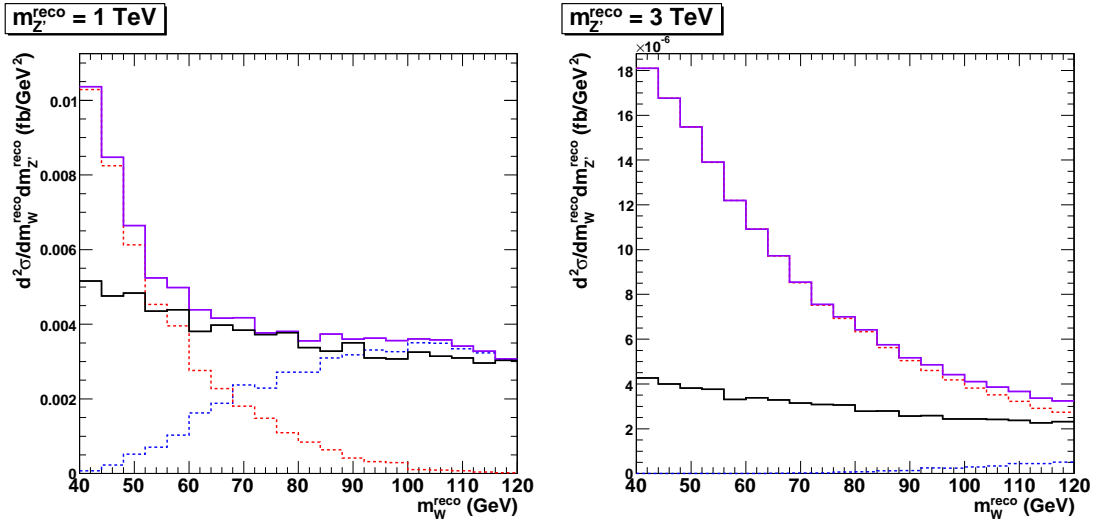
FIG. 2: *Slices of the showered leading-order double-differential cross section for W+jets $\rightarrow$ $(l\nu)$+jets vs (particle-level) reconstructed W mass, fixing the reconstructed Z' mass at 1 and 3 TeV. Displayed are the nominal BDRS procedure (solid black), traditional dijets (dashed blue), traditional monojets (dashed red), and combined dijet+monojet (solid purple).*

background, the 1 TeV background illustrates a more subtle potential advantage. In the traditional analysis, we inevitably face the issue of combining dijet and monojet analyses. There is a broad crossover region between dijet-dominance and monojet-dominance as we scan through potential Z' masses.[11] In contrast, the BDRS procedure is largely free of any distance or energy scales beyond the fat-jet clustering size. It operates smoothly and (mostly) without dimensionful thresholds. This feature may be advantageous, both in simplifying the analysis and in avoiding possible systematic effects on the background shape in the $(m_W^{\rm reco}, m_{Z'}^{\rm reco})$ plane. The relative flatness over $m_W^{\rm reco}$ may also be seen as an advantage. The background processed by BDRS is only steeply falling along the $m_{Z'}^{\rm reco}$ direction, possibly allowing a signal bump to stand out more clearly in the mass plane than in the more

---

[11] The detailed behavior of this crossover depends on the jet definition and cuts. For example, when we focus on the region $m_{Z'}^{\rm reco} = 1$ TeV, the reconstructed Ws will have $p_T \sim 500$ GeV. In our dijet analysis, the minimum ΔR is 0.4, and the minimum y is $y_{\rm cut} = 0.09$. The minimum dijet mass can be roughly approximated as $\sqrt{0.09}(0.4)(500\text{ GeV}) = 60$ GeV, which corresponds reasonably well with the left panel of Fig. 2. Note that changing (or removing) the symmetry cut will simply move the location of the crossover. Taking a larger R for the jet is another option. However this subjects the reconstructed Ws to more contamination, and is less discriminating against highly asymmetric QCD configurations.
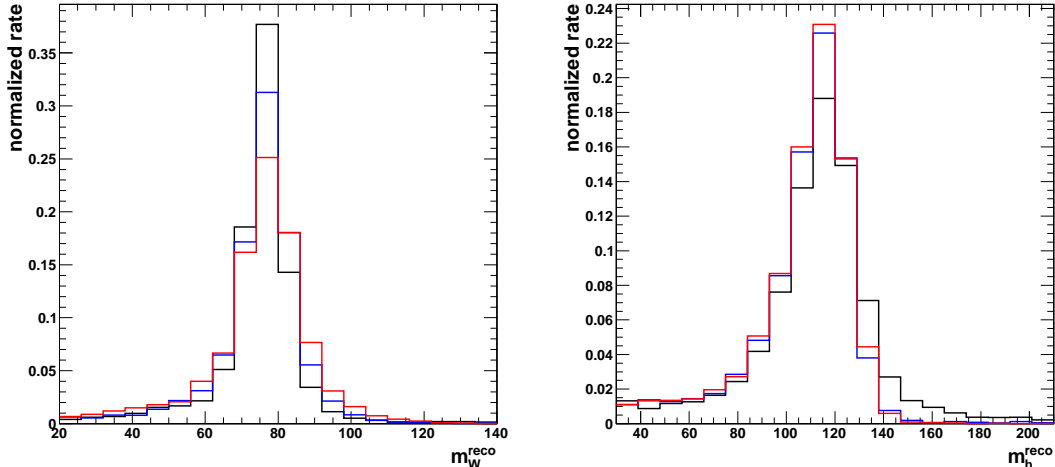
FIG. 3: *Reconstructed W-jet (left) and h-jet (right) invariant mass distributions for Z′ masses of 1 TeV (black), 2 TeV (blue), and 3 TeV (red). The rescaled ECAL model of appendix B has been applied.*

	$p_T \simeq 500$ GeV	$p_T \simeq 1000$ GeV	$p_T \simeq 1500$ GeV
W	76%	77	72
h	59	61	62

TABLE I: *Tag rates (in percent) for boosted electroweak bosons. (Absolute statistical errors are of order 2%.)*

traditional analysis.

C. Performance and robustness

We now categorize the performance of the BDRS procedure for discriminating boosted $W/Z/h$. To make our estimates somewhat more realistic, we apply a simple detector model, which is described in detail in appendix B. This model hybridizes ECAL spatial resolution with ECAL+HCAL energy measurements, and introduces rudimentary energy smearing. The simple trick of using the ECAL to trace the energy flow significantly decreases the impact of detector granularity at very high p_T. Event reconstruction follows section IV.

For the signal, we use W-jets and (non-b-tagged) h-jets reconstructed from the $Z′ \rightarrow$

16

	$p_T \simeq 500$ GeV	$p_T \simeq 1000$ GeV	$p_T \simeq 1500$ GeV
quark $\to W$	6.5%	6.5	5.9
quark $\to h$	6.8	5.6	5.8
gluon $\to W$	10.4	8.3	7.4
gluon $\to h$	10.5	8.8	7.4

TABLE II: *Mistag rates (in percent) for quarks and gluons faking boosted electroweak bosons. (Absolute statistical errors are of order 0.3%.)*

$WW \to (l\nu)(q\bar{q}')$ and $Z' \to Zh \to (l^+l^-)(b\bar{b})$ samples, respectively. (We do not present explicit numbers for Z-jets, but these will be very similar to W-jets.) The boosted bosons produced in the Z' decays are strongly peaked toward $p_T = m_{Z'}/2$, and are usually quite central. The tag rate is defined as the fraction of reconstructed events where the summed boson subjets fall within a prespecified mass window chosen by eye: $m_{\text{subjets}} = [65, 95]$ GeV for Ws, and $m_{\text{subjets}} = [100, 140]$ GeV for Higgses.[12] The reconstructed W and Higgs mass distributions are shown in fig. 3, and tag rates are displayed in table I. Qualitatively, we can see that the rates are $O(1)$, are weak functions of p_T, and that the Higgs rates tend to be smaller. This last feature comes from the fact that the higgs is scalar, versus a longitudinal spin-1 boson, and that the mass peak is slightly degraded by neutrinos from semileptonic bottom and charm decays. It is also worth noting that the the tag rates for transverse W would be smaller than those for longitudinal, though we have not investigated this quantitatively.[13]

The mistag rates are displayed in table II, for $Z(q/g)$ backgrounds generated in $\hat{p}_T$ bins of $[400, 600]$ GeV, $[800, 1200]$ GeV, and $[1200, 1800]$ GeV. ($W(q/g)$ simulations yield identical results.) These rates are also fairly weak functions of p_T, and are similar for our W and Higgs mass windows. Quarks fake heavy bosons less often than gluons, as their probability for radiating is smaller. The final numbers are about 6-7% for quarks, and 7-11% for gluons.

---

[12] Sometimes a boosted boson will fail reconstruction because the BDRS procedure deconstructs its fat-jet down to a single calorimeter cell. We have checked that this is very rare for signal, at the few percent level at most. It can be more important for QCD fat-jets, which tend to have less substructure. The loss rate in that case is still typically less than 10%. Effectively, these represent an additional contribution to the zero-mass bin which we have not accounted for.

[13] Determining this would be important, for example, for W bosons produced in decays of spin-0 or spin-2 resonances.

