Tagging new physics with charm

S. Iwamoto, G. Lee, Y. Shadmi and Y. Weiss

Physics Department, Technion — Israel Institute of Technology,
Haifa 32000, Israel

E-mail: sho@physics.technion.ac.il, leeg@physics.technion.ac.il,
yshadmi@physics.technion.ac.il, yanivwe@tx.technion.ac.il

ABSTRACT: We propose a new variable, the charm fraction, for collider searches for new physics. We analyze this variable in the context of searches for simplified supersymmetry models with squarks, the gluino, and the bino, assuming that only the lightest mass-degenerate squarks can be produced at the high-luminosity LHC. The charm fraction complements event counting and kinematic information, increasing the sensitivity of the searches for models with heavy gluinos, for which squark production is flavor-blind. If squarks are discovered at the LHC, this variable can help discriminate between different underlying models. In particular, with improved charm tagging, the charm fraction can provide information on the gluino mass, and in some scenarios, on whether this mass is within the reach of a future 100 TeV hadron collider.

Contents

1   Introduction

Charm tagging at the LHC is a topic of intense study [1, 2], with future advances expected with the implementation and improvement of machine learning algorithms. It has recently been added to the menu of supersymmetry searches [3–5], along with the much more mature bottom tagging, leading to improved limits on the masses of charm squarks [6]. In the context of searches for physics Beyond the Standard Model (BSM), there is a fundamental difference between bottom and charm tagging. The third generation is expected to play a special role in extensions of the Standard Model (SM): light top partners are motivated by naturalness, and at the same time, the masses of top and bottom partners may be significantly affected by their large Yukawa couplings. In contrast, because of the smallness of the relevant Yukawa couplings, BSM physics may plausibly be first- and second-generation flavor blind. BSM processes at the LHC may then produce similar amounts of first- and second-generation quarks, whereas SM processes are dominated by first-generation quarks. The fraction of charm quarks in BSM candidate events can thus be a useful discriminator between new physics and the SM, and between different BSM models.

To examine this, we study the charm fraction in the production of mass-degenerate squark pairs. We use simplified models, containing only the gluino, some subset of the first- and second-generation squarks, and a bino as the lightest supersymmetric particle (LSP). These models are the "bread and butter" of supersymmetry searches, predicting two or more hard jets and large missing energy ($\not{E}_{\mathrm{T}}$). Generically, the cross section for squark-pair production drops steeply with the gluino mass. Thus, these searches become more challenging for heavy gluinos. However, with 8-fold squark degeneracy, squark production becomes more flavor-blind as the gluino mass increases. For a decoupled gluino,

pairs of charm squarks constitute 25% of the supersymmetry sample. Measuring this fraction can thus increase the sensitivity of LHC squark searches to scenarios with very heavy gluinos, which are challenging due to their smaller production cross sections. As the gluino mass decreases, t-channel gluino-exchange diagrams with quarks in the initial state become increasingly important, and the fraction of charm quarks from pair-produced squarks goes down.

We therefore assume that the gluino is beyond the discovery reach of the high-luminosity LHC (HL-LHC), and study the charm fraction in squark pair production. If an excess is observed in jets plus missing energy searches, significant effort would be required in order to determine the multiplicity, the SU(3) charge, the mass, and the spin of the particles produced. In addition, a key question is whether additional particles with masses beyond the LHC reach exist. The simplified models we study here are characterized by four parameters: the number of squark flavors produced, the gluino mass, the squark mass, and the bino LSP mass. The first three determine the BSM production cross section, while the latter two — and in particular, their difference — determine the event kinematics and consequently the efficiency of the search. As is well known, measurements of various kinematic observables such as the effective mass m_{eff} and the stransverse mass m_{T2} can be used to extract some information on the squark and bino masses [7]. The charm fraction, which is a qualitatively different observable, can yield new information on the underlying model, and in particular on the gluino mass. The latter is important input for the planning of future accelerators like the 100 TeV proton-proton collider, which is anticipated to be sensitive to gluino masses up to 10–15 TeV [8, 9].

While we use squark production as a concrete example, we believe it is important to approach LHC searches with as few theory biases as possible. Thus for example, the new states produced could be colored Kaluza-Klein fermions, and the "squarks" should be thought of merely as new fundamental colored scalars. The charm fraction may help determine whether these new produced states are the end of the story.

