Study of LHC Searches for a Lepton and Many Jets

Mariangela Lisanti,[1, *] Philip Schuster,[2, 3, †] Matthew Strassler,[4, ‡] and Natalia Toro[2, 3, §]

1Princeton Center for Theoretical Science, Princeton, NJ 08542, USA
2Perimeter Institute for Theoretical Physics, Ontario, Canada, N2L 2Y5
3Institute for Advanced Study, Princeton, New Jersey 08540, USA
4Rutgers University, New Brunswick, New Jersey 08854, USA
(Dated: July 27, 2011)

Searches for new physics in high-multiplicity events with little or no missing energy are an important component of the LHC program, complementary to analyses that rely on missing energy. We consider the potential reach of searches for events with a lepton and six or more jets, and show they can provide increased sensitivity to many supersymmetric and exotic models that would not be detected through standard missing-energy analyses. Among these are supersymmetric models with gauge mediation, R-parity violation, and light hidden sectors. Moreover, ATLAS and CMS measurements suggest the primary background in this channel is from $t\bar{t}$, rather than W+jets or QCD, which reduces the complexity of background modeling necessary for such a search. We also comment on related searches where the lepton is replaced with another visible object, such as a Z boson.

I. INTRODUCTION

Most searches for supersymmetry (SUSY) at hadron colliders focus on signatures with missing transverse energy ($\not{E}_T$). These signatures are motivated by theories with stable, invisible dark matter candidates or long-lived neutral particles, such as the lightest super-partner (LSP). In the first 36 pb^{-1} of LHC data, searches for SUSY in mainly hadronic channels imposed minimum missing energy requirements ranging from 100 to 250 GeV [1–9]. Demanding significant $\not{E}_T$ is a powerful strategy for rejecting Standard Model (SM) backgrounds, but sacrifices sensitivity to a variety of Beyond-the-Standard-Model (BSM) signals with smaller $\not{E}_T$, including many types of weak-scale supersymmetric models. Many of these theories are also unobservable in existing exotica searches, which tend to focus on few-particle resonances. In this work, we consider a search that complements current high-$\not{E}_T$ hadronic analyses by providing sensitivity to models with little or no missing energy.

Large classes of SUSY and exotica models with low $\not{E}_T$ share characteristic features, including (1) high object multiplicity, particularly in scenarios where long decay chains deplete $\not{E}_T$, (2) high $\sum |p_T|$ of reconstructed objects, associated with the produced-particle mass scale, and (3) strong-interaction production cross sections, with modest suppression of lepton or electroweak gauge boson emission in decay chains. We will show that searches that take into account all three of these features have an important role to play at the LHC.

---

*Electronic address: mlisanti@princeton.edu
†Electronic address: pschuster@perimeterinstitute.ca
‡Electronic address: strassler@physics.rutgers.edu
§Electronic address: ntoro@perimeterinstitute.ca

Even for models with large $\not{E}_T$, such searches can have sensitivity comparable to existing jets+$\not{E}_T$ analyses, and when effects significantly reduce the $\not{E}_T$ signal, they can retain and often gain sensitivity. In this sense, they provide excellent complementarity to the $\not{E}_T$-based searches, covering additional large domains of parameter space.

The approach that we will consider involves searching for events with a visible non-jet object(s), a high multiplicity of jets, and S_T above a threshold of $\mathcal{O}(\,\mathrm{TeV})$, where

$$S_T \equiv \not{E}_T + \sum_{\substack{\text{visible} \\ \text{objects}}} |p_T| \; . \tag{1}$$

Such a strategy is certainly not new. High object-multiplicity with large S_T is already used in TeV-scale black hole searches [10–12], where the signal is expected to dominate at $S_T \gtrsim 2$ TeV, as seen in Fig. 1. In this case, QCD multi-jets is the dominant background, orders of magnitude larger than electroweak and top production, and to remove it requires an S_T cut of multiple TeV. While such a cut is reasonable for black holes, whose partonic cross-section grows with energy and whose signal peaks at ultra-high S_T, it is too extreme for more general low-$\not{E}_T$, high-multiplicity new physics scenarios. We will consider a complementary approach, reducing the QCD background by requiring at least one lepton or photon, and thus taking advantage of feature (3) above — the relative ease of producing electroweak bosons or leptons in new physics events. In this way, one gains sensitivity to signals peaked at more moderate $S_T \sim 1$ TeV.

Searches that are much closer to what we advocate include the case of a leptonic Z accompanied by many jets, which has been studied at the Tevatron in [13]. Similar strategies have been used in broad model-independent searches, such as SLEUTH at the Tevatron [14–18], which scanned through many different final states for excesses in the high-S_T tail. The MUSiC analysis at CMS also involves a fairly exhaustive scan seeking discrepancies between data and Monte Carlo in high S_T distributions [19, 20]. Other black-hole searches have examined top [21] or di-lepton final states [22, 23]. However, as yet there does not seem to exist a comprehensive program of dedicated high-multiplicity, high-S_T searches in samples with leptons and/or photons. Our aim is to advocate for such a program, by stressing the robustness and scientific importance of searches of this type.

In this article, we do this by focusing on a specific case, which we believe to be the most powerful of the set – searches requiring a lepton, accompanied by six or more jets above a p_T threshold, and by a small amount of m_T (as might be expected from a W) but with no explicit $\not{E}_T$ cut. This search strategy has a high efficiency for many signals, and it appears to have a relatively simple background, dominated, for suitable cuts, by $t\bar{t}$ plus a small number of extra jets.

The following section expands on our motivation, describing how low-$\not{E}_T$, high-multiplicity signals easily arise in SUSY and in other new-physics models. We discuss backgrounds for a lepton-plus-many-jet search in §III. In §IV, we expand on the features of one family of supersymmetric models, focusing in §IV B on the sensitivity that might be achievable with further analysis of the 36 pb^{-1} of 2010 data. We argue in §IV C that an expanded analysis in 2011 data will significantly enhance this sensitivity, even accounting for increases in p_T thresholds. Cross-checks that could help test whether an observed excess is due to mismodeling of the top background are suggested in §V. We extend our discussion to other models in §VI, thereby demonstrating that our strategy is widely applicable, and conclude in §VII with a few comments.
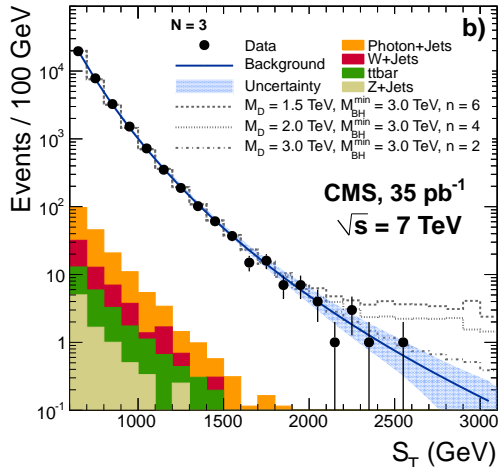
FIG. 1: Plot from CMS [12] of the Standard Model S_T distribution for events with precisely three reconstructed objects of $p_T > 50$. S_T is as defined in Eq. (1). Note the ratio of the electroweak and top pair rates (colored histograms) to the QCD-dominated total S_T distribution.

II.   MOTIVATION FOR THE MULTI-JETS + X STRATEGY

It is quite possible that new physics will first manifest itself as a large signal with many jets, little or no missing energy, and an occasional lepton, photon, or other relatively rare visible object. In this section we review some mechanisms by which this may naturally occur. While this possibility may be realized in a broad set of theories, including models with extra dimensions [24], Higgs and heavy-flavor compositeness [25, 26] and a Little Higgs [27–29], we take supersymmetry as our main example, and focus on the case where the rare object is a lepton. We will see at the end of this section how the use of a lepton-plus-many-jet search strategy can recover sensitivity to models whose $\not{E}_T$ signal is too low to be found by standard means.

A generic signature of supersymmetry involves jets and missing energy. By $SU(3)$ color conservation, jets always accompany strong production of squarks and gluinos. If R-parity is exact, and if the supersymmetric spectrum is minimal with the gravitino heavier than the lightest superpartner, then the lightest SM superpartner is also the lightest R-parity-odd particle — the "LSP" — and is stable. If, in addition, the LSP is produced with moderately high momentum, then a high-$\not{E}_T$ search is very effective. However, the assumptions underlying the high-$\not{E}_T$ strategy are quite model-dependent. Several effects can lead to a reduction in the $\not{E}_T$ signal, and a corresponding increase in the jet multiplicity:

- **Decay of the lightest SM superpartner to a partly-visible final state.** The most obvious feature that can reduce the $\not{E}_T$ signal is a change in the structure of the electroweak decay topology. For instance, in low-scale gauge-mediated SUSY breaking, the gravitino is the lightest R-parity-odd particle. The lightest SM superpartner is the next-to-lightest R-parity-odd particle (the "NLSP"), and can decay to a gravitino in association with its partner. A neutralino NLSP can decay to a photon, Z or Higgs boson. Many of the other possible choices for the NLSP often produce jets and/or taus. Compared to similar models with no such decay, the $\not{E}_T$ in the signal is reduced, and replaced with other objects, which often include two or more jets. Among numerous other classes of examples, the next-to-minimal supersymmetric Standard
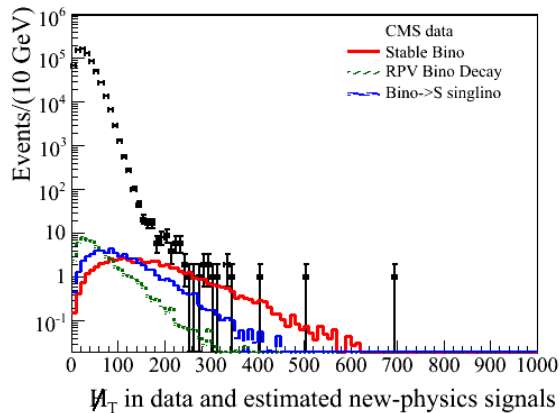
3

FIG. 2: Missing energy distributions for three variations on the low-mass benchmark model described in §IV, in which the bino is either a stable LSP (red solid curve), decays to an NMSSM-like singlet and singlino, followed by singlet decay to two jets (green dotted curve), or decays through R-parity violation to three quarks (blue dashed). A lepton veto, angular cuts, and $H_T > 300$ GeV have been imposed (similar to the base selection of [31]). Shown for comparison is the data from [31], which agree with the data-driven estimate of Standard Model background.