Interestingly, the mistag rates show a decreasing trend as we increase p_T, particularly for gluons. This may be indicating some of the global color effects mentioned in subsection III A. In particular, our very large fat-jet size will pick up FSR emissions at wide angles, which are present for QCD jets but not for electroweak boson jets. This hardens the background jet-mass spectrum, sending some fraction of background events beyond the electroweak boson mass windows. To get a better sense for what this implies, we re-ran the W-tag analysis on the $m_{Z'} = 3$ TeV WW signal and $p_T = [1200, 1800]$ backgrounds with $R = 0.3$. This jet size is just large enough to contain the subjets from almost all of the signal events within our W mass window, given the value of y_{cut}. We find that the signal efficiency increases slightly, from 72% to 74%. This represents some recaptured events, which were lost in the quasi-hemispheric analysis due to high-mass combinations where one subject is the entire boosted W and the other is an ISR jet. (The original fraction of such "bad" combinations is at the 5% level for all Z' samples.) However, the quark and gluon mistag rates increase by a factor of roughly 1.7, to 9.4% and 13.0%, respectively. Optimization of the fat-jet and subject-finding procedures to best capitalize on this physics is an interesting topic which we defer to future work. Still, it is already clear that using smaller jets at higher p_T may, counterintuitively, be *detrimental*.

We have also investigated the theoretical robustness of these mistag rate estimates by comparing our results against samples generated with `HERWIG v6.510` [71] and `MadGraph/MadEvent+PYTHIA v4.4.44` [72] with jet-parton matching. The `HERWIG` samples are of identical composition to the default `PYTHIA` samples. They display no statistically significant difference in any of the mistag rates. The `MadGraph` sample consists of Z+jets events MLM-matched [73] up to three jets using p_T and ΔR thresholds of 30 GeV and 0.2, respectively. We have generated the $p_T = [400, 600]$ GeV bin, and find rather good agreement with the `PYTHIA` and `HERWIG` results.[14] The mistag rates from `MadGraph` do appear to be 10-20% lower, relatively speaking, but this would have to be confirmed with higher-statistics running.

Detector modeling (see appendix B) is largely unimportant for the mistag rates given

---

[14] When working at higher-order it becomes somewhat ambiguous whether a jet originates from a quark or a gluon. As an operational definition, we search out hard progenitor partons within $\Delta R = 1.4$ of the main fat-jet axis, tallying individual (signed) quark flavors. Jets with leftover quark-number are deemed "quark jets," and are otherwise deemed "gluon jets." This procedure yields relative "Zq" and "Zg" cross sections quite comparable to what is obtained with simple $2 \to 2$ matrix elements.

fixed jet-mass windows, as the QCD spectrum processed through BDRS is not sharply featured. We have explicitly checked that eliminating the calorimeter geometry model and/or the energy smearing has no significant effect. The jet-mass distributions for signal are of course much more sensitive. Practically, we can parametrize our ignorance with a single p_T-dependent factor: the amount by which we have mis-estimated the jet-mass resolution. If a more realistic estimate tells us that we must expand our mass windows by a factor x to accept the same signal, then our mistag rates will increase by approximately the same factor (since the background spectra are fairly flat). Signal significance will scale like $1/\sqrt{x}$, and signal-to-background like $1/x$. Assuming that x is an $O(1)$ number, then our estimates for Z' reach should also be good to $O(1)$.

IV. Z' SEARCHES

In this section, we estimate the discovery reach for Z' resonances decaying to W^+W^- and Zh, applying the jet substructure techniques in section III. As discussed in section II, l+jets is the most promising final state for W^+W^-, and we investigate this mode. This has previously been studied in [6, 29]. For Zh, we focus exclusively on the case of a light Higgs (120 GeV) decaying to $b\bar{b}$. The channel with associated leptonic Z decay has traditionally been considered the most promising [6, 28] (see also [31]), and we also investigate this mode. The channel with invisible Z has been mentioned in [6], but not investigated in detail. This is the final mode which we study here. Perhaps not without coincidence, these three topologies — $W^+W^- \to (l\nu)(q\bar{q}')$, $Zh \to (l^+l^-)(b\bar{b})$, and $Zh \to (\nu\bar{\nu})(b\bar{b})$ — are nearly the same as those studied in [8], though with Wh replaced by WW.[15]

The Z' samples for this study are generated with 2→4 matrix elements in `MadGraph/MadEvent v4.4.32` [72] and showered/hadronized in `PYTHIA` (plugin version `2.1.3`). The Z' width is taken to be 3%, which is smaller than instrumental width and the signal windows that we use below. (Our estimates should work reasonably well for resonances up to about 15% natural width.) Our background samples are generated with `PYTHIA v6.4.11` [74]

---

[15] We note that our results for WW can be translated with minor modifications to the case of $W' \to Wh \to (l\nu)(b\bar{b})$ without b-tags. Our non-tagged Zh results can also be translated for searches for $ZZ \to (l^+l^-)(q\bar{q})$ and/or $ZW \to (l^+l^-)(q\bar{q}')$, though some care should be taken in accounting for spin effects.

with default settings.[16] All processes are evaluated at leading-order.[17] As discussed in appendix A, we do not incorporate pileup into the analysis, but have checked that its effects would be quite manageable. The final-state particles are processed through a primitive ECAL/HCAL hybrid calorimeter model, described in detail in appendix B, to capture the main spatial resolution effects which we face in reality. The resulting calorimeter cells are then clustered/declustered using `FastJet 2.4.1` [78]. Energy smearing is applied to the final reconstructed objects (leptons and subjets). Subjet energy smearing is described in appendix B. Electrons are smeared by 2%, and muons by $(5\%)\sqrt{E/\text{TeV}}$.

We estimate discovery reach for 1, 2, and 3 TeV Z' by using simple counting. For each sample, we construct a box in the $(m_{W/h}^{\text{reco}}, m_{Z'}^{\text{reco}})$ plane, centered on the signal. The box is not optimized, other than coarsely by eye. We claim that discovery is possible if two criteria are met: $N_S/\sqrt{N_B} > 5$ and $N_S > 10$.

Several of our reconstruction criteria are common between all of the analyses. We accept leptons (e and μ) with $p_T > 30$ GeV and $|\eta| < 2.5$. The leptons are required to be isolated from surrounding hadronic activity (neglecting photons and other leptons) within a cone of $R = 0.4$: $p_T(l)/(p_T(l) + p_T(\text{hadrons})) > 0.9$. Jets are constructed, as described in section III, using the Cambridge/Aachen algorithm with $R = 1.4$. We consider jets with $p_T > 20$ GeV and $|\eta| < 2.5$. The leading jet, which we take to be the hadronic electroweak boson candidate, must have $p_T > 200$ GeV and $|\eta| < 1.5$. (The last requirement forces the candidate into the better-segmented barrel region of the detector.) This jet is declustered, and the two subjets used to reconstruct the boson.

A. W^+W^- in l+jets mode: $(l\nu)(q\bar{q}')$

Our basic selection cuts are as follows. We demand exactly one isolated lepton, which should be located in the semicylinder opposite the hadronic W candidate jet. We model missing energy by taking the transverse vector that balances the (energy-smeared) lepton

---

[16] `PYTHIA` accounts for spin effects in the electroweak boson decay distributions for these backgrounds. However, they would not be accounted for in Z' decays.

[17] The NLO QCD K-factor for the Z' signal is 1.3 [75]. The QCD K-factor for high-p_T W/Z+jets, which is our dominant background in all cases, is 1.6 [76]. The net effect of these corrections on $S/\sqrt{B}$ is therefore small. Electroweak corrections for high-p_T processes can also be significant [77]. These tend to reduce the W/Z+jets rate.
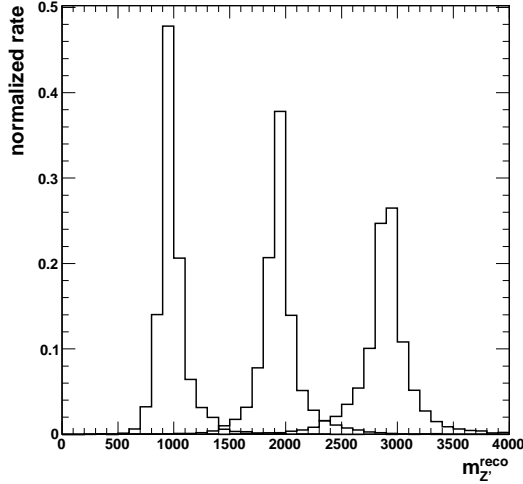
FIG. 4: *Reconstructed Z' invariant masses for 1, 2, and 3 TeV in $WW \to (l\nu)(q\bar{q}')$.*

	$m_{Z'} = 1$ TeV	$m_{Z'} = 2$ TeV	$m_{Z'} = 3$ TeV
Signal Eff.	12.3%	15.8%	15.6%
$\sigma(Wj:\ qg \to qW)$	43.8 fb	1.47 fb	0.11 fb
$\sigma(Wj:\ q\bar{q} \to gW)$	13.0 fb	0.57 fb	0.050 fb
$\sigma(t\bar{t})$	23 fb	0.57 fb	0.011 fb
$\sigma(WW)$	3.77 fb	0.33 fb	0.042 fb
$\sigma(WZ)$	1.01 fb	0.068 fb	0.0068 fb
$\sigma(Wh)$	0.029 fb	0.0017 fb	0.00013 fb

TABLE III: *Signal efficiency and background cross sections after all cuts in $WW \to (l\nu)(q\bar{q}')$. Note that the signal efficiency includes the branching fractions for the W decays.*

and two subjets. We then approximately reconstruct the complete leptonic W boson by exploiting the quasi-collinearity of the lepton and neutrino: $\eta_\nu \equiv \eta_l$. After this step, we can construct the complete event mass, $m_{Z'}^{\mathrm{reco}}$. To reduce W+jets background, which tends to be produced at shallower angles than the signal for a given $\hat{s}$, we demand that the hadronic W satisfy $p_T > m_{Z'}^{\mathrm{reco}}/3$. Finally, in order to control top backgrounds, we apply a jet veto:

Luminosity	$m_{Z'} = 1$ TeV	$m_{Z'} = 2$ TeV	$m_{Z'} = 3$ TeV
$\mathcal{L} = 30$ fb^{-1}	58.3 fb	9.02 fb	2.70 fb
$\mathcal{L} = 100$ fb^{-1}	31.9 fb	4.94 fb	1.48 fb
$\mathcal{L} = 300$ fb^{-1}	18.4 fb	2.85 fb	0.85 fb
$S/B = 1$	501.5 fb	15.44 fb	1.36 fb

TABLE IV: $\sigma(Z') \times BR(Z' \to W^+W^-)$ required for discovery or $S/B = 1$ in the $(l\nu)(q\bar{q}')$ mode.
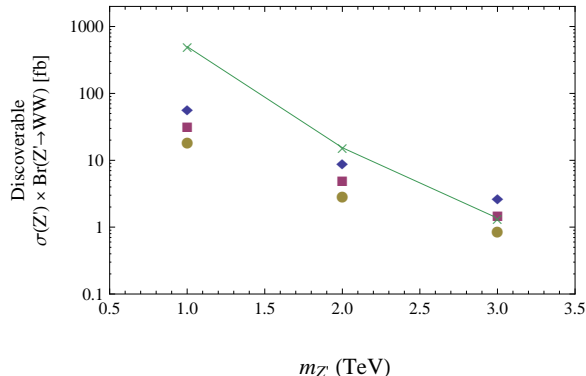
FIG. 5: *Discovery reach for $WW \to (l\nu)(q\bar{q}')$. The three symbols (diamond, box, circle) correspond to (30, 100, 300) fb^{-1} of integrated luminosity. The $\times$ refers to the value that is required to achieve $S/B = 1$.*

the subleading jet in the event (if found) should satisfy $p_T < p_T(l)/2$.[18] The reconstructed Z' resonance, after application of all cuts, is shown in figure 4.

Our signal box is $m_W^{\text{reco}} = [65, 95]$ GeV and $m_{Z'}^{\text{reco}} = [m_{Z'} - 15\%, m_{Z'} + 15\%]$. The final signal efficiencies and background cross sections are shown in table III. The discovery reach, i.e. the minimum Z' cross section times $BR(W^+W^-)$ required to claim discovery, is listed for integrated luminosities of 30, 100, and 300 fb^{-1} in table IV, and displayed graphically in fig 5. We also list/display the $\sigma \times BR$ required to achieve $S/B = 1$.

---

[18] This is a very simplistic cut, with many caveats. In particular, since the jets we use are very large, pileup can significantly increase the subleading jet's momentum. However, even a slightly more sophisticated analysis could likely work around this, for example reclustering with a smaller jet size or trimming the fat-jet [63]. We suspect that a more dedicated top "anti-tag," for example attempting to fully reconstruct and veto the leptonic and/or hadronic top, could perform even better. We also note that a jet veto of this type rejects W+jets backgrounds generated in the soft W-strahlung-like regions of phase space.
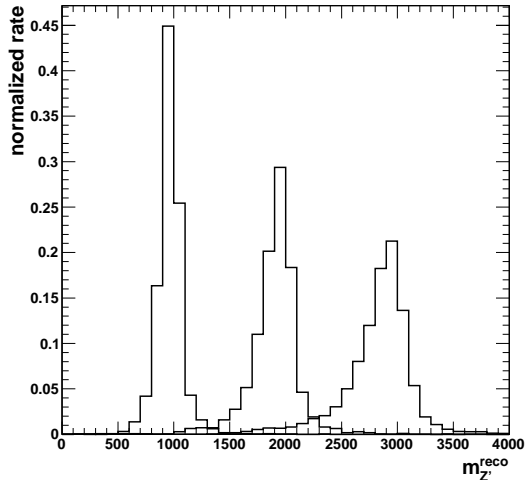
FIG. 6: *Reconstructed Z' invariant masses for 1, 2, and 3 TeV in the $Zh \to (l^+l^-)(b\bar{b})$ mode.*

As one benchmark to interpret the results, we point out that a 3 TeV Z' with $\sigma \times BR \simeq$ 1.5 fb, such as the warped 5D case discussed in section II, can be discovered with about 100 fb^{-1} of data (and with $O(1)$ S/B ratio). Previous estimates had placed discovery at the ab^{-1} scale [6] or higher [29].

B. Zh with leptonic Z: $(l^+l^-)(b\bar{b})$

For this analysis, we reconstruct the leptonic Z by demanding exactly two isolated leptons with opposite-sign and same-flavor, and which fall into the mass window $m_{l^+l^-} = [80, 100]$ GeV. These should be within $|\Delta\phi| < \pi/2$ of each other, and the leptonic Z reconstructed from them should be $|\Delta\phi| > \pi/2$ away from the Higgs candidate jet. The Z' is reconstructed as a simple four-vector sum of the leptonic Z with the Higgs-subjets. As with the W^+W^- analysis above, we demand that the Higgs candidate satisfy $p_T > m_{Z'}^{\rm reco}/3$. Figure 6 displays the reconstructed resonance peaks after application of cuts.

Searches for a light Higgs decaying to $b\bar{b}$ traditionally rely heavily on b-tagging. However, the quality of vertex-based b-tagging is expected to quickly degrade at high p_T, and we do not here have the ability to model this. Nonetheless, we can ask whether a more robust form of b-tagging can be applied, or whether b-tagging needs to be applied at all. In lieu of capitalizing on the displaced vertices of B hadrons, we use the well-exploited fact that

23

	$m_{Z'} = 1$ TeV	$m_{Z'} = 2$ TeV	$m_{Z'} = 3$ TeV
Signal Eff.	1.3% (0.39%)	1.7% (0.53%)	1.7% (0.57%)
$\sigma(Zj: qg \to qZ)$	5.68 fb (0.16 fb)	0.18 fb (0.0066 fb)	0.014 fb (0.0006 fb)
$\sigma(Zj: q\bar{q} \to gZ)$	1.52 fb (0.09 fb)	0.067 fb (0.0053 fb)	0.006 fb (0.0005 fb)
$\sigma(Z(W/Z))$	0.093 fb (0.009 fb)	0.0042 fb (0.00011 fb)	0.00044 fb (0.000014 fb)
$\sigma(Zh)$	0.046 fb (0.013 fb)	0.0024 fb (0.0007 fb)	0.00025 fb (0.000076 fb)

TABLE V: *Signal efficiency and background cross sections after all cuts in $Zh \to (l^+l^-)(b\bar{b})$. Note that the signal efficiency includes the branching fractions for the Z and h decays. The numbers in parentheses are after μ-tagging.*

Luminosity	$m_{Z'} = 1$ TeV	$m_{Z'} = 2$ TeV	$m_{Z'} = 3$ TeV
$\mathcal{L} = 30$ fb^{-1}	186.2 fb (121.9 fb)	27.8 fb (62.4 fb)	19.7 fb (59.0 fb)
$\mathcal{L} = 100$ fb^{-1}	102 fb (66.8 fb)	15.2 fb (18.7 fb)	5.9 fb (17.7 fb)
$\mathcal{L} = 300$ fb^{-1}	58.9 fb (38.6 fb)	8.8 fb (6.2 fb)	2.4 fb (5.9 fb)
$S/B = 1$	552.6 fb (69.8 fb)	15.3 fb (2.4 fb)	1.2 fb (0.2 fb)

TABLE VI: $\sigma(Z') \times BR(Z' \to Zh)$ *required for discovery or $S/B = 1$ in the $(l^+l^-)(q\bar{q}')$ mode. The numbers in parantheses are after μ-tagging.*

they have fairly high semi-leptonic branching fractions, specifically to the very clean case of muons. A muon can be produced either promptly from the b quark decay, or from a subsequent charm decay, leading to a net BR of about 20% for $B \to \mu$+anything. Since the Higgs decay gives us two chances to find one of these decays, our net tag rate for ≥ 1 muon embedded within the Higgs-subjets should approach 35%, independent of p_T. Physical backgrounds to this include prompt heavy flavor (b- and c-jets), gluons splitting to heavy flavor within light QCD jets, and decays-in-flight of pions and kaons (which we include in our PYTHIA simulations). We implement the tag by searching out muons with $p_T > 5$ GeV within cones of size $R = \Delta R_{\text{subjets}}$ around each subjet. In practice, we find tag rates of roughly 30%, and mistag rates at the level of 3-8%, increasing slighly with p_T and generally somewhat higher for gluons. There will also be backgrounds from fake muons, but we assume that these can be brought to a level below the physical mistag rate.