This paper is organized as follows. In section 2, we specify the simplified models we use, review the basics of squark searches, describe the Monte Carlo numerical analysis, and expand on the treatment of charm tagging in our analysis. In section 3, we proceed to study the charm fraction in the models and discuss the results. We end with some remarks in section 4.

2   General framework: models and overview of searches

In order to demonstrate the use of the charm fraction, we consider simplified models, consisting of the first- and second-generation squarks, the gluino, and a bino LSP. We assume that the squark spectrum is flavor-blind: squarks of the same gauge quantum numbers are mass degenerate, while some hierarchies may exist between left-handed and right-handed squarks, and/or between up- and down-type squarks. We imagine a scenario in which the gluino is beyond the reach of the 14 TeV LHC, with mass above 4 TeV [10], and assume that only the lightest squarks can be directly produced. Note that the latter assumption requires only mild hierarchies among the squark masses.

$N_{\tilde{q}}$	$m_{\tilde{q}}$ [GeV]	$m_{\tilde{\chi}_1^0}$ [GeV]	$m_{\tilde{g}}$ [TeV]
8 ($\tilde{u}_{\mathrm{L,R}}, \tilde{c}_{\mathrm{L,R}}, \tilde{d}_{\mathrm{L,R}}, \tilde{s}_{\mathrm{L,R}}$)	1600	500	4, 5, 6, 7.5, 10, 13, 450
	1600	300	
	1500	300	
	1500	1	
4 ($\tilde{u}_{\mathrm{R}}, \tilde{c}_{\mathrm{R}}, \tilde{d}_{\mathrm{R}}, \tilde{s}_{\mathrm{R}}$)	1600	300	
	1400	1	
2 ($\tilde{u}_{\mathrm{R}}, \tilde{c}_{\mathrm{R}}$)	1600	300	
	1500	300	
	1500	1	
	1400	1	

Table 1. Simplified supersymmetry models used in this paper.

Since the gluino mass affects the production of squark pairs, the models are characterized by the squark mass $m_{\tilde{q}}$, the bino mass $m_{\tilde{\chi}_1^0}$, the gluino mass $m_{\tilde{g}}$, and the number of squark flavors produced, which we denote by $N_{\tilde{q}}$. We consider three scenarios:

- $N_{\tilde{q}} = 8$ models, in which all the left- and right-handed squarks are degenerate;

- $N_{\tilde{q}} = 4$ models, in which only the right-handed squarks can be directly produced (with the left-handed squarks beyond LHC reach);

- $N_{\tilde{q}} = 2$ models, in which only the right-handed up-type squarks, $\tilde{u}_{\mathrm{R}}$ and $\tilde{c}_{\mathrm{R}}$, can be directly produced (with all remaining squarks beyond LHC reach).

The parameters of the different models are summarized in table 1.

2.1 Search basics

The simplified models we consider predict squark-pair production at the LHC, yielding events with at least two hard jets, large missing energy, and no electron or muon. Our analysis below closely follows ATLAS analyses of this topology. As we focus on the HL-LHC with $\sqrt{s} = 14\,\mathrm{TeV}$ and an integrated luminosity $\int \mathcal{L} = 3000\,\mathrm{fb}^{-1}$, we employ the Meff-2j-3100 signal region (SR) of ref. [10], which discusses HL-LHC reaches based on this event topology. This SR selects events with two or more jets, missing energy above 160 GeV, and inclusive effective mass above 3100 GeV. Some of the model points we will display in the discussion of the charm fraction are already excluded by 13 TeV data. In order to determine whether a model is already excluded, we use the Meff-2j-2000 SR of ref. [11], which is based on 13 TeV LHC data with $\int \mathcal{L} = 13.3\,\mathrm{fb}^{-1}$. The full sets of cuts defining both SRs are reviewed in table 2.