Model (NMSSM) often allows the lightest standard-model superpartner to decay into a mostly singlet Higgs-like scalar and its superpartner, with the scalar in turn decaying to jets (often b's). In SUSY-like models, such as universal or partly-universal extra dimensions with a KK-parity, or Little Higgs with a T-parity, similar considerations apply.

- **Cascade decays or squeezed spectrum.** Even within the minimal supersymmetric Standard Model (MSSM), the kinematics of the decay topology can be unfavorable for $\not{E}_T$-based searches. This happens when the spectrum is modestly squeezed, or when cascade decays into W or Z carry off energy, leaving the LSP with lower p_T and thus reducing the $\not{E}_T$ signal. Again, analogous issues can affect other SUSY-like models, such as universal extra dimensions.

- **Weakly-broken global symmetries (e.g., R-parity violation.)** If R-parity is sufficiently violated, the $\not{E}_T$ signal in a SUSY model can be almost entirely lost, as the LSP decays into two or three jets. The same effect can occur in a Little Higgs model, where T-parity plays a similar role to R-parity, and may be violated in some cases by an anomaly [30]. The KK-symmetries of an extra-dimensional model may also suffer some amount of violation.

- **Top-rich signals.** SUSY models with light stops or sbottoms, or with light Higgsinos, are likely to produce a t and a $\bar{t}$ in a large fraction of the events, if kinematics allow. The same applies for many models of strong dynamics at the electroweak scale, to which the top quark often couples more strongly than other quarks. Compared to a model that produces other quarks, one that produces a t and a $\bar{t}$ will have two to four extra jets (and often a lepton) in the final state.

Meanwhile, often independently of whether the $\not{E}_T$ signal is small or large, leptons and/or photons commonly appear in the cascade decays of the colored particles. The case of a single
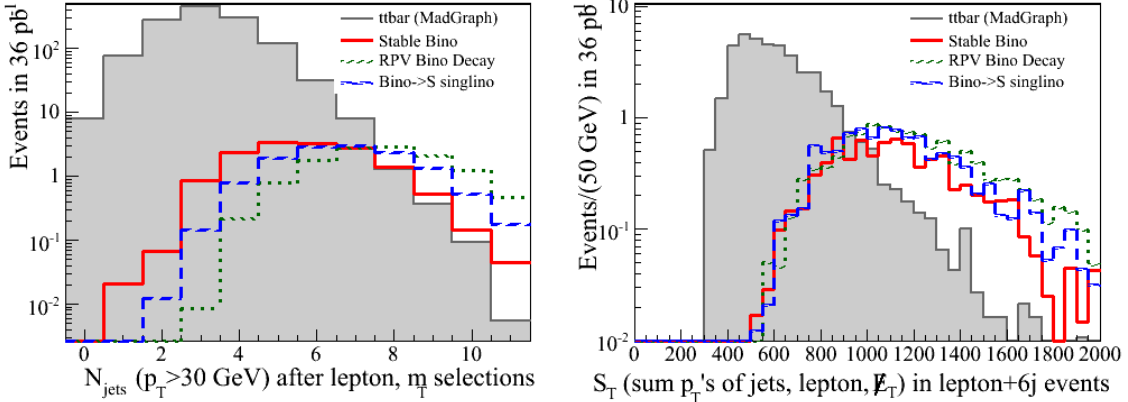
4

FIG. 3: As in Fig. 2, distributions for the low-mass benchmark signals described in §IV. The MadGraph Monte Carlo estimate of the $t\bar{t}$ background, which dominates for $N_{\text{jets}} \gtrsim 4$, is shown in gray. Left: Number of jets ($p_T > 30$ GeV) in each signal after lepton and m_T requirements. Right: The S_T distribution [see Eq. (1)] after requiring 6 jets with $p_T > 30$ GeV.

lepton is of particular interest, because leptons often arise as daughters of W or Z bosons emitted in colored-particle decays, of lepton partners (such as sleptons), or of top quarks. This motivates us to focus our attention in this work on the potential of a lepton-plus-many-jet search.

To demonstrate the impact of reduced $\not{E}_T$ in this case, we show an illustrative trio of examples in Figures 2 and 3. These models will be considered in detail in §IV, to which we refer the reader for a more complete discussion. All three share the same nondescript spectrum of MSSM particles: a 550 GeV gluino, squarks at 800 GeV, and other gauginos obeying approximate mass unification relations. The three models differ only in the fate of the bino, which is taken either to be stable (red solid curves), to decay to a singlet, which decays to two jets, and a stable singlino of the NMSSM (green dotted curves), or to decay to three partons through R-parity violation (blue dashed curves). Figure 2 shows the missing energy distribution for these three models, using base selection cuts similar to those in the CMS jets+$\not{E}_T$ study [31], which require a lepton veto, angular cuts, and $H_T > 300$ GeV. The singlino and RPV scenarios have reduced $\not{E}_T$ and are buried under the background, making them difficult to discover with the standard jets+$\not{E}_T$ searches. Figure 3 shows an alternate approach that requires one lepton, $m_T > 20$ GeV, and no $\not{E}_T$ requirement. In this case, the singlino and RPV scenarios are quite distinct from the background in distributions of jet multiplicity and S_T. The lepton-plus-many-jet strategy is therefore complementary to the $\not{E}_T$-based search. It provides considerable sensitivity even for the stable bino scenario, and has greatly improved sensitivity to the singlino and RPV scenarios, which would be missed by standard searches. More details will be given in §IV.

III.   BACKGROUNDS TO A LEPTON-PLUS-MANY-JET SAMPLE

The relevant backgrounds for a lepton-plus-many-jet search are covered in this section. A discussion of the Monte Carlo modeling procedure and suggestions for future data-driven background estimates are also included. More details of the Monte Carlo validation and
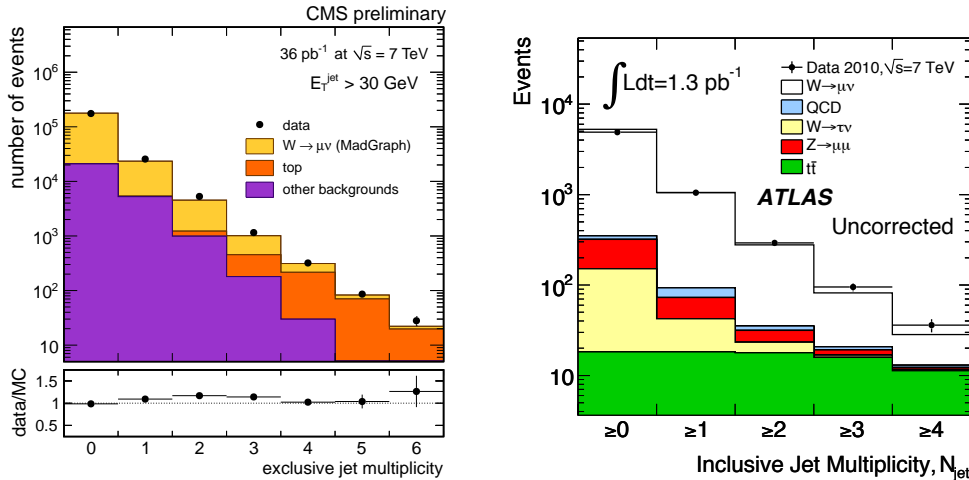
FIG. 4: Distribution of events versus number of jets, taken from W-plus-jets studies at CMS [32] (left) and ATLAS [33] (right). Both plots shown are for the muon channel ($p_T^\mu > 20$ GeV); plots for electrons are found in the references. The CMS (ATLAS) plot uses a jet p_T threshold of 30 (20) GeV and requires $m_T > 20$ (40) GeV; the ATLAS plot further requires $\not{E}_T > 25$ GeV.

detector mock-up can be found in Appendix A.