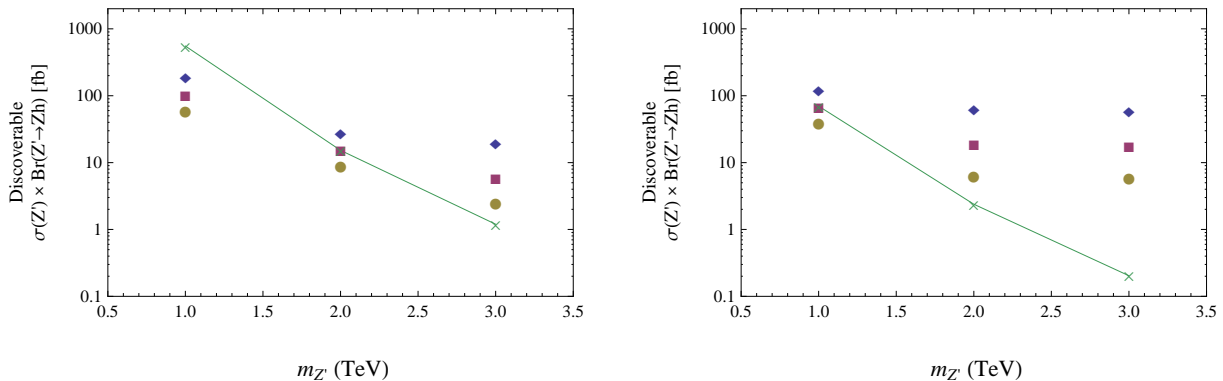We run two concurrent analyses, one without a b-tag and one with the μ-based b-tag.

FIG. 7: *Discovery reach for $Zh \to (l^+l^-)(b\bar{b})$ without (left) and with (right) μ-tagging. The three symbols (diamond, box, circle) correspond to (30, 100, 300) fb^{-1} of integrated luminosity. The $\times$ refers to the value that is required to achieve $S/B = 1$.*

The signal box is $m_h^{\mathrm{reco}} = [100, 140]$ GeV and $m_{Z'}^{\mathrm{reco}} = [m_{Z'} - 15\%, m_{Z'} + 15\%]$. The final signal efficiencies and background cross sections are shown in table V. The discovery reach is listed in table VI, and displayed graphically in fig 7, along with the $\sigma \times BR$ required to achieve $S/B = 1$.

We can see that the statistical reach and S/B are clearly improved by the μ-tag for lower masses, but it becomes overkill at higher masses, where the untagged background is only $O(1)$ events even at 300 fb^{-1}. For 2 and 3 TeV, the tagged discovery reach is essentially controlled by the $N_S = 10$ requirement, with near-vanishing background. Clearly, b-tagging is most important at sub-TeV masses, where the background can still be substantial. Fortuitously, sub-TeV masses are exactly where traditional b-tagging techniques would operate well. But we learn here that b-tagging in this mode at multi-TeV mass may in fact be counterproductive, at least from a strictly statistical standpoint.

Using the same benchmark model as in the previous subsection, 3 TeV and 1.5 fb^{-1}, we find that discovery would require a bit less than 400 fb^{-1}. This can be contrasted with [6], which required slightly less than 1 ab^{-1}.

C. Zh with invisible Z: $(\nu\bar{\nu})(b\bar{b})$

At low p_T, processes involving an invisible Z are challenging because the neutrinos are nearly back-to-back. In our case, they are nearly collinear, leading to striking monojet
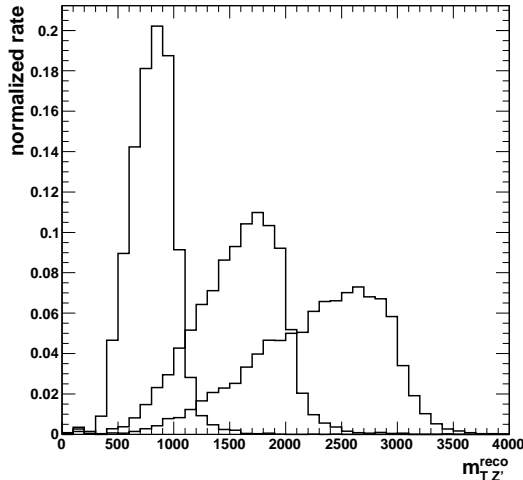
25

FIG. 8: *Reconstructed Z' transverse masses for 1, 2, and 3 TeV in the $Zh \to (l^+l^-)(b\bar{b})$ mode.*

signals with a single TeV-scale jet and almost no other activity. Honestly modeling the background in this case can be somewhat subtle, as non-gaussian fluctuations in energy measurements can create fake missing energy. However experimental simulation studies (see e.g. [79]) indicate that generic QCD backgrounds, as well as top and W backgrounds, can be readily controlled. Consequently, we focus exclusively on Z+jets and SM diboson backgrounds.

We use our default analysis cuts, and apply no special cuts to insist on large missing energy. A more realistic analysis would require some kind of veto on activity opposite the Higgs candidate, but this would still pass the signal with very high efficiency. We define missing energy as simply the transverse momentum vector that balances the Higgs, $\vec{\rlap{E}}_T \equiv -\vec{p}_T(h)$. A more refined measurement would also take into account energy deposited by ISR jets, etc.

Proper reconstruction of the Z' mass peak is impossible, as there is no unambiguous way to know the p_Z of the invisible Z. However, the p_T of the Z' decay products has a Jacobian peak which is spin-enhanced. In figure 8, we show the distribution of the reconstructed Z'

26

	$m_{Z'} = 1$ TeV	$m_{Z'} = 2$ TeV	$m_{Z'} = 3$ TeV
Signal Eff.	5.4% (1.7%)	6.8% (2.2%)	7.1% (2.3%)
$\sigma(Zj:\ qg \to qZ)$	43.0 fb (1.54 fb)	1.42 fb (0.049 fb)	0.11 fb (0.0058 fb)
$\sigma(Zj:\ q\bar{q} \to gZ)$	9.2 fb (0.46 fb)	0.38 fb (0.03 fb)	0.037 fb (0.003 fb)
$\sigma(Z(W/Z))$	0.87 fb (0.044 fb)	0.064 fb (0.0024 fb)	0.0037 fb (0.00095 fb)
$\sigma(Zh)$	0.073 fb (0.025 fb)	0.0035 fb (0.00097 fb)	0.00041 fb (0.0001 fb)

TABLE VII: *Signal efficiency and background cross sections after all cuts in $Zh \to (\nu\bar{\nu})(b\bar{b})$. Note that the signal efficiency includes the branching fractions for the Z and h decays. The numbers in parentheses are after μ-tagging.*

Luminosity	$m_{Z'} = 1$ TeV	$m_{Z'} = 2$ TeV	$m_{Z'} = 3$ TeV
$\mathcal{L} = 30$ fb^{-1}	123.7 fb (78.6 fb)	18.2 fb (15.5 fb)	4.97 fb (14.8 fb)
$\mathcal{L} = 100$ fb^{-1}	67.7 fb (43.1 fb)	9.97 fb (6.64 fb)	2.7 fb (4.4 fb)
$\mathcal{L} = 300$ fb^{-1}	39.1 fb (24.9 fb)	5.76 fb (3.83 fb)	1.57 fb (1.48 fb)
$S/B = 1$	987.7 fb (123.8 fb)	27.2 fb (3.8 fb)	2.1 fb (0.43fb)

TABLE VIII: *$\sigma(Z') \times BR(Z' \to Zh)$ required for discovery or $S/B = 1$ in the $(\nu\bar{\nu})(b\bar{b})$ mode. The numbers in parentheses are after μ-tagging.*

transverse mass,[19]

$$m_{T\,Z'}^{\text{reco}} \equiv \sqrt{p_T(h)^2 + m(h)^2} + \sqrt{\not{E}_T^2 + m_Z^2} \simeq 2p_T(h). \tag{3}$$

As a signal box, we take $m_h^{\text{reco}} = [100, 140]$ GeV and $m_{T\,Z'}^{\text{reco}} = [m_{Z'} - 30\%, m_{Z'} + 10\%]$. As in the leptonic Z analysis, we consider cases with/without μ-tagging. The final signal efficiencies and background cross sections are shown in table VII. The discovery reach is listed in table VIII, and displayed graphically in fig 9, along with the $\sigma \times BR$ required to achieve $S/B = 1$.

---

[19] We also explored two other types of reconstruction, exploiting the fact that the Z' is produced approximately at rest in the lab frame. We considered the mass formed by assuming that the Z and h are produced back-to-back in three dimensions in the lab frame, as well as a mixture between this and the transverse mass by using the full h three-vector and the invisible Z transverse vector. Of our three Z' reconstruction methods, we find that the transverse mass works the best for all of the physical $m_{Z'}$ that we consider.
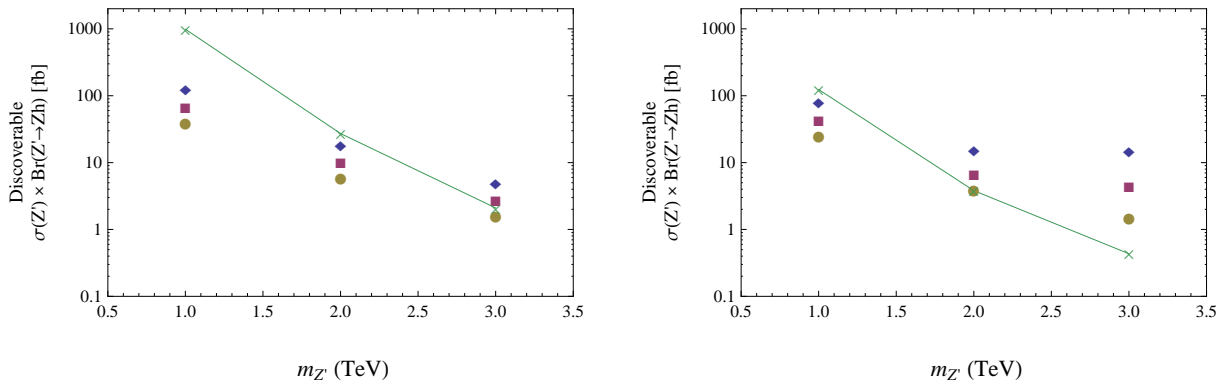
FIG. 9: *Discovery reach for $Zh \to (\nu\bar{\nu})(b\bar{b})$ without (left) and with (right) μ-tagging. The three symbols (diamond, box, circle) correspond to (30, 100, 300) fb^{-1} of integrated luminosity. The $\times$ refers to the value that is required to achieve $S/B = 1$.*

From the perspective of pure statistics, our analysis of Zh with invisible Z suggests that it may be even more sensitive than the corresponding analysis with leptonic Z, mainly owing to the three times higher event rate (as well as our lack of reconstruction criteria on the Z). Of course, this comes at a price of lower S/B, and less distinctive signal shape, as the kinematic recontruction is less precise. Still, μ-tagging is no longer as detrimental above 2 TeV, for example allowing discovery of the 3 TeV warped Z' at about 300 fb^{-1} with $S/B \simeq 3$. We believe that this mode merits a more prominent role for resonance searches at the LHC.