2.2 SM backgrounds and charm production

The main SM background for the two-jets plus missing energy search is Z + jets with the Z boson decaying into neutrinos (see figure 1(a) and figure 1(b)). The next source of

		Meff-2j-2000	Meff-2j-3100
Number of jets, electrons, muons		$\geq 2, = 0, = 0$	
$\not{E}_T$ [GeV]	$>$	250	160
$p_T(j_1), p_T(j_2)$ [GeV]	$>$	250, 250	160, 60
$|\eta(j_1, j_2)|$	$<$	1.2	—
$\Delta\phi(j_{1,2,(3)}, \not{E}_T)_{\min}$	$>$	0.8	0.4
$\Delta\phi(j_{i>3}, \not{E}_T)_{\min}$	$>$	0.4	0.2
$\not{E}_T/\sqrt{H_T}$ [GeV$^{1/2}$]	$>$	20	15
$m_{\rm eff}$(incl.) [GeV]	$>$	2000	3100

Table 2. Definitions of our signal regions. SR Meff-2j-2000 is from the ATLAS analysis based on $13.3\,\mathrm{fb}^{-1}$ data at the $13\,\mathrm{TeV}$ LHC [11], and Meff-2j-3100 is based on the HL-LHC study [10]. In Meff-2j-2000 (Meff-2j-3100), jets are required to satisfy $p_T > 50\,\mathrm{GeV}$ and $|\eta| < 2.8$ ($p_T > 20\,\mathrm{GeV}$ and $|\eta| < 4.5$), and $\Delta\phi$ cuts are applied to all the jets with $p_T > 50\,\mathrm{GeV}$ ($p_T > 40\,\mathrm{GeV}$). H_T is the scalar sum of p_T of all the jets, and $m_{\rm eff}$(incl.) is the sum of $\not{E}_T$ and H_T. Events are vetoed if electrons and/or muons with $p_T > 10\,\mathrm{GeV}$ are present.

background is $W + \mathrm{jets}$ production, which we return to below. Dibosons and $t\bar{t}$ production give smaller contributions (see, e.g., ref. [11]).

The dominant W-background is from W decays to tau plus neutrino, with the tau decaying hadronically, predominantly to light jets. These processes receive large contributions from diagrams like figure 1(b), with the Z replaced by a W, leading to a τ, neutrino and two jets in the final state, as well as from diagrams like figure 2(a), which lead to a τ, neutrino and one jet. Another type of background from W production is processes in which the W decays into a light lepton (electron or muon) plus neutrinos, and the lepton is lost in the reconstruction (see, e.g., figure 2(b)).

In the invisible-Z background, the leading source of charm quarks is QCD production of $c\bar{c}$ pairs as shown in figure 1(a). Another important source of charm quarks in the SM background comes from higher-order, but log-enhanced, processes, and in particular "gluon splitting" into a $c\bar{c}$ pair (see, e.g., ref. [12]). While this is not included in our leading order (LO) simulation of the hard processes, some component of charm pairs from gluon splitting is generated by `Pythia`. Gluon-charm initial states in figure 1(b) and gluon-strange initial states in figures 2(a) and 2(b) give small contributions since they are PDF-suppressed. In the latter two figures, processes with an initial-state down-quark are CKM-suppressed.

2.3 Monte Carlo simulation

Signal samples are generated by `MadGraph_aMC@NLO 5` [13] at LO, with the PDF set `NNPDF2.3QED` at LO with $\alpha_s = 0.13$ [14]. The baseline selections described in table 3 are applied based on the missing transverse energy and jet p_T. Parton showering and hadronization are performed by `Pythia 6.4` [15]. Tau decays are simulated by `TAUOLA` [16]. For detector simulation, `Delphes 3.3.0` [17] is utilized with the default detector card, where
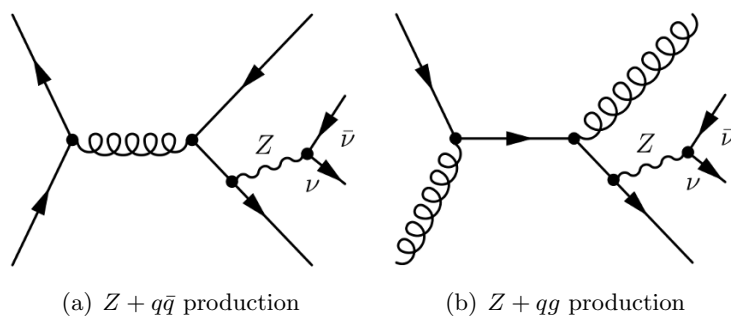
(a) $Z + q\bar{q}$ production   (b) $Z + qg$ production

Figure 1. Diagrams contributing to the $Z + \text{jets}$ background.