Production rates for a single lepton plus n jets have been measured at the LHC [32, 33]. Figure 4 shows the distribution of jet multiplicities in the μ plus n jets sample at CMS, for 36 pb^{-1} (left) and at ATLAS for 1.3 pb^{-1} (right). In both cases, there is good correspondence between the normalization of the data and the Monte Carlo predictions, within statistics. Note the rate for SM events in the six-jet single lepton channel is O(pb), as seen in the CMS plot in Fig. 4. At low jet multiplicity, the events are dominated by $W^\pm$ plus n jets, followed by QCD multi-jets where one jet is misidentified as a lepton. In contrast, the five- and six-jet bins are dominated, for jets of p_T of at least 30 GeV, by $t\bar{t}$ plus jets. This is because the jets in $W^\pm$-plus-jets production are generated through perturbative QCD, and are often forward and/or soft, whereas $t\bar{t}$ production can produce up to four parton-level central jets with sizeable p_T, all for the price of α_s^2.

The apparent dominance of the top background is useful and important, since $t\bar{t}$ plus one or two jets can be modeled and measured with somewhat more confidence than $W^\pm$ plus five or six jets.[1] Although Fig. 4 shows that the $W^\pm$-plus-jets background is subdominant at high multiplicities, it could potentially become important again on the high S_T tail, where our search is focused.[2] At high-S_T the top quarks are increasingly boosted and their daughter jets can merge, or ruin lepton isolation. This reduces the $t\bar{t}$ background relative to W-plus-jets. Although we are not aware of evidence that this would make the two backgrounds comparable, this should be checked in data, perhaps using 4- and 5-jet samples. In addition to the absence of reconstructible top quarks in the W-plus-jets background, the rapidity distribution for the lepton is also a handle, as it is central for $t\bar{t}$-plus-jets (and most signals) and flatter for W-plus-jets. (For a recent study out to high multiplicity, see [34].) For the

---

[1] We have checked that $t\bar{t}$ accompanied by other heavy particles, such as W, Z, and h, are a small contribution.

[2] We thank G. Salam for raising this question.

remainder of this work, we assume that the top background remains dominant at the values of S_T (around $1 - 1.5$ TeV) relevant for our studies.

To study this background, a matched $t\bar{t}$-plus-n-jets sample ($n \leq 2$) was generated using `MadGraph 4.4.49` [35] for matrix element generation, `Pythia 6.4` [36] for parton showering and hadronization, and an MLM matching procedure [37] in combination with a shower-$k_\perp$ scheme introduced in [38–40]. Further details on the event generation and matching are provided in the Appendix. The total cross section for the tops is normalized to 150 pb, consistent with the theoretical next-to-leading-order prediction and recent measurements [41, 42].

The selection cuts used here to study the 36 pb^{-1} samples are modeled after those in the CMS study [32]. (When studying the 1 fb^{-1} samples, we raise these cuts; see §IV C.) One lepton is required with $p_T > 20$ GeV and $|\eta| < 2.1$ for a muon and 2.5 for an electron. No direct $\not{E}_T$ cut is applied; however, it is required that $m_T = \sqrt{2p_T \not{E}_T(1 - \cos \Delta\phi)} > 20$ GeV. Jets are formed using an anti-k_T clustering algorithm [43] in `FastJet` [43] with $R = 0.5$. Jets with $p_T > 30$ GeV and $|\eta| < 2.4$ are counted; those that fall within $\Delta R < 0.2$ of an electron are not included in the jet count. Our Monte Carlo model has been tuned and validated against data wherever possible (see the Appendix for details). In particular, the overall rates of three- to six-jet events from $t\bar{t}$ agree with both data and the CMS full simulation Monte Carlo in Fig. 4 (see Fig. 12 for comparison). However, we rely on Monte Carlo for the shape of the S_T distributions and the normalization of the higher-multiplicity jet bins.

It should eventually be possible to model the shape of the distribution using data-driven techniques, with little reliance on Monte Carlo. Preliminary studies in `MadGraph` and `Pythia` Monte Carlo suggest that 4- and 5-jet S_T distributions may be a useful tool in estimating the S_T distribution of events with six or more jets. In particular, the measured power-law of the S_T tails in 4- and 5-jet events can be used to estimate the tail of the higher-multiplicity sample. To model the full distribution, including its rise and turn-over, a refined method, such as kinematics-dependent reweighting of 5-jet events, is needed. Both approaches are illustrated in Figure 5, in the context of a $t\bar{t}$+jets Monte Carlo sample (left) and in the presence of a new-physics signal (right). For background, the lepton plus 4- and 5-jet S_T distribution (dashed purple) models the tail of the 6-jet distribution (thick gray), though it has a much lower threshold and a larger population of low-S_T events. This difference arises largely from the limited phase-space for lower-S_T events to share this energy among six jets with a fixed p_T threshold of 30 GeV. A simple reweighting of five-jet events to account for this effect[3] (orange) yields better agreement with the bulk of the S_T distribution. A new-physics signal then appears as an excess on the tail of the high-multiplicity S_T distribution relative to the lower-multiplicity curves.

---

[3] Five-jet events are weighted according to a "splitting factor" $\sum_{i=1,5} \max\left(0, 1 - \frac{2p_{T,min}}{p_T(j_i)}\right)$, where $p_{T,min} = 30$ GeV is the jet p_T threshold. This function parametrizes the fraction of phase-space in which the splitting of any of the five jets into two lower-p_T jets would produce a six-jet event above threshold. This naive weight could presumably be improved by QCD-splitting-motivated corrections, and by separately modeling the splitting into 7 or more jets.
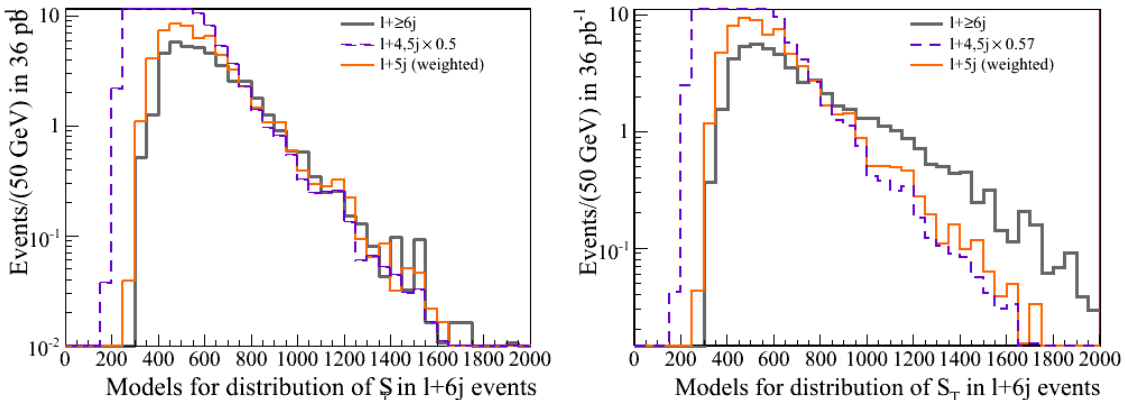
FIG. 5: Suggestion of how lower-multiplicity S_T distributions could be useful for a data-driven model of the $t\bar{t}$ background. Left: The S_T distribution in events with a lepton and $\geq$ 6-jet events (thick gray), and two distributions from which it could be modeled (see text): the S_T distribution of 4 and 5-jet events (purple dashed) after lepton and m_T selections, and 5-jet events reweighted (solid orange) as described in Footnote 3. Right: Same distributions, in the presence of the low-mass singlino benchmark [see Eq. (4)], which distorts the 6-jet S_T distribution but has limited impact at lower multiplicity.

IV.   AN ILLUSTRATIVE CASE STUDY

This section explores how the lepton-plus-many-jet search strategy can be applied in a particular case study. The focus will be on a simplified model (described in §IV A) within the MSSM, and two variants thereof in which the lightest MSSM neutralino decays, decreasing the $\not{E}_T$ and increasing the number of jets in a typical event. §IV B considers the lepton-plus-many-jet strategy in the context of the 2010 data set of 36 pb^{-1}, for which the existence of published data on $t\bar{t}$ and other backgrounds reduces our dependence on Monte Carlo simulations. Specifically, the Monte Carlo background generated in this work can be cross-checked against the published W and Z+jets studies at CMS [32] and ATLAS [33]. In this context, evidence is given that there is complementarity between search strategies based on missing energy and one based on jet multiplicity and S_T. A low-mass benchmark point is considered in some detail, and rough estimates of exclusion reach across the parameter space are given.

§IV C then presents prospects for the lepton-plus-many-jet searches with 1 fb^{-1} of data, accounting for increased p_T thresholds. A high-mass benchmark point is considered in some detail, and it is checked that sensitivity to the low-mass benchmark point is retained.

A.   The Fiducial Models

We consider a subspace of the MSSM, parametrized by the gluino pole mass $M_{\tilde{g}}$ and flavor-universal squark mass $M_{\tilde{q}}$. The bino and wino soft masses are given by $M_1 : M_2 :$

8

$M_{\tilde{g}} = 1 : 2 : 6$. The sleptons and Higgsino are heavy enough that they play no role.[4] This simplified model is loosely reminiscent of the CMSSM and has similar phenomenology.