V. CONCLUSIONS

In this paper, we have explored the utility of jet substructure methods to aid in the search for a Z' decaying to electroweak bosons at the LHC. Our findings can be naturally divided into two parts. First, we have made a detailed categorization of the performance of the BDRS substructure algorithm for identifying boosted hadronically-decaying electroweak bosons in the TeV regime. Second, we have estimated the Z' discovery potential in several final-state channels, demonstrating significant improvements over previous analyses.

In order to clearly justify the use of substructure, we have made a comparison between BDRS and more traditional jet-based reconstructions. While the difference is modest for genuine boosted electroweak bosons, we isolated several advantages in the treatment of the background. For QCD jets with mass near the electroweak boson mass, BDRS achieves

$O(1)$ greater discrimination by exploiting the jets' internal kinematics. Beyond this, since BDRS is free of artificial ΔR thresholds, the background mass spectrum is free of artificial mass thresholds, and it is also much flatter.

Another important observation is the stability and benefit of using a large fat-jet clustering radius. We chose a radius of $R = 1.4$, which covers an area much larger than a traditional jet, and in particular much larger than a TeV-scale electroweak boson-jet. In some sense, we are no longer applying "jet" substructure but rather approaching "event" substructure. While the entire event is contaminated with plentiful uncorrelated radiation, we nonetheless find that BDRS is extremely efficient at zooming-in to the most interesting region of activity. The scale-invariant picture of the event offered by this procedure offers several advantages. As noted above, the background mass spectrum becomes modestly-sloped and largely featureless. The large catchment area of the fat-jet allows a single analysis strategy to interpolate smoothly between resonance searches in the several-hundred GeV range up to the multi-TeV range. Finally, a nontrivial amount of global color-discrimination is incorporated for free.

Ultimately, this scale-invariant picture breaks down once we hit the size of the individual detector elements. We incorporated into our analysis a primitive detector model which highlights this limitation, and suggests how to overcome it. We used the highly-segmented ECAL as a tracer of spatial energy flow, and incorporated the full ECAL+HCAL energy measurements. While a procedure like this is not so crucial for a 1 TeV Z', it becomes very important for 3 TeV, where the quarks from the secondary decay of a W will usually sit in adjacent HCAL cells. We do not expect that our model gives a truly accurate picture of the jet mass resolution at very high p_T, but it does demonstrate that the LHC detectors have enough information to resolve the individual quarks. Our resonance search results should be largely robust if the true mass resolution can come within $O(1)$ of our estimate of roughly 12%.

For our resonance searches, we explored decays into Zh and W^+W^-, the former under the assumption of a 120 GeV SM-like Higgs decaying primarily to $b\bar{b}$. In general, we find much better reach than what has previously been claimed. For example, a 3 TeV Z' with fb-scale cross section need not be relegated to the Super-LHC. This specific mass has become a kind of benchmark for composite/5D Higgs models, as a minimum to pass electroweak precision tests. We find that discovery should be possible with as little as 100 fb^{-1} in the W^+W^-

mode, with $O(3)$ times higher luminosity required for Zh to catch up.

Previous studies of Zh have focused exclusively on the $(l^+l^-)(b\bar{b})$ final state utilizing b-tags. It is well-known that b-tagging performance degrades at high p_T, owing to the increasing collinearity of hits in the inner trackers. In order to bypass the complicated modeling of this behavior, we have focused exclusively on a muon-based b-tagger, which should extrapolate robustly to high energies. We found that this tagger improves discovery reach below 2 TeV, but that it becomes counterproductive at higher masses due to the $O(1/3)$ signal tagging efficiency and the small number of expected untagged background events at normal LHC luminosities. Of course, it is possible that a more sophisiticated high-p_T b-tagger could perform better.

We have also studied the largely neglected mode $(\nu\bar{\nu})(b\bar{b})$. By reconstructing the Z' transverse mass, we find that this channel is highly competitive. The three times higher statistics appears to more than compensate for the less precise kinematics. In the case that a discovery is claimed of $Z' \to Zh$, visible and invisible Z modes will serve as important cross-checks of each other.

We conclude that jet substructure techniques are very promising for identifying highly boosted electroweak bosons, and that these techniques can be fruitfully applied to Z' searches. (Though we have not explored the following point in detail, our methods can also be immediately applied to searches for other objects that decay into electroweak bosons, such as W' and KK gravitons, as well as searches in weak boson fusion.) The scope of our analysis was limited to the case of two-body boson decays into quarks. This is adequate for W (and Z) bosons, but Higgses clearly present a much richer set of possible hadronic and semi-hadronic final states. Identifying these decays will require a broader set of substructure tools than what we have implemented here. We relegate their study to future work [25, 26].

Appendix A: Pileup and Filtering

In the analyses of this paper, we consider Z' masses starting at 1 TeV. In this kinematic region, we find that the effects of underlying event contamination within the subjets are modest. However, searches extending to multi-TeV Z' masses will only be made possible by high-luminosity running, where we inevitably face the issue of approximately 20 min-bias pileup collisions superimposed on each Z' candidate event. Determining the best way to remove the pileup contamination (and the underlying event contamination at lower $m_{Z'}$) is beyond the scope of this paper. However, we do pause here to consider how large the effect is and how effective filtering is at correcting it.

We simulate pileup by adding particles from PYTHIA min-bias events to our Z' events. The number of min-bias collisions sampled is drawn, on an event-by-event basis, from a Poisson distribution with mean of 20. In principle, charged pileup can be tracked back to event vertices distinct from the Z' production and subtracted, leaving over an irreducible neutral pileup component. On the other hand, the detectors themselves perform part of this subtraction "for free," by sweeping away soft charged particles in the magnetic field. To coarsely model this, we also consider events where all charged particles below $p_T = 1$ GeV, including those from the primary hard collision, are removed by-hand prior to jet clustering.[20] To factorize this discussion from the possible degrading effect of the magnetic field on the mass reconstruction (as well as all other detector-related issues, which we discuss in appendix B), we work at particle-level without altering the particles' trajectories.

We study the effect of pileup on the reconstructed Higgs mass in the 1 and 2 TeV $Z' \to Zh \to (l^+l^-)(b\bar{b})$ samples. (The WW mode is less sensitive to this contamination at a given Z' mass because $\Delta R_{\text{subjets}}$ tends to be smaller. Partially this is because the W is lighter than the Higgs, but also because the W is highly longitudinally polarized.) The results are presented in Fig. 10.[21] The effects of pileup are, unsurprisingly, most pronounced in

---

[20] The magnetic field and inner HCAL radius of the ATLAS and CMS experiments are, respectively, 2.3m/2T and 1.8m/4T. With these parameters, charged hadrons with $p_T < 0.7$ GeV and $p_T < 1.1$ GeV will not reach the HCAL.

[21] Note that the uncontaminated mass peak already displays an asymmetric broadening. This is due to physics, not a flaw in the reconstruction. Most Higgs decays through $b\bar{b}$ contain at least one neutrino in
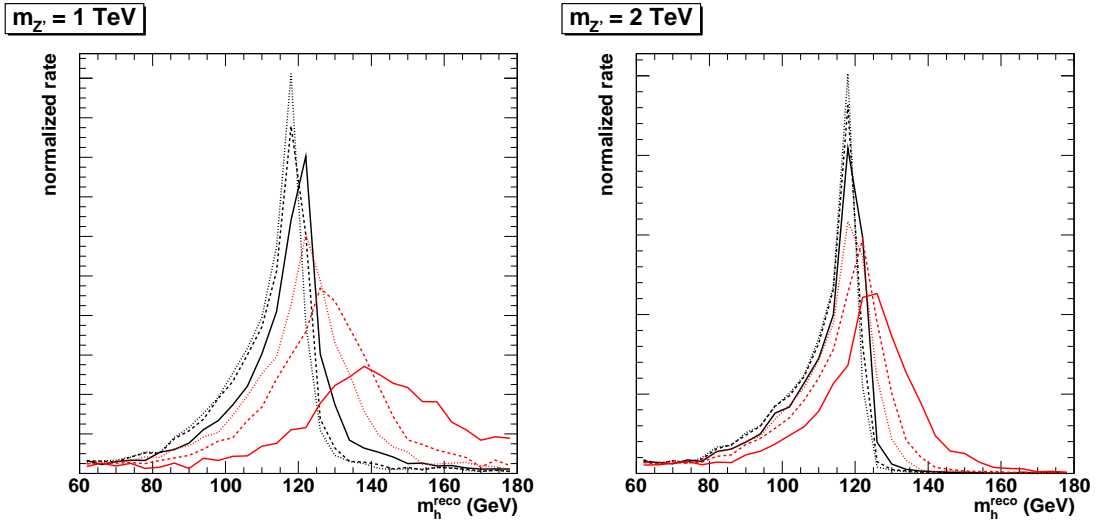
FIG. 10: *Distributions of the (particle-level) reconstructed Higgs mass for 1 and 2 TeV $Z' \to Zh \to (l^+l^-)(b\bar{b})$. Displayed are results without pileup (black) and with pileup (red). Results are further subdivided into: before filtering (solid), after filtering (dashed), and after filtering with a soft track cutoff (dotted).*

the 1 TeV sample. The Higgs mass peak becomes shifted upward by about 20 GeV, and broadened by a comparable amount. The effect is roughly cut in half by filtering, and cut in half again by applying the soft track cutoff. This reduces the shifting/broadening down to a level that is very likely below that of the realistic experimental resolution. We have also investigated the case with purely neutral pileup (not shown in Fig. 10), and find that it looks quite similar to the track-cutoff case.