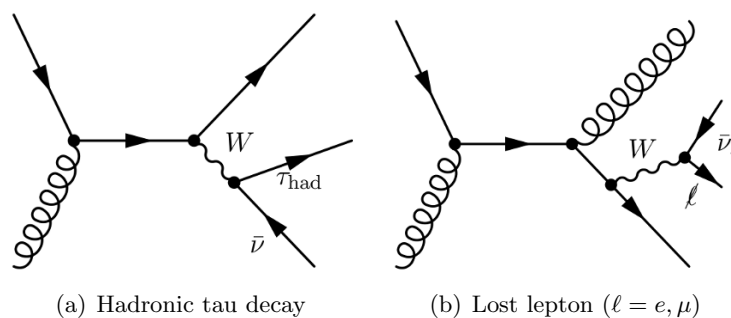(a) Hadronic tau decay   (b) Lost lepton ($\ell = e, \mu$)

Figure 2. Diagrams contributing to the $W + \text{jets}$ background. ℓ denotes an electron or muon that is lost in the reconstruction.

the parameter of anti-k_{T} algorithm [18, 19] for jet clustering is replaced by $R = 0.4$ to match the ATLAS studies. Pile-up effects are not considered. These signal samples are rescaled by next-to-leading-order (NLO) K-factors, which are calculated by `Prospino 2` [20].[1] We then apply the selection cuts for SR Meff-2j-2000 and, using the K-factors for 13 TeV colli-sions, and compare to the upper bound obtained by the ATLAS analysis [11] to determine whether the model point is excluded.

To obtain the charm fraction at the HL-LHC, background events are also generated by the same procedure. Our selection cuts, especially the high m_{eff} cut of 3100 GeV, suppress the $W + \text{jets}$ background so that it is about a third of the $Z + \text{jets}$ background. At the same time, the $Z + \text{jets}$ background is easier to calculate compared to the $W + \text{jets}$ background, since the latter comprises different components (e.g., $j\tau\nu$, $jj\tau\nu$). We simulate the $Z + \text{jets}$ background and the different components of the $W + \text{jets}$ background at leading order. We find that the fractions of (truth-level) charm quarks in each of these are similar. We therefore obtain the total number of background events by reweighting the $Z + \text{jets}$ sample (with the baseline selection in table 3) to match the number of events from $Z + \text{jets}$ and

---

[1]`Prospino` does not handle non-degenerate squarks from the first two generations. However, the NLO correction is dominated by QCD contributions (light quarks and gluons) [21]. Therefore, additional heavy squarks only contribute at next-to-next-to-leading order and the `Prospino` K-factors are a good approxi-mation even in this case. As previously mentioned, we only require mild hierarchies between the masses of the lightest and other squarks, so we will ignore leading-log corrections.

		Meff-2j-2000	Meff-2j-3100	
		Signals	Z + jets	Signals
$|\not{p}_{\mathrm{T}}|$ [GeV]	>	—	150	—
Leading jet p_{T} [GeV]	>	150	500	150
Subleading jet p_{T} [GeV]	>	—	60	—

Table 3. Baseline event selection we use in the simulation. Variables are calculated at the `MG5_aMC` level. The missing transverse momentum $\not{p}_{\mathrm{T}}$ is defined by the sum of neutrino momenta.

W + jets processes in the ATLAS analysis [10] (figure 8b), and approximate the fraction of charm quark events (whether these contain a single charm quark or a pair of charm quarks) in this sample by its value in our simulated Z + jets sample.

Events are selected by the cuts of the Meff-2j-3100 SR. For each event in the SR, the two jets with leading p_{T} are considered in the calculation of the charm fraction. Initially, `Delphes 3.3.0` is utilized to determine the "truth-level" jet flavor. In this algorithm, a jet is considered bottom-flavored, or a "truth-level bottom jet", if one or more bottom quarks (b or $\bar{b}$) exist in the jet cone as a `Pythia`-level parton. If no bottom partons are found but charm partons exist, the jet is labeled as a "truth-level charm jet". Otherwise, the jet is treated as a light jet at the truth level. The detector-level charm tagging is performed based on this "truth-level" information as we describe below.