The dominant new-physics processes are gluino pair-production, squark pair-production and squark-gluino production. When $M_{\tilde{q}} > M_{\tilde{g}}$, the gluinos decay predominantly via a two-step cascade through the wino or Higgsino,

$$\tilde{g} \to 2j + \widetilde{W} \to 2j + W^{\pm}/Z^0 + \widetilde{B} \,, \tag{2}$$

and a significant fraction of squarks decay into the gluino. Therefore, it is common to obtain more than six jets, a lepton from W decay, and (if the bino is stable) missing energy. The squarks also have a less active decay:

$$\tilde{q} \to j + \widetilde{W} \to j + W^{\pm}/Z^0 + \widetilde{B} \,, \tag{3}$$

which dominates when $M_{\tilde{q}} < M_{\tilde{g}}$. In this case, high-multiplicity states are not generic unless the bino undergoes further decay.

Three possibilities for the fate of the bino are explored here:

- A stable bino LSP, giving rise to missing energy.

- An NMSSM-like decay of the bino into a singlet S and its superpartner, the "singlino" $\tilde{S}$; the S then decays to $b\bar{b}$. ($M_S = 0.5 \cdot M_1$, $M_{\tilde{S}} = 0.25 \cdot M_1$ are taken throughout). Decays of Higgsino- or wino-like NLSPs to Z or h in gauge mediation [44], or of any neutral NLSP in a suitable hidden valley [45], can have similar kinematics. In each of these cases, the energy of the NLSP is split between jets and missing energy.

- Fully hadronic R-parity violation, in which the bino decays to three quarks. Only decays to light-flavor quarks will be considered here, though heavy-flavor-rich decays are also possible. The only missing energy in this scenario is from W and Z decays to neutrinos.

These three variants of the fiducial model will be compared in the studies presented below.

In addition, two specific points in parameter space will be used to illustrate the effectiveness of the search strategy:

$$\begin{aligned} \textbf{low-mass benchmark point:} \quad & (M_{\tilde{g}}, M_{\tilde{q}}) = (550, 800) \text{ GeV} \\ \textbf{high-mass benchmark point:} \quad & (M_{\tilde{g}}, M_{\tilde{q}}) = (700, 1000) \text{ GeV} \end{aligned} \tag{4}$$

The signal events are generated in `Pythia 6.4` [36]. Background simulation is discussed in §III and Appendix A 1. A detailed description of the detector mock-up can be found in Appendix A 2.

B. The 2010 data sample of 36 pb^{-1}

This section explores the effectiveness of the lepton-plus-many-jet search in 36 pb^{-1}. The three variants of the low-mass benchmark — stable bino, bino→ $S\tilde{S}$, and RPV bino

---

[4] Specifically, $\mu = M_{\tilde{g}} + 100$ GeV, while slepton masses of 1500 GeV are chosen for convenience.

decay — were already considered in Figs. 2 and 3. Figure 6 shows the integrated S_T and $\not{E}_T$ distributions for the three variants, compared to the matched top background generated with `MadGraph` and normalized (see §III) against data [32, 33]. For both plots (unlike Fig. 2), the p_T, m_T and $n_{jets} \geq 6$ cuts presented in §III are imposed. While the stable bino scenario produces a small excess on the $\not{E}_T$ tail, the two cases with a decaying bino have suppressed missing energy and are not distinguishable from background. In contrast, all three variants are clearly separated from the background in S_T.

Table I summarizes the estimated efficiencies at each stage of the lepton-plus-many-jet selection for each variant of the low-mass benchmark point. After imposing p_T requirements on the lepton and jets, a multiplicity requirement of six or more jets, and an S_T cut of 1000 GeV, there are 4.9, 6.3, and 7.7 signal events remaining for the stable $\widetilde{B}$, $\widetilde{B} \to S\widetilde{S}$, and $\widetilde{B}$ RPV decay variants, respectively, compared to 2.2 events for the top background. Even with no assumptions about the top background, the models with decaying binos can be easily excluded if only the expected 2 events are observed. For models where the excess over background is not as significant, a robust understanding of the top background must be obtained through a reliable data-driven model (or a hybrid Monte-Carlo-and-data-driven model). Issues relevant to this step are discussed in §III; see also §V.

Table I also summarizes the cut-efficiencies for two high-$\not{E}_T$ selections modeled after jets+$\not{E}_T$ analyses recently published by CMS (the high-H_T selection from [31]) and AT-LAS (selection 'D' from [4]). (Though we have attempted to mimic the selections in these references, we emphasize that our detector mock-up (described in Appendix A 2) is highly simplified, and the results are only approximate.) Even with a stable LSP, the lepton-plus-many-jet search has sensitivity only slightly weaker than the existing high-$\not{E}_T$ searches. And the lepton-plus-many-jet search gains sensitivity if the LSP decays, while the other search strategies rapidly lose their power.

Figure 7 generalizes these results to a broad range of gluino and squark masses. Rough estimates of the expected sensitivity of the lepton-plus-many-jet and existing high-$\not{E}_T$ searches at 36 pb^{-1} are shown as a function of $M_{\tilde{g}}$ and $M_{\tilde{q}}$ for the stable bino (left), bino $\to$ singlino (center), and RPV (right) variations. The orange solid (red dashed) line is the projection
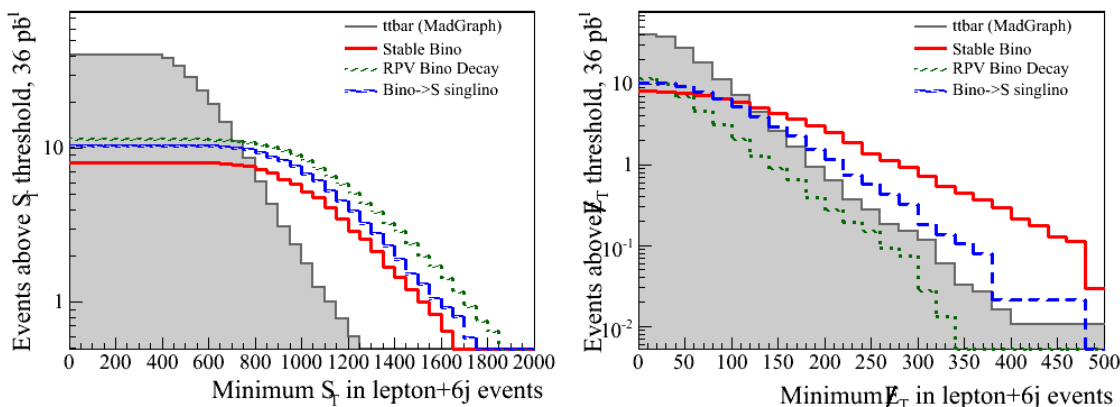

FIG. 6: Left: Number of events expected above an S_T cut (*i.e.*, the integrated upper tail of the S_T histogram in Fig. 3), after lepton p_T, m_T and 6-jet requirements from §III. Right: The integrated missing energy tail after the same selections. Both plots correspond to the low-mass benchmark.

	Stable $\widetilde{B}$	$\widetilde{B} \to S\,\widetilde{S}$	$\widetilde{B} \to 3j$ (RPV)	Background
Total Rate (36 pb^{-1})	86	86	86	
Lepton+6j				
1 lepton ($p_T > 20$)	15	15	13	*1400*
$m_T(\ell,\nu) > 20$	14	12	12	*1200*
≥ 4 jets ($p_T > 30$)	13	12	12	*460*
≥ 6 jets ($p_T > 30$)	7.5	9.7	11	*43*
$S_T > 1000$ GeV	4.9	6.3	7.7	*2.2*
CMS high-H_T				
Lepton veto	70	71	73	
≥ 3 jets ($p_T > 50$)	69	71	73	
$d\phi(j, \not{E}_T)$	54	51	51	
$H_T > 500$ GeV	50	50	50	
$\not{E}_T > 150$ GeV	25	9.1	1.8	**43.8 ± 9.2**
ATLAS jets+MET D				
3 jets, $p_T >$120, 40, 40	85	86	86	
Lepton veto	64	65	68	
$\not{E}_T > 100$	47	26	7.5	
$d\phi(j, \not{E}_T)$	33	16	4	
$\not{E}_T / M_{eff} > 0.25$	15	3.1	0.3	
$M_{eff} > 1000$ GeV	3.1	0.4	0.04	**2.5 ± 1.0$^{+1.0}_{-0.6}$ ± 0.2**

TABLE I: Total expected event rate in 36 pb^{-1} for the low-mass benchmark models discussed in §IV B, and the rates surviving several cut flows in Monte Carlo. The top section of the table displays the cut-flow efficiencies for the lepton-plus-many-jet strategy, with background rates in italics corresponding to $t\bar{t}$ only, as computed using `MadGraph` Monte Carlo. The lower two sections give estimated efficiencies for the same models, through cut-flows based on the hadronic jets+$\not{E}_T$ analyses done by CMS (the high-H_T selection used in [31]) and ATLAS (selection 'D' used in [4]). For the latter two cases, the (bold-faced) expected background is the full background prediction quoted in the respective search papers.

of the likely limits for the lepton-plus-many-jet search with $S_T > 600$ (1000) GeV. For the 600 (1000) GeV cut, the expected top background is 15 (2) events, and it is assumed that a signal of 12 (6) expected new-physics events passing the selection could be excluded. This exclusion threshold assumes that systematic errors on the $t\bar{t}$ background are small enough to be neglected compared to statistical errors. A significant systematic error would reduce the overall sensitivity of a lepton-plus-many-jet search, but would not alter our conclusion that the search is complementary to missing-energy-based searches.