The effect of pileup on background reconstruction is more subtle. QCD subjet configurations with high mass prefer to be highly asymmetric in energy, and the most extreme cases are automatically ignored by the BDRS procedure. However, such configurations can be pushed above the symmetry threshold y_{cut} by the addition of pileup particles to the softer subjet. This can be particularly problematic with our very large fat-jet size ($R = 1.4$). Filtering, even combined with a low-p_T track cutoff, is not powerful enough to fully counteract this. As an example, we show in the left panel of Fig. 11 the m_h^{reco} spectrum from our Z+jet

---

the final state. Some of the broadening is also due to underlying event contamination, particularly at 1 TeV. Filtering partially corrects this off.
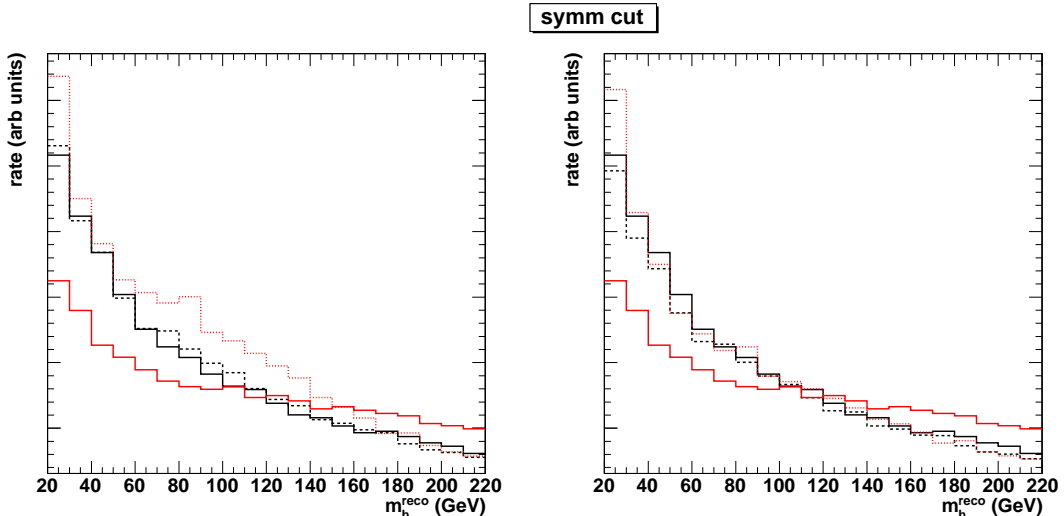
FIG. 11: *Distributions of the (particle-level) reconstructed Higgs mass for $Z+jets \to (l^+l^-)+jets$ backgrounds generated at $\sqrt{\hat{s}} = 1$ TeV, without (left) and with (right) a symmetry cut applied after filtering. Displayed are results without pileup (black) and with pileup (red). Results are further subdivided into: before filtering (solid), after filtering (dashed), and after filtering with a soft track cutoff (dotted). (Omitted for clarity are no-pilup with filtering and track cutoff (black-dotted), and pileup with filtering and without track cutoff (red-dashed).) Note that the arbitrarily-normalized vertical scales on the two plots are identical.*

background generated near $\sqrt{\hat{s}} = 1$ TeV.[22] After application of the track cutoff and filtering, there is a leftover enhancement of the background spectrum near 100 GeV. We find that the excess is dominated by the cases described above, but that these can easily be removed by insisting that the second-hardest subjet remain fairly hard after filtering. For example, imposing a post-filtering symmetry cut p_T(filtered subjet #2)$/p_T$(filtered jet) > 0.1 is adequate, as shown in the right panel of Fig. 11. Without pileup, this cut is largely redundant with $y_{\rm cut}$ and has small effect. The effect on the Z' signal with pileup is at the 1% level.

---

[22] For the purpose of this study, these events are processed without b-tagging and without isolation requirements on the leptons, the latter in order to factorize out changes in lepton identification efficiency in the presence of pileup. We pick a fixed slice of parton-level $\sqrt{\hat{s}}$ instead of a fixed slice of $m_{Z'}^{\rm reco}$ in order to disentangle biases introduced by pileup on the latter quantity. (Such bias is largely corrected away by filtering, but this is difficult to see simultaneously without a full two-dimensional plot.) While we do not show results at $\sqrt{\hat{s}}$ above 1 TeV, we indeed find that the effects of pileup become milder, as we did for the Z' samples.
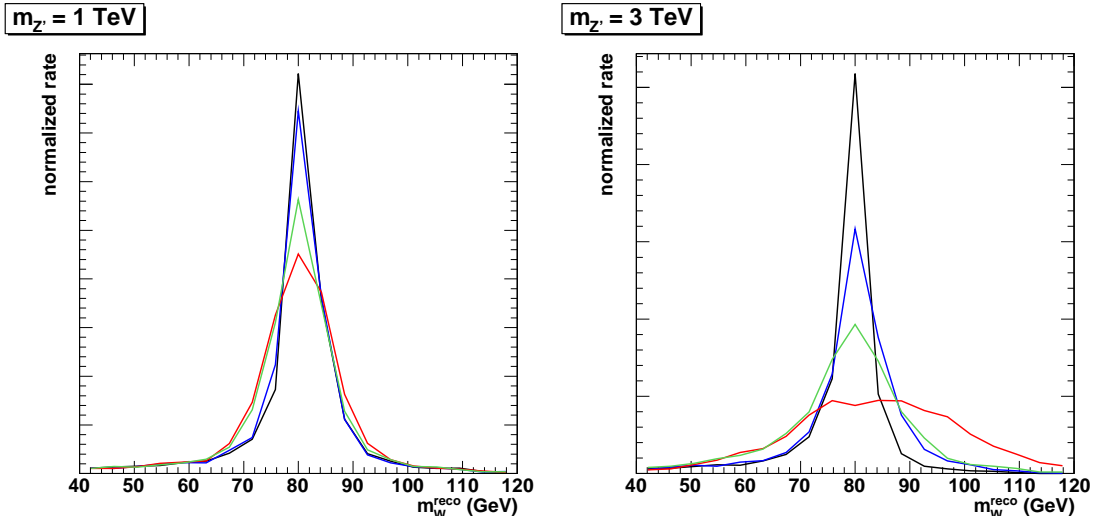
FIG. 12: *Distributions of the reconstructed hadronic W mass for 1 and 3 TeV $Z' \to WW \to (l\nu)(q\bar{q}')$. Displayed are particle-level (black), idealized particle-flow (blue), rescaled ECAL (green), and pure HCAL (red). Detector models are described in more detail in the text.*

Having demonstrated that the degrading effects of pileup can probably be mostly removed for TeV-scale Z' masses, we do not incorporate pileup or filtering into our boosted electroweak boson studies. However, we do emphasize that any realistic broad-spectrum Z' search should ideally utilize a pileup/UE subtraction method that works well in the sub-TeV mass range but applies smoothly up to trans-TeV masses without adversely affecting the reconstruction there. We also note that the alternative procedures of pruning [13] and trimming [63] are also worth exploring, possibly in combination with filtering and with each other [80].

Appendix B: Detector Effects

While we do not have the tools to fully address the impact of the detector on the quality of boosted boson reconstruction, we can at least make plausibility arguments and highlight what issues might affect the jet-mass measurement. Probably the biggest worry is that for TeV-scale W-jets, the spatial resolution of the detector will be too poor to resolve the mass peak. To get a feeling for how bad the situation might be, we consider four toy models (inspired by the CMS detector) with decreasing spatial resolution:

34

1. Simple particle-level as a baseline.

2. An idealization of particle-flow: charged particles are measured perfectly, photon energy is deposited into perfect $\Delta\eta \times \Delta\phi = 0.02 \times 0.02$ ECAL cells, and neutral hadron energy is deposited into perfect 0.1×0.1 HCAL cells. The calorimeter cells then serve as massless particles (with momentum vector oriented along the cell's center in η-ϕ) for the purposes of jet clustering.

3. An idealized ECAL+HCAL model where ECAL cells are used to trace the energy flow. Photons and non-isolated electrons are deposited into the ECAL, and all hadrons are deposited into the HCAL. (Muons and isolated electrons are kept as tracks.) Each HCAL cell has an associated 5×5 block of ECAL cells. The ECAL energy is rescaled to match the ECAL+HCAL energy, and the HCAL cells are discarded. The rescaled ECAL cells are used in jet clustering.