For comparison, the estimated reach of jets+$\not{E}_T$ searches for these models is also shown in Fig. 7. The thick black lines on the plots are a rough estimate of a combined limit from the ATLAS and CMS jets-plus-$\not{E}_T$ searches, in particular [4] (region "D") and CMS [31] (high-H_T and high-$\not{H}_T$ regions). These estimates neglect any signal contamination of control regions, which could weaken the projections by up to ~50 GeV.

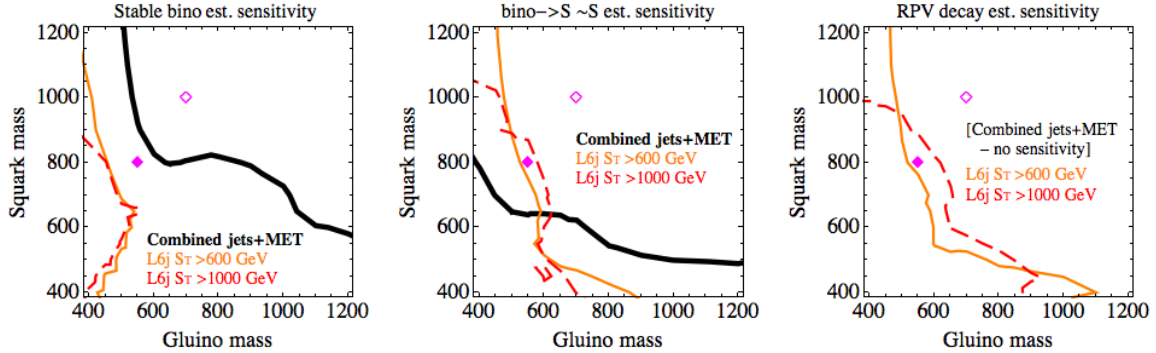The combination of jets+$\not{E}_T$ analyses gives the best coverage for the stable bino scenario,

FIG. 7: Estimated 36 pb^{-1} sensitivity of lepton-plus-many-jet searches, with S_T thresholds of 600 GeV (orange) and 1000 GeV (red dashed), for the fiducial models as a function of different squark and gluino masses. Each panel corresponds to a different LSP scenario: stable $\tilde{B}$ (left), $\tilde{B} \rightarrow S + \tilde{S}$ (middle), and hadronic RPV decay of $\tilde{B}$ (right). For comparison, we estimate the combined sensitivity of three ≈ 36 pb^{-1} hadronic SUSY searches in ATLAS [4] (region "D") and CMS [31] (high-H_T and high-$\not{H}_T$) (thick black). The low- and high-mass benchmark points are marked by solid and open magenta diamonds.

where the mass splitting between the gluino and bino is large and there is considerable missing energy. When the bino decays to a singlino, the jets+$\not{E}_T$ searches lose their reach; fewer events survive the strong $\not{E}_T$ cut, and limits can only be set when the gluino and squark masses are light enough to maintain a large production rate. In contrast, the lepton-plus-many-jet analysis has significant coverage for gluino masses less than 600 GeV and all squark masses up to 1200 GeV, and can reach up to $m_{\tilde{g}} \sim 900$ GeV for lower squark masses. With an RPV decay of the LSP, the missing energy arises only from the $W^{\pm}$ that produces the lepton, and the jets+$\not{E}_T$ search has essentially no sensitivity. The lepton-plus-many-jet analysis largely compensates for this loss.

B-tagging and anti-tagging are important handles for both classes of searches. Two ATLAS searches requiring b-tags [3], not shown, may extend the light-squark sensitivity for the singlino model by ~ 50 GeV. The use of b-tagging fractions in a lepton-plus-many-jet search is discussed in §V A.

C. Prospects for Searches in 2011 Data

In this section, the potential sensitivity of lepton-plus-many-jet searches in 1 fb^{-1} of 2011 data is briefly considered. The complementarity between lepton-plus-many-jet and jets+$\not{E}_T$ searches, illustrated at 36 pb^{-1} in §IV B, should persist at higher luminosities, but as there is as yet no data with which to normalize Monte Carlo estimates, this cannot be studied reliably. The discussion here is therefore limited to a semi-quantitative and preliminary examination of the two benchmark points.

LHC data-taking during the ongoing 2011 run is at much higher luminosities than in 2010, leading to higher trigger thresholds (on H_T, jet and lepton p_T, etc.) and significant pile-up. Consequently, a lepton-plus-many-jet search in 2011 data will presumably require tighter cuts than those used in §IV B. The following discussion therefore proceeds with lepton p_T

and m_T requirements raised from 20 up to 30 GeV, and with jet p_T threshold raised from 30 to 45 GeV.

Despite the higher thresholds, the separation of signal from background *improves* relative to what is seen in Table I. Under the new cuts, combined with a raised S_T cut of 1000 GeV, the efficiency for the top background is reduced by a factor of 3, while the efficiency of the low-mass benchmark signals, for all three variants, drops by no more than a factor of 2. The much greater integrated luminosity of 1 fb^{-1} easily compensates for the lower signal efficiency, and allows use of seven-jet events, as well as cross-checks in other distributions (see §V). Thus, higher luminosity will enhance sensitivity to lower-mass scenarios, despite the increased thresholds.

Of course, higher luminosity also allows the use of the lepton-plus-many-jet approach for much heavier gluino and squark masses than the low-mass benchmark. Figure 8 shows multiplicity and kinematic distributions for the three variants of the high-mass benchmark point, with $(M_{\tilde{g}}, M_{\tilde{q}}) = (700, 1000)$ GeV. With a decaying bino, the shapes of the signal-plus-background distributions, both in n_{jet} after an S_T cut and in S_T after an n_{jet} cut, are quite different from the top background. Furthermore, with the larger gluino and squark masses, the signal S_T distribution peaks at higher values, and is thus more easily separated from the $t\bar{t}$ background.

V. POTENTIAL CROSS-CHECKS ON BACKGROUNDS AND SIGNALS

In this short section, we consider several variables that could distinguish a new signal from a mismodeled background. Among these, the most powerful (though not for all signals) appears to be the b-tag multiplicity distribution, discussed in §V A. A number of other variables that are also worth considering are briefly mentioned in §V B.
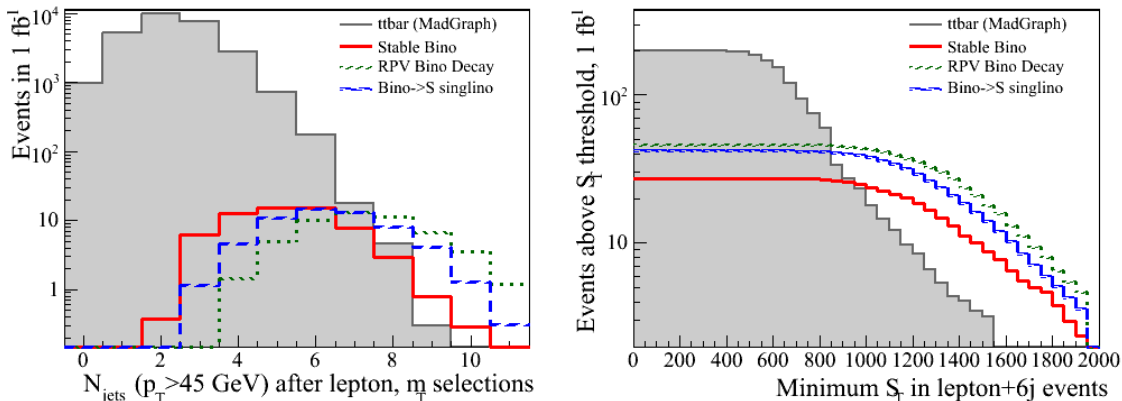

FIG. 8: Kinematic distributions in the signal region, for the *high-mass* benchmark models, with the tighter cuts (jet $p_T > 45$ GeV) and 1 fb^{-1} integrated luminosity: (left) $n_{jet}(p_T \geq 45$ GeV) after lepton and m_T cuts and (right) integrated number of events above an S_T cut, after additionally requiring at least 6 jets.

13

A.   Cross-checks using heavy-flavor-tagged jet multiplicities

Top quark samples have well-measured and well-understood b-tagging rates, arising from the ubiquitous b and $\bar{b}$ and the presence of a c quark in half the mixed leptonic-hadronic events. A distortion in the expected ratios of tag-multiplicities is evidence for a signal. Of course, some signals, such as those that typically have a $t\bar{t}$ or $b\bar{b}$ pair in every event, have similar b-content to background and show little distortion. In this case, other variables, such as those mentioned in §V B, may be needed.

Estimates of the b-tagging multiplicity distributions for the three variants of our high-mass benchmark model are shown in Fig. 9, for 1 fb^{-1} integrated luminosity and a minimum jet p_T of 45 GeV. A similar estimate for the top background is shown. These numbers are obtained very naively, by requiring that the b jet be in the tracking volume and assuming a 60% tagging efficiency. Charm tags and mistags are not accounted for, but should not be important for signal. For background, results such as Figures 1d and 2d in the auxiliary plots of [46] suggest that charm tagging would contribute of order one event to the three-tag bin. Other backgrounds such as $t\bar{t}b\bar{b}$, $t\bar{t}h$, etc., appear negligible. The different variant models exhibit significant differences in b-tagging multiplicities. The stable-LSP and RPV models, which mainly produce light quark jets, give a large contribution to the 0-tag bin relative to the 1-tag bin. The singlino model, with its four extra b quarks per event, has a significant number of events with more than one b tag. In the following section, we will see other examples of models that give even larger excesses in the 3-tag bin. Clearly, distributions of b-tag multiplicities can be very helpful in separating an excess from signal from a mismodeled top background.

One challenge for this method is that at high S_T some of the jets can be quite hard, and tagging fractions become both smaller and less precisely known. Fortunately it may be possible to use data to determine the tagging fractions in the top background. The extra jets in $t\bar{t}$ plus two jets are rarely b or c jets, and so the tagging fractions in $t\bar{t}$ plus zero, one and two jets may be quite similar. However, this suggestion has not been checked in Monte Carlo.
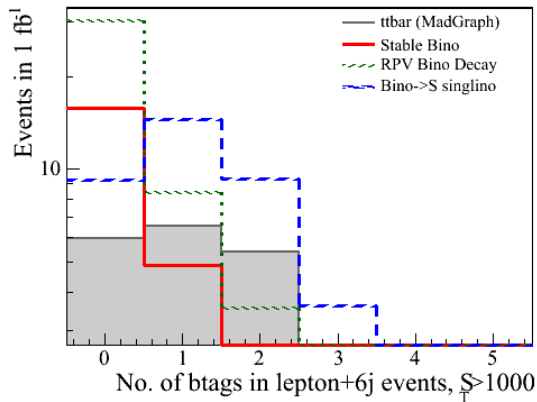

FIG. 9:   Estimate of b-tag multiplicities for `MadGraph` $t\bar{t}$ (gray shaded, not including c tags) and the three high-mass benchmark models: (red solid) stable LSP, (green dotted) LSP decay through singlino and singlet, (blue dashed) RPV decay of LSP. We assume 'naive' tagging fractions, as described in the text.

B. Other Signal Cross-Checks

There are a number of other features that can make a sample with signal-plus-background look different from a background that is mainly from $t\bar{t}$-plus-jets. These include the following:

- The jet multiplicity distribution, which for $n_{jets} \geq 6$ falls off gradually in the top background, but can be nearly flat in the presence of signal from 6 out to 7 and 8 jets.

- The η distribution, event-by-event, of the leading six jets. In signal, the jets are more tightly clustered around their average η than in background, where one or two of the jets tend to be from ISR.

- The total integrated lepton-charge asymmetry, which is small in $t\bar{t}$ (perhaps of order a few percent) but can be as large as $2:1$ in some signals, for instance one dominated by associated squark-gluino production.

- The efficiency for reconstructing top quarks, which can be greatly reduced for signals in which top quarks are rare.

- Angular variables that are characteristically correlated with one another in top quark backgrounds. For instance, the observables ΔR between the two leading jets, ΔR between the leading jet and the lepton, and $\Delta \phi$ between the lepton and the $\not{E}_T$ vector show strong correlations in background. These correlations are often reduced in signals.

Note that some of these variables, in particular the first and second, are also useful in rejecting W-plus-jets backgrounds.

For our high-mass benchmarks, 1 fb^{-1} may be too small for these features to be significant individually, though the first two seem promising even at rather low statistics. With several times more data, all of these observables should become useful. Combining these variables might also prove powerful, though this has not been studied here.

VI. EXAMPLES OF OTHER MODELS

Having considered a particular class of fiducial models in detail, we now check that the lepton-plus-many-jet search is sensitive to a wide range of signals. We have studied a number of different classes of models. For some classes, the results are similar to those in the fiducial models, while for others some new features arise, particularly in the context of b-tagging.

A. t-rich SUSY Models

Many SUSY models are a large source for top quarks, which naturally result in high-jet-multiplicity signals. We have studied one class of examples, with relatively light stops and sbottoms, a bino LSP, and a gluino somewhat heavier than the stops and sbottoms but lighter than the other squarks. Although such models often have large cross-sections for electroweak gaugino production, much of the relevant cross-section for high-multiplicity events comes from associated squark-gluino production, with gluino and squark pair production also playing a role. As before, we have chosen three variants of these models, with
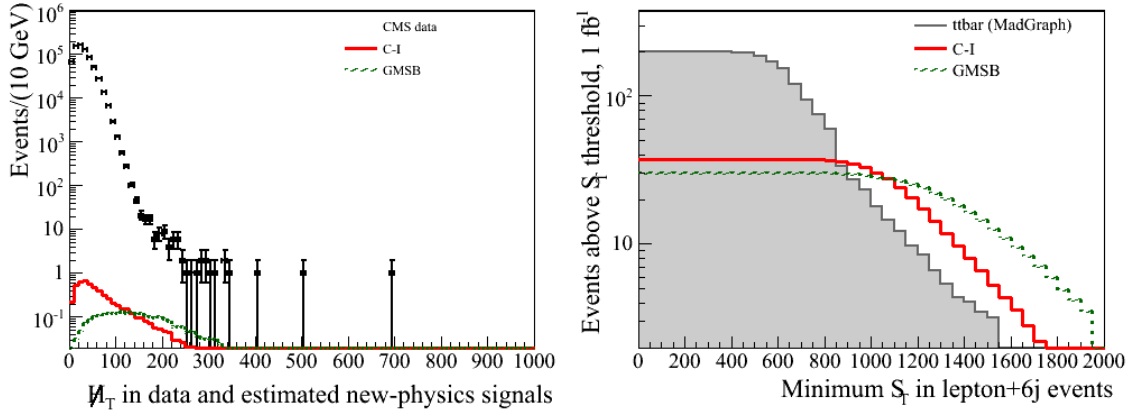
FIG. 10: Distributions for the gauge-mediated model (green dashed) and the compositeness-inspired model (red solid) discussed in §VI B and §VI C respectively. Left: Missing energy distributions, as in Fig. 2 (36 pb^{-1}). Right: Number of events expected above an S_T cut (i.e. integrated upper tail of S_T histogram) after lepton p_T, m_T and 6-jet requirements, as in Fig. 8 (1 fb^{-1}).

a stable bino, a bino decaying to a singlet and singlino, and a bino decaying to light quarks via R-parity violation.

Almost all of our plots (omitted for brevity) are qualitatively the same as those for the fiducial models shown in §IV B and §IV C, and our conclusions remain the same. The one exception is in b-tagging, which, because of the presence of top and bottom quarks in many of the events, is shifted to larger tag multiplicity relative to the models shown in Fig. 9. In the following, more examples of theories with an excess of ≥ 3 tagged jets will be given.

B. GMSB with Z or H decays

In gauge-mediated SUSY breaking (GMSB), the lightest SM superpartner is the NLSP and decays to a gravitino, the true LSP. Such models can also give high-multiplicity signals with reduced $\slashed{E}_T$. Here, we consider the case of a Higgsino NLSP decaying [44] mainly to a Z or h plus a gravitino, with a very small branching fraction to decay to a photon plus gravitino. Similar models, possibly with even higher multiplicity, can arise in the presence of a Hidden Valley sector [45].

As an illustration, we choose a gauge-mediated model with a Higgs boson at 120 GeV and Higgsinos $\tilde{h}$ at 198–208 GeV. The Higgsino NLSP has branching fractions of 78%, 20% and 2% to h, Z and γ plus a gravitino. We consider a model in the class of General Gauge Mediation (or, more simply, with a non-minimal set of messenger fields for which colored and colorless messengers are not mass-degenerate). In this model, the gluino is at 800 GeV, the bino at 404 GeV and the wino at 856 GeV; the squarks are at 1100 GeV. Sleptons are near 500 GeV, too heavy to play any role. Dominant production modes involve gluino pairs, with $\tilde{g} \to t\bar{b}\tilde{h}^-, \bar{t}b\tilde{h}^+, t\bar{t}\tilde{h}^0$, with a minority of gluinos decaying to light quark pairs plus a Higgsino. The large number of b quarks in the final state is further enhanced when the Higgsino decays to a Higgs boson.

The model produces numerous same-sign di-lepton events, but we estimate that the number of events in 1 fb^{-1} that pass the stringent requirements of the 2010 CMS search [8]
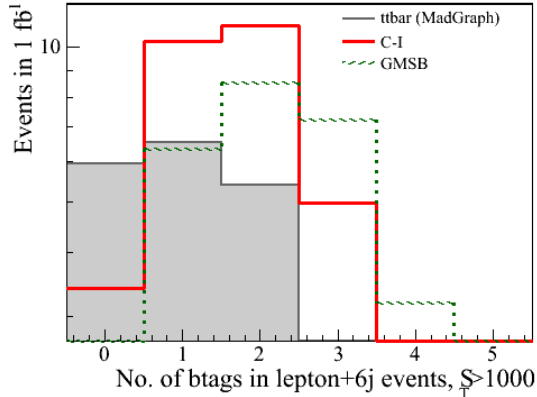
FIG. 11: As in Fig. 9, estimated b-tag multiplicities at 1 fb^{-1} for `MadGraph` $t\bar{t}$ (gray shaded, not including charm tags) and for the gauge-mediated model (green dashed) and the compositeness-inspired model (red solid) discussed in §VI B and §VI C respectively.

is still rather small, making our current search at worst complementary. In fact, because trigger and analysis thresholds for the same-sign search must presumably be raised for 2011, our search strategy might prove more sensitive.