4. An idealized pure HCAL, where all particles (except for muons and isolated electrons) are deposited in 0.1×0.1 cells.

We apply these four models to the 1 and 3 TeV $Z' \to WW \to (l\nu)(q\bar{q}')$ signal samples, as shown in Fig. 12. As one would expect, the impact of the detector model becomes greater at higher Z' mass. For 1 TeV Z', the typical $\Delta R_{\text{subjets}}$ is about 0.35, and our detector segmentation is not a major issue. For 3 TeV Z', the typical $\Delta R_{\text{subjets}}$ is about 0.12, barely large enough for the two subjets to sit in individual HCAL cells. Nonetheless, even the pure HCAL detector manages to reconstruct the W peak, with about 15 GeV resolution. The simple trick of using the ECAL to trace the energy flow improves resolution by nearly a factor of 2, down to 8 GeV, and idealized particle flow incorporating the tracker could in principle reduce the smearing to the 5 GeV level. In what follows, we will take the rescaled ECAL detector as our working model, as a compromise.

Of course, even this model, which neglects the possible utilization of the tracker, is likely optimistic. This is mainly for three reasons: the magnetic field will to some extent spread out the charged particles, calorimeter energy deposits will spread laterally as the showers develop inside the cells, and energy measurements in individual cells are subject to sampling errors and electronics noise. We address each of these in turn.

To coarsely model the effect of the magnetic field, we rotate the momentum vectors of charged particles in ϕ so that they are oriented toward the point at which they would impact the innermost part of the calorimeter on an arced trajectory. We take an inner calorimeter radius of 2 meters, and a magnetic field strength of 4T. Particles of $p_T < 1.2$ GeV spiral-out and are lost. Using the rescaled ECAL model and 1 TeV Z', the reconstructed W mass is shifted upward by about 3 GeV and broadened by a comparable amount. As this is a small effect, and can probably be largely corrected off using particle flow techniques,[23] we continue to model the detector without a magnetic field in what follows.

The energy deposited in the calorimeter from a given particle is not usually contained in a single cell, but will leak to some extent into neighboring cells. To get a sense for how sensitive we might be to this leakage, we rerun the 3 TeV Z' sample through the rescaled ECAL model with ECAL and HCAL cells both doubled in lateral size. The net effect is a roughly 50% increase in the $m_W^{\rm reco}$ smearing, to about 12 GeV. This is comparable to the pure 0.1×0.1 HCAL model, and is still very reasonable.[24]

Finally, we also consider the the effects of smearing out the energy measurements. The energy resolution of subjets has not been well-studied in full simulation, so as a rough guess we simply use the resolution curve for $R = 0.5$ cone jets in the CMS detector [81]: $\Delta E/E = 5.6/E \oplus 1.25/\sqrt{E} \oplus 0.033$, with $\oplus$ indicating quadrature sum and E measured in GeV. The effect on the rescaled ECAL results at 3 TeV is modest, whereas the relative effects at smaller Z' mass are more pronounced. The net result is that the W mass resolution comes out to about 10 GeV (12%) for all of our Z' samples. The spatial segmentation effects and energy-sampling effects roughly compensate for each other.

Our nominal choice of detector model for the Z' search is the rescaled ECAL model with energy smearing applied as above. This roughly approximates some of the dominant resolution effects. We note that an ATLAS substructure-based study of a 1.1 TeV WW

---

[23] The largest degrading effect is from the softer particles, which separate out from the core of the jet. These are less susceptible to crowding of hits, which can frustrate precision tracking.

[24] A further concern is that a very energetic QCD background jet will appear artificially massive due to its energy spreading into several cells. For example, a very tight 1 TeV jet that shares its energy equally between two adjacent 0.1×0.1 HCAL cells would appear to have a mass of 50 GeV, even if its true mass is much smaller. Incorporating the energy pattern in the ECAL then becomes crucial. As long as this is done, we do not expect that spatial smearing effects will have such a significant impact on the background jet-mass distribution.

resonance in vector boson fusion [30] quotes a resolution for m_W^{reco} of 7.4 GeV, which is better than what we obtain with our model. However, a full simulation study at higher resonance mass is clearly warrented.

[1] D. Rainwater, *Searching for the Higgs Boson* (2007), hep-ph/0702124.

[2] J. Bagger et al., *CERN LHC Analysis of the Strongly Interacting W W System: Gold Plated Modes*, Phys. Rev. **D52**, 3878 (1995), hep-ph/9504426.

[3] J. M. Butterworth, B. E. Cox, and J. R. Forshaw, *WW Scattering at the CERN LHC*, Phys. Rev. **D65**, 096014 (2002), hep-ph/0201098.

[4] M. Schmaltz and D. Tucker-Smith, *Little Higgs Review*, Ann. Rev. Nucl. Part. Sci. **55**, 229 (2005), hep-ph/0502182.

[5] H. Davoudiasl, S. Gopalakrishna, E. Ponton, and J. Santiago, *Warped 5-Dimensional Models: Phenomenological Status and Experimental Prospects*, New J. Phys. **12**, 075011 (2010), 0908.1968.

[6] K. Agashe et al., *LHC Signals for Warped Electroweak Neutral Gauge Bosons*, Phys. Rev. **D76**, 115015 (2007), 0709.0007.

[7] T. Han, H. E. Logan, B. McElrath, and L.-T. Wang, *Phenomenology of the Little Higgs model*, Phys. Rev. **D67**, 095004 (2003), hep-ph/0301040.

[8] J. M. Butterworth, A. R. Davison, M. Rubin, and G. P. Salam, *Jet Substructure as a New Higgs Search Channel at the LHC*, Phys.Rev.Lett. **100**, 242001 (2008), arXiv:0802.2470.

[9] D. E. Kaplan, K. Rehermann, M. D. Schwartz, and B. Tweedie, *Top Tagging: A Method for Identifying Boosted Hadronically Decaying Top Quarks*, Phys. Rev. Lett. **101**, 142001 (2008), 0806.0848.

[10] T. Plehn, G. P. Salam, and M. Spannowsky, *Fat Jets for a Light Higgs*, Phys. Rev. Lett. **104**, 111801 (2010), 0910.5472.

[11] K. Rehermann and B. Tweedie, *Efficient Identification of Boosted Semileptonic Top Quarks at the LHC* (2010), 1007.2221.

[12] M. H. Seymour, *Searches for New Particles Using Cone and Cluster Jet Algorithms: A Comparative Study*, Z. Phys. **C62**, 127 (1994).

[13] S. D. Ellis, C. K. Vermilion, and J. R. Walsh, *Techniques for Improved Heavy Particle Searches*

with Jet Substructure, Phys. Rev. **D80**, 051501 (2009), 0903.5081.

[14] L. G. Almeida et al., *Substructure of High-p_T Jets at the LHC*, Phys. Rev. **D79**, 074017 (2009), 0807.0234.

[15] L. G. Almeida, S. J. Lee, G. Perez, G. Sterman, and I. Sung, *Template Overlap Method for Massive Jets*, Phys. Rev. **D82**, 054034 (2010), 1006.2035.

[16] G. Brooijmans, *High pT Hadronic Top Quark Identification. Part I: Jet Mass and YSplitter* (2008), ATL-PHYS-CONF-2008-008.

[17] J. Thaler and L.-T. Wang, *Strategies to Identify Boosted Tops*, JHEP **07**, 092 (2008), 0806.0023.

[18] J. M. Butterworth, J. R. Ellis, A. R. Raklev, and G. P. Salam, *Discovering Baryon-Number Violating Neutralino Decays at the LHC*, Phys. Rev. Lett. **103**, 241803 (2009), 0906.0728.

[19] G. D. Kribs, A. Martin, T. S. Roy, and M. Spannowsky, *Discovering Higgs Bosons of the MSSM Using Jet Substructure* (2010), arXiv:1006.1656.

[20] T. Plehn, M. Spannowsky, M. Takeuchi, and D. Zerwas, *Stop Reconstruction with Tagged Tops* (2010), 1006.2833.

[21] G. Cacciapaglia, C. Csaki, G. Marandella, and J. Terning, *The Gaugephobic Higgs*, JHEP **0702**, 036 (2007), hep-ph/0611358.

[22] J. Galloway, B. McElrath, and J. McRaven, *Signatures of Extra Dimensions from Upsilon Decays with a Light Gaugephobic Higgs Boson*, Phys. Lett. **B670**, 363 (2009), 0807.2657.

[23] J. Galloway, B. McElrath, J. McRaven, and J. Terning, *Gaugephobic Higgs Signals at the LHC*, JHEP **11**, 031 (2009), 0908.0532.

[24] M. A. Diaz and T. J. Weiler, *Decays of a Fermiophobic Higgs* (1994), hep-ph/9401259.

[25] A. Katz, M. Son, and B. Tweedie, to appear.

[26] A. Katz, M. Son, and B. Tweedie, to appear.

[27] V. Barger, P. Langacker, and H.-S. Lee, *Six-Lepton Z' Resonance at the LHC*, Phys. Rev. Lett. **103**, 251802 (2009), 0909.2641.

[28] J. E. Garcia, M. Lechowski, E. Ros, and D. Rousseau (The Atlas), *Search for the Decays $Z_H \to Zh$ and $W_H \to Wh$ in the Little Higgs Model Assuming $m(h) = 120$ GeV* (2004), ATL-PHYS-2004-001.