The integrated S_T distribution for this model is presented in Fig. 10 and demonstrates that this model well exceeds the top background using our methods. The $\not{E}_T$ distribution, also shown in Fig. 10 (where the background, as in Fig. 2, is the data taken from [31]), confirms that this model is not easily seen using the simplest jet-plus-$\not{E}_T$ searches. Meanwhile, the b-tag distribution, shown in Fig. 11, demonstrates that the excess is quite different from the expected top background, with over a third of the excess events carrying three or more b tags.

C. Non-SUSY models with top quarks plus other particles

Certain strong-dynamics BSM models [25] may contain new composite fermions ψ in higher-representations of $SU(3)$, such as sextets or octets of color. Other models may have multiple triplets. These fermions may couple most strongly to the third generation, and may largely decay to a top quark plus an exotic scalar or pseudoscalar ϕ. The field ϕ may be an octet (triplet) [singlet] of color if ψ is a sextet or triplet (octet) [triplet]. This scalar may then in turn decay to gluons or to heavy-quark pairs. Alternatively, this scalar may be too heavy to be produced on-shell, in which case the decay of ψ may be a three-body decay to a top quark plus two other hard partons.

Another class of models with similar phenomenology can arise in the context of R-parity-violating SUSY. A gluino (playing the role of ψ) may decay to the top squark $\tilde{t}$ (playing the role of ϕ) which then decays $\tilde{t} \to \bar{b}\bar{s}$. Alternatively, the gluino may decay directly to a triplet tbs of quarks, as a three-body decay.

For technical reasons, we have chosen to display the RPV SUSY model with $\tilde{g} \to \tilde{t}\bar{t}$, followed by $\tilde{t} \to bs$, though again we emphasize that there is nothing especially supersymmetric about this signature. The other models are almost identical in most kinematic distributions, except for normalization, on which we comment below. The integrated S_T dis-

tribution, and the $\not{E}_T$ distribution, are shown in Fig. 10, while the b-tagged jet multiplicity for this particular model (with 4 b quarks in each event) is shown in Fig. 11.

In comparing the particular model simulated here to the non-SUSY compositeness-inspired models mentioned in the first paragraph of this section, there are a few things to keep in mind. First, Dirac octet fermions, as complex representations, have double the cross-section of a Majorana gluino at the same mass scale. Second, sextet fermions and Dirac octet fermions have almost identical cross-sections and kinematics in their decays. Third, although Majorana gluinos will also produce a same-sign lepton signal, sextets or Dirac octets, being complex, will not. Thus, while the Majorana gluino case shown in the figures is somewhat borderline for detection using our method, the sextet and Dirac octet would be more easily detectable, and invisible to the same-sign di-lepton search.

Meanwhile, triplets, such as a t', have a cross-section several times smaller than the gluino. Our methods would need further optimization to detect one such quark, for a wide range of masses. However, a model with several roughly-degenerate quarks should be observable using our proposed analysis.

Finally, while the RPV SUSY model in our plots produces four bottom quarks in each event, which will further distinguish it from backgrounds, the non-SUSY models mentioned above might produce as few as two and as many as six, depending on the decay mode of the scalar ϕ. Consequently, b-tagging is very useful as a cross-check in some models of this type, though not in all.

D.   Departures from Perturbative Quantum Field Theory

The lepton-plus-many-jet strategy is quite inclusive, so it is also potentially sensitive to new physics that is not well understood theoretically. For example, a breakdown of perturbative quantum field theory that leads to large partonic cross-sections would potentially be detectable through such a search. In particular, models that have partonic cross-sections that rise at high S_T, but less dramatically than black holes, might not be detectable without requring a lepton in order to reduce QCD backgrounds (see Fig. 1). While we have not studied specific models, it appears that a lepton-plus-many-jet search would fill a gap between standard black hole searches and other exotica and SUSY searches.

VII.   DISCUSSION AND CONCLUSIONS

In this work, we have argued that a search requiring high jet multiplicity and high S_T, along with a lepton, and with a limited m_T requirement (for the purpose of reducing fake leptons), has sensitivity to phenomena that are currently not being covered by existing $\not{E}_T$-based SUSY searches or by any published exotica searches. A key observation is that the background is apparently dominated by top-quark backgrounds, not by W-plus-jets. Checking our work where possible against existing public results, we have argued (see Fig. 7) that this strategy, applied to the 2010 data, would be sensitive to large classes of models that are not excluded by the existing searches. We have further argued that this strategy still works with cuts raised somewhat to account for higher trigger thresholds and pile-up in the 2011 data. Furthermore, this search is sensitive to many different types of high-cross-section physics, including a large variety of SUSY models and compositeness models. We hope that searches along these lines are already underway, or will soon be undertaken.

18

Before focusing on the lepton-plus-many-jet sample, we argued more generally that strategies based on high-multiplicity, high-S_T, and a non-jet visible object could be powerful. It is natural to consider replacing the lepton in our search with a photon or Z, or to consider something a bit more elaborate, such as same-flavor opposite-sign leptons off the Z peak (non-Z di-leptons, or "NZDL".) The NZDL case seems most likely to extend directly from our current discussion. In some kinematic regimes, the background is again dominated by top, and is relatively simple to model. For signals that dominantly produce leptons in pairs, through an off-shell Z or in a cascade decay with a slepton-like intermediate state, a strategy requiring an NZDL pair and several jets (perhaps as few as four) might do better than the single-lepton-plus-six-jet strategy we have described in this paper.

Searches involving a Z boson, on the other hand, may have either Z-plus-many-jet or $t\bar{t}$ as a background, depending on cuts. In the latter case, the extra Z bosons themselves are a signal; in the former case, however, a strategy such as ours might apply. The challenge here is to determine the background, perhaps using photon-plus-many-jet to estimate the Z-plus-many-jet distributions.

For a search in the photon-plus-many-jet channel itself, however, background modeling may be difficult, as there is no obvious data-driven method. The strategy suggested in [34], of looking at high jet-multiplicity for an excess of photons or Z bosons at central rapidity, would apply, but it would be interesting to seek an alternate approach.

More generally, we strongly encourage the experimental groups, whenever practical, to make public the S_T distributions for their samples, including Standard Model measurements and control regions. Although we recognize that validating such plots is by no means simple, they are of great physical interest. Among the various possible kinematic variables, it appears that S_T is especially robust and informative. Distributions in S_T would assist theorists in confirming their estimates of backgrounds, and estimating whether high-mass signals of interest are within reach of existing data. It may even be possible to exclude some models with large cross-sections, based only on the overall high-S_T rate without a dedicated search.

Clearly, a diverse array of LHC searches is necessary so that new physics can be found wherever it may be, and to assure maximal use is made of LHC data. To develop a sufficiently broad search program requires identifying many regions of phase-space where Standard Model backgrounds are small and plausible new-physics signals could be present at detectable levels. We have studied one such region (and suggested a few others) in this paper, and argued that reasonable signals may be large where the backgrounds are small. Meanwhile, if one is to draw any general scientific lesson (such as the absence of superpartners below some mass scale) from a collection of null results, it is crucial that this set of searches be robust to modifications of the theory in question. We have argued that combining the lepton-plus-many-jet search (along with other similar searches) with existing $\not{E}_T$-based searches significantly improves the robustness of the set.

For any study of this type, Standard Model backgrounds are the central concern, and thus careful measurements of the Standard Model at the LHC are the key to the effort. This is especially true when the leading backgrounds are difficult to calculate reliably. For us, the W- and Z-plus-jets measurements at ATLAS and CMS [32, 33] were invaluable, both in motivating our study of the lepton-plus-many-jets strategy and in allowing us to think through certain subtleties that such an analysis would encounter. We expect that the growing body of LHC Standard Model measurements will suggest many new avenues for future searches.
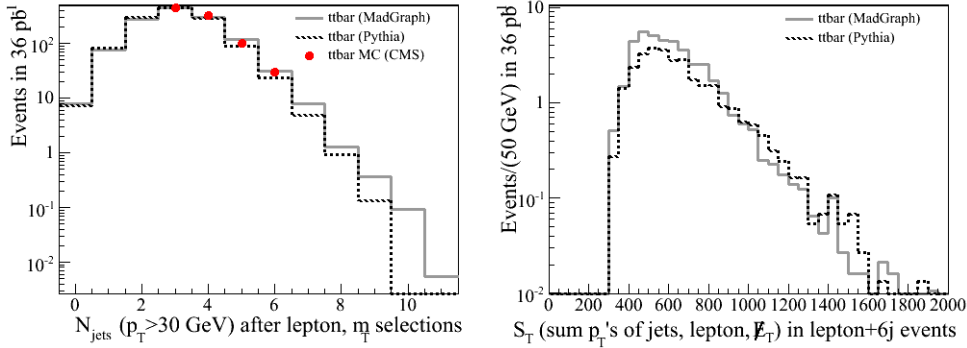
FIG. 12: Left: Our computation of the jet multiplicity distribution for matched `MadGraph` $t\bar{t}$ (gray) and Pythia $t\bar{t}$ (black dotted) Monte Carlo samples, for electrons and muons combined. The red dots indicate the CMS Monte Carlo (consistent with 36 pb^{-1} data) in each bin, as shown in Fig. 4. Right: Comparison of S_T distributions for the matched `MadGraph` and unmatched `Pythia` $t\bar{t}$ samples.

Appendix A: Monte Carlo and Detector Mock-Up

1. Modeling the $t\bar{t}$ plus jets background

A matched $t\bar{t}$ background was generated with `MadGraph 4.4.49` [35] with CTEQ6L1 parton distributions [47]. To guarantee that the high S_T tail of the top distribution is well-populated, 50K events were generated in each of three regimes, where the top with the highest p_T in the event was required to either be less than 100 GeV, between 100 - 300 GeV, or greater than 300 GeV. `Pythia 6.4.22` [36] was used for parton showering and hadronization. An MLM matching procedure was implemented using `MadGraph` and `Pythia` in combination with a shower-$k_\perp$ scheme introduced in [38–40]. The matching scale was $Q_{\mathrm{Match}} = 100$ GeV for the top sample. The results were then passed through a private detector mock-up as described below.

Figure 12 (left) compares our matched $t\bar{t}$ sample with the Monte Carlo expectation given by CMS in [32]. Our sample is normalized to the top NLO cross-section of 150 pb, giving a K factor of 1.64. The three and four jet channels agree perfectly, but our Monte Carlo over-predicts the five and six-jet channel by approximately 20 and 30%, respectively.

We also generated an unmatched top sample using `Pythia`. Figure 12 (right) shows the S_T distributions for the `Pythia` and `MadGraph` samples. `MadGraph` with matching gives more six-jet events than does `Pythia`; this is expected, as the matching procedure includes two additional ISR jets in the matrix element, as compared to `Pythia`, which only includes one by default. The p_T distributions for the hardest four jets in each Monte Carlo sample correspond well with each other, leading to good agreement on the high-S_T tail in Fig. 12. Meanwhile, although the p_T distributions of the fifth and sixth jet differ between the matched and unmatched samples, this has a significant effect only on the low-S_T bins of Fig. 12.

2. Reconstruction and Analysis/Detector Mock-Up

We employ a simple detector mock-up that allows flexibility in tuning the simulated detector response and reconstruction efficiency for any given analysis.

Starting from hadronized Monte Carlo truth, we build jet objects with `FastJet` [43], using anti-k_T with $\Delta R = 0.5$. All hadrons with $|\eta| < 3.0$ and leptons (electrons and muons) and photons with $|\eta| < 2.5$ are included in the jet reconstruction. A two-dimensional missing energy vector is constructed as $-\sum_i \vec{p}_T(i)$, where i runs over hadrons, leptons and photons within the same η ranges (the resulting missing energy is comparable to both truth- and jet-level missing energies constructed from the same objects). Prompt leptons and those from tau decay are treated as lepton candidates, provided they survive isolation. Jets that match a lepton within $\Delta R < 0.2$ and carry less than twice the lepton p_T are discarded. Following this, we apply a naive geometric isolation requirement to leptons, so that leptons within $\Delta R < 0.4$ of higher-p_T jets are rejected. Finally, we apply a parametrized identification and reconstruction efficiency to lepton objects. Electrons and muons are treated alike, with a parametrized efficiency that reflects their average. Similarly, b-tags are applied by first matching each b-parton in an event with $|\eta| < 3$ to its nearest jet, then applying a parametrized efficiency.

For comparison, we also use PGS [48] to establish a rough estimate of systematic errors involved in mocking-up a given set of object-level selections.

The procedure above is clearly not a faithful reproduction of the actual reconstruction performed by LHC experiments, nor is it intended to be. It has been designed to model new-physics signals and physics backgrounds simply. For example, no attempt is made to model resolution-dependent tails, fakes, or other instrumental effects. We have compared signal efficiencies obtained with this mockup with those reported in several new-physics searches at 36 pb^{-1} at ATLAS and CMS, and in most cases they are compatible (after combining e and μ channels) within about 10%.

[1] ATLAS, G. Aad *et al.*, Eur. Phys. J. **C71**, 1682 (2011), 1103.6214.

[2] ATLAS, T. A. Collaboration, Eur. Phys. J. **C71**, 1647 (2011), 1103.6208.

[3] ATLAS, G. Aad *et al.*, Phys. Lett. **B701**, 398 (2011), 1103.4344.

[4] Atlas, J. B. G. da Costa *et al.*, Phys. Lett. **B701**, 186 (2011), 1102.5290.

[5] Atlas, G. Aad *et al.*, Phys. Rev. Lett. **106**, 131802 (2011), 1102.2357.

[6] CMS Collaboration, V. Khachatryan *et al.*, Phys.Lett. **B698**, 196 (2011), 1101.1628.

[7] CMS, S. Chatrchyan *et al.*, JHEP **06**, 026 (2011), 1103.1348.

[8] CMS, S. Chatrchyan *et al.*, JHEP **06**, 077 (2011), 1104.3168.

[9] CMS, S. Chatrchyan *et al.*, JHEP **06**, 093 (2011), 1105.3152.

[10] ATLAS Collaboration, ATLAS CONF 2011 068 .

[11] ATLAS Collaboration, ATLAS CONF 2010 088 .

[12] CMS Collaboration, V. Khachatryan *et al.*, Phys.Lett. **B697**, 434 (2011), 1012.3375.

[13] CDF, T. Aaltonen *et al.*, Phys. Rev. **D76**, 072006 (2007), 0706.3264.

[14] D0 Collaboration, B. Abbott *et al.*, Phys.Rev. **D64**, 012004 (2001), hep-ex/0011067.

[15] D0 Collaboration, B. Abbott *et al.*, Phys.Rev.Lett. **86**, 3712 (2001), hep-ex/0011071.

[16] CDF Collaboration, T. Aaltonen *et al.*, Phys.Rev. **D78**, 012002 (2008), 0712.1311.

[17] CDF Collaboration, T. Aaltonen *et al.*, Phys. Rev. **D79**, 011101 (2009), 0809.3781.

[18] CDF Collaboration, T. Aaltonen *et al.*, Phys.Rev.Lett. (2007), 0712.2534.

[19] CMS Collaboration, CMS PAS EXO-08-005 .

[20] CMS Collaboration, CMS PAS EXO-10-021 .

[21] CERN Report No. ATLAS-CONF-2011-070, 2011 (unpublished).

[22] ATLAS Collaboration, ATLAS CONF 2011 065 .

[23] CMS Collaboration, CMS PAS EXO 10 024 .

[24] T. Appelquist, H.-C. Cheng, and B. A. Dobrescu, Phys. Rev. **D64**, 035002 (2001), hep-ph/0012100.

[25] E. Katz, private communication.

[26] T. Gregoire and E. Katz, JHEP **0812**, 084 (2008), 0801.4799.

[27] N. Arkani-Hamed, A. G. Cohen, and H. Georgi, Phys. Lett. **B513**, 232 (2001), hep-ph/0105239.

[28] N. Arkani-Hamed *et al.*, JHEP **0208**, 021 (2002), hep-ph/0206020.

[29] E. Katz, J.-y. Lee, A. E. Nelson, and D. G. Walker, JHEP **0510**, 088 (2005), hep-ph/0312287.

[30] C. T. Hill and R. J. Hill, Phys.Rev. **D76**, 115014 (2007), 0705.0697.

[31] CMS, S. Chatrchyan *et al.*, (2011), 1106.4503.

[32] CMS Collaboration, CMS PAS EWK-10-012 .

[33] ATLAS Collaboration, G. Aad *et al.*, Phys.Lett. **B698**, 325 (2011), 1012.5382, auxiliary plots at `https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/WPlusJets_01`.

[34] V. Pavlunin and D. Stuart, Phys.Rev. **D78**, 035012 (2008), 0806.2338.

[35] J. Alwall *et al.*, JHEP **0709**, 028 (2007), 0706.2334.

[36] T. Sjostrand, S. Mrenna, and P. Z. Skands, JHEP **05**, 026 (2006), hep-ph/0603175.

[37] M. Mangano, Fermilab Monte Carlo Workshop, Oct. 2002 (unpublished).

[38] T. Plehn, D. Rainwater, and P. Z. Skands, Phys. Lett. **B645**, 217 (2007), hep-ph/0510144.

[39] A. Papaefstathiou and B. Webber, JHEP **06**, 069 (2009), 0903.2013.

[40] J. Alwall, K. Hiramatsu, M. M. Nojiri, and Y. Shimizu, Phys. Rev. Lett. **103**, 151802 (2009), 0905.1201.

[41] CMS Collaboration, CMS PAS TOP-11-001 .

[42] ATLAS Collaboration, ATLAS CONF 2011 034 .

[43] M. Cacciari, G. P. Salam, and G. Soyez, JHEP **04**, 063 (2008), 0802.1189.

[44] K. T. Matchev and S. D. Thomas, Phys.Rev. **D62**, 077702 (2000), hep-ph/9908482.

[45] M. J. Strassler, (2006), hep-ph/0607160.

[46] Atlas Collaboration, G. Aad *et al.*, Phys.Rev.Lett. **106**, 131802 (2011), 1102.2357, auxiliary plots at `https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/susy-1lepton_01`.

[47] J. Pumplin *et al.*, JHEP **07**, 012 (2002), hep-ph/0201195.

[48] J. Conway *et al.*, `http://physics.ucdavis.edu/~conway/research/software/pgs/pgs4-general.htm`.