[29] D. Benchekroun, C. Driouichi, and A. Hoummada, *Search for $Z' \to WW$ at LHC with the ATLAS Detector*, Eur. Phys. J. direct **C3**, N3 (2001).

[30] G. Aad et al. (The ATLAS), *Expected Performance of the ATLAS Experiment - Detector, Trigger and Physics* (2009), 0901.0512.

[31] C. Hackstein and M. Spannowsky, *Boosting Higgs Discovery - The Forgotten Channel* (2010), 1008.2202.

[32] P. Langacker, *The Physics of Heavy Z' Gauge Bosons*, Rev. Mod. Phys. **81**, 1199 (2009), 0801.1345.

[33] E. Gildener and S. Weinberg, *Symmetry Breaking and Scalar Bosons*, Phys. Rev. **D13**, 3333 (1976).

[34] E. Farhi and L. Susskind, *A Technicolored G.U.T*, Phys. Rev. **D20**, 3404 (1979).

[35] B. Holdom, *Techniodor*, Phys. Lett. **B150**, 301 (1985).

[36] T. W. Appelquist, D. Karabali, and L. C. R. Wijewardhana, *Chiral Hierarchies and the Flavor Changing Neutral Current Problem in Technicolor*, Phys. Rev. Lett. **57**, 957 (1986).

[37] M. Bando, T. Morozumi, H. So, and K. Yamawaki, *Discriminating Technicolor Theories through Flavor Changing Neutral Currents: Walking or Standing Coupling Constants?*, Phys. Rev. Lett. **59**, 389 (1987).

[38] V. A. Miransky, *Dynamics of Spontaneous Chiral Symmetry Breaking and Continuum Limit in Quantum Electrodynamics*, Nuovo Cim. **A90**, 149 (1985).

[39] T. Appelquist and L. C. R. Wijewardhana, *Chiral Hierarchies from Slowly Running Couplings in Technicolor Theories*, Phys. Rev. **D36**, 568 (1987).

[40] C. Csaki, C. Grojean, L. Pilo, and J. Terning, *Towards a Realistic model of Higgsless Electroweak Symmetry Breaking*, Phys. Rev. Lett. **92**, 101802 (2004), hep-ph/0308038.

[41] M. A. Luty and T. Okui, *Conformal Technicolor*, JHEP **09**, 070 (2006), hep-ph/0409274.

[42] D. B. Kaplan and H. Georgi, *SU(2) x U(1) Breaking by Vacuum Misalignment*, Phys. Lett. **B136**, 183 (1984).

[43] D. B. Kaplan, H. Georgi, and S. Dimopoulos, *Composite Higgs Scalars*, Phys. Lett. **B136**, 187 (1984).

[44] H. Georgi and D. B. Kaplan, *Composite Higgs and Custodial SU(2)*, Phys. Lett. **B145**, 216 (1984).

[45] M. J. Dugan, H. Georgi, and D. B. Kaplan, *Anatomy of a Composite Higgs Model*, Nucl. Phys. **B254**, 299 (1985).

[46] R. Contino, Y. Nomura, and A. Pomarol, *Higgs as a Holographic Pseudo-Goldstone Boson*,

Nucl. Phys. **B671**, 148 (2003), hep-ph/0306259.

[47] D. B. Kaplan, *Flavor at SSC Energies: A New Mechanism for Dynamically Generated Fermion Masses*, Nucl. Phys. **B365**, 259 (1991).

[48] L. Randall and R. Sundrum, *An Alternative to Compactification*, Phys. Rev. Lett. **83**, 4690 (1999), hep-th/9906064.

[49] H. Davoudiasl, J. L. Hewett, and T. G. Rizzo, *Bulk Gauge Fields in the Randall-Sundrum model*, Phys. Lett. **B473**, 43 (2000), hep-ph/9911262.

[50] A. Pomarol, *Gauge bosons in a Five-Dimensional Theory with Localized Gravity*, Phys. Lett. **B486**, 153 (2000), hep-ph/9911294.

[51] K. Agashe, A. Delgado, M. J. May, and R. Sundrum, *RS1, Custodial Isospin and Precision Tests*, JHEP **08**, 050 (2003), hep-ph/0308036.

[52] N. Arkani-Hamed, A. G. Cohen, and H. Georgi, *Electroweak Symmetry Breaking from Dimensional Deconstruction*, Phys. Lett. **B513**, 232 (2001), hep-ph/0105239.

[53] J. M. Maldacena, *The Large N limit of Superconformal Field Theories and Supergravity*, Adv. Theor. Math. Phys. **2**, 231 (1998), hep-th/9711200.

[54] N. Arkani-Hamed, M. Porrati, and L. Randall, *Holography and Phenomenology*, JHEP **08**, 017 (2001), hep-th/0012148.

[55] R. Rattazzi and A. Zaffaroni, *Comments on the Holographic Picture of the Randall-Sundrum Model*, JHEP **04**, 021 (2001), hep-th/0012248.

[56] K. Agashe, R. Contino, L. Da Rold, and A. Pomarol, *A Custodial Symmetry for $Zb\bar{b}$*, Phys. Lett. **B641**, 62 (2006), hep-ph/0605341.

[57] R. Contino, Z', Z_{KK}, Z^* *and all that: Current bounds and theoretical prejudices on heavy neutral vector bosons*, Nuovo Cim. **123B**, 511 (2008), 0804.3195.

[58] R. Contino, T. Kramer, M. Son, and R. Sundrum, *Warped/Composite Phenomenology Simplified*, JHEP **05**, 074 (2007), hep-ph/0612180.

[59] O. Gedalia and G. Perez, *TASI 2009 Lectures - Flavor Physics* (2010), 1005.3106.

[60] C. Csaki, J. Hubisz, G. D. Kribs, P. Meade, and J. Terning, *Big Corrections from a Little Higgs*, Phys. Rev. **D67**, 115002 (2003), hep-ph/0211124.

[61] C. Csaki, J. Hubisz, G. D. Kribs, P. Meade, and J. Terning, *Variations of Little Higgs Models and their Electroweak Constraints*, Phys. Rev. **D68**, 035009 (2003), hep-ph/0303236.

[62] N. Arkani-Hamed, A. G. Cohen, E. Katz, and A. E. Nelson, *The Littlest Higgs*, JHEP **07**, 034

(2002), hep-ph/0206021.

[63] D. Krohn, J. Thaler, and L.-T. Wang, *Jet Trimming*, JHEP **02**, 084 (2010), 0912.1342.

[64] The ATLAS Collaboration, *ATLAS Sensitivity to the Standard Model Higgs in the HW and HZ Channels at High Transverse Momenta* (2009), ATL-PHYS-PUB-2009-088.

[65] Y. L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber, *Better Jet Clustering Algorithms*, JHEP **08**, 001 (1997), hep-ph/9707323.

[66] M. Wobisch and T. Wengler, *Hadronization Corrections to Jet Cross Sections in Deep- Inelastic Scattering* (1998), hep-ph/9907280.

[67] M. Rubin, *Non-Global Logarithms in Filtered Jet Algorithms*, JHEP **05**, 005 (2010), 1002.4557.

[68] J. Gallicchio and M. D. Schwartz, *Seeing in Color: Jet Superstructure*, Phys. Rev. Lett. **105**, 022001 (2010), 1001.5027.

[69] Y. Hatta and T. Ueda, *Jet Energy Flow at the LHC*, Phys. Rev. **D80**, 074018 (2009), 0909.0056.

[70] Z. Han, private communication.

[71] G. Corcella et al., *HERWIG 6.5: an event generator for Hadron Emission Reactions With Interfering Gluons (including supersymmetric processes)*, JHEP **01**, 010 (2001), hep-ph/0011363.

[72] J. Alwall et al., *MadGraph/MadEvent v4: The New Web Generation*, JHEP **09**, 028 (2007), 0706.2334.

[73] S. Hoeche et al., *Matching Parton Showers and Matrix Elements* (2006), hep-ph/0602031.

[74] T. Sjostrand, S. Mrenna, and P. Skands, *PYTHIA 6.4 Physics and Manual*, JHEP **05**, 026 (2006), hep-ph/0603175.

[75] B. Fuks, M. Klasen, F. Ledroit, Q. Li, and J. Morel, *Precision Predictions for Z' - Production at the CERN LHC: QCD Matrix Elements, Parton Showers, and Joint Resummation*, Nucl. Phys. **B797**, 322 (2008), 0711.0749.

[76] M. Rubin, G. P. Salam, and S. Sapeta, *Giant QCD K-factors Beyond NLO*, JHEP **09**, 084 (2010), 1006.2144.

[77] J. H. Kuhn, A. Kulesza, S. Pozzorini, and M. Schulze, *Electroweak corrections to hadronic production of W bosons at large transverse momenta*, Nucl. Phys. **B797**, 27 (2008), 0708.0476.

[78] M. Cacciari and G. P. Salam, *Dispelling the N^3 Myth for the k_t Jet-Finder*, Phys. Lett. **B641**, 57 (2006), hep-ph/0512210.

[79] The CMS Collaboration, *Search for Mono-Jet Final States from ADD Extra-Dimensions at*

$\sqrt{s} = 10$ *TeV* (2009), CMS PAS EXO-09-013.

[80] D. E. Soper and M. Spannowsky, *Combining Subjet Algorithms to Enhance ZH Detection at the LHC*, JHEP **08**, 029 (2010), 1005.0417.

[81] The CMS Collaboration, *CMS Technical Design Report* (2006).