2.4 Charm tagging

The main limiting factor in measurements of the charm fraction is the charm tagging capabilities and in particular the fake rates. Current analyses at the LHC experiments utilize charm-tagging algorithms based on the working points $(\epsilon_c, \epsilon_b, \epsilon_l) = (0.19, 0.2, 0.005)$ [1] or $(0.2, 0.24, 0.02)$ [2]. Here ϵ_c is the tagging efficiency of charm quarks, while ϵ_b and ϵ_l are the mistag rates for bottom and light jets, respectively. These taggers are primarily trained on $t\bar{t}$ samples and thus the maximal jet transverse momentum does not exceed $300\,\mathrm{GeV}$ [10]. In contrast, the average p_{T} for the simplified models presented here is $\sim 500\,\mathrm{GeV}$, so charm-tagging would be more challenging. Still, charm tagging algorithms will likely undergo significant improvement by the end of the HL-LHC program. Thus, we consider two optimistic scenarios with efficiencies $\epsilon_c = 0.5$ or 0.3 and mistag rates $\epsilon_b = 0.2$ and $\epsilon_l = 0.005$ (see also ref. [22]). Since we cannot reliably estimate the p_{T} and η dependence of the various efficiencies, we take them to be constant over the entire ranges.

For a given set of tagging parameters, $(\epsilon_c, \epsilon_b, \epsilon_l)$, we can divide the sample of events passing the cuts as follows. In each event, we examine the truth-level flavor of each of the two hardest jets. A truth-level charm jet is "tagged" as a charm jet with probability ϵ_c. Similarly, each truth-level l (b) jet is "tagged" as a charm jet with probability ϵ_l (ϵ_b). We denote the number of these "charm-tagged" jets by N_c, and the total number of events in the sample by N_{ev}. For high-efficiency, high-purity taggers, it would also be useful to separately consider events in which the two hardest jets are both tagged as charm jets. We denote the number of these double-tagged events by $N^{\text{2-tag}}$.
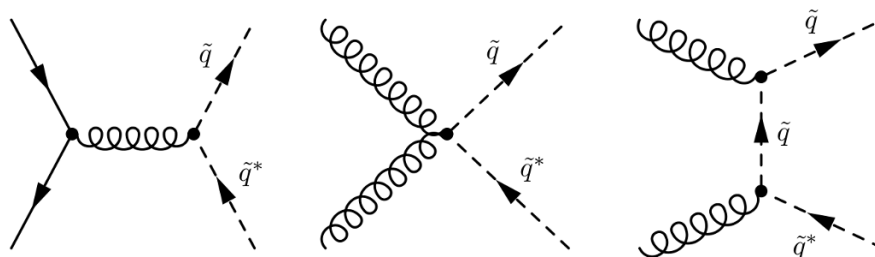
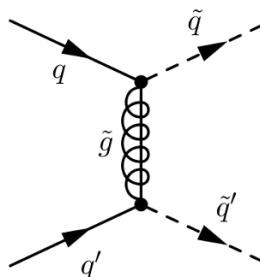**Figure 3**. Diagrams contributing to flavor-democratic squark-pair production.

Figure 4. A diagram contributing to flavor-undemocratic squark-pair production.

One caveat of our simulation is the fact that the `Delphes` algorithm that we utilize treats $c\bar{c}$ pairs originating from gluons or from taus as charm jets, even when the pair is clustered into a single jet. We return to this issue in the discussion of the results.

3   The charm fraction

Squark pairs would be produced at the LHC either through flavor-democratic processes (see figure 3), or through gluino-mediated processes (see figure 4), which are sensitive to the proton PDF's and are thus flavor-dependent. As the gluino mass is increased, the latter processes become less significant.

We define the *charm fraction* F_c as the ratio

$$F_c \equiv \frac{N_c}{2N_{\rm ev}} \,, \tag{3.1}$$

where $N_{\rm ev}$, N_c were defined in section 2.4. For events coming from squark pair production, we expect this fraction to increase as the gluino mass increases. This behavior is exhibited in figure 5(a), where we plot the charm fraction for a model with $N_{\tilde{q}} = 8$ squarks with $m_{\tilde{q}} = 1.5\,{\rm TeV}$ and a massless bino, assuming an ideal tagger ($\epsilon_c = 100\%, \epsilon_b = 0, \epsilon_l = 0$). Error bars are the Monte Carlo uncertainties. The fraction rises from 0.10 to 0.16 as the gluino mass varies from 4 TeV to 13 TeV, and asymptotes to 0.25 for a decoupled gluino. Repeating this for the SM background yields a charm fraction of 0.09. The hollow points are already excluded by the Meff-2j-2000 SR of [11] (see section 2.3 for details).

In figure 5(b), we show the fraction of double-tagged events,

$$F_c^{\text{2-tag}} \equiv \frac{N^{\text{2-tag}}}{N_{\rm ev}} \,. \tag{3.2}$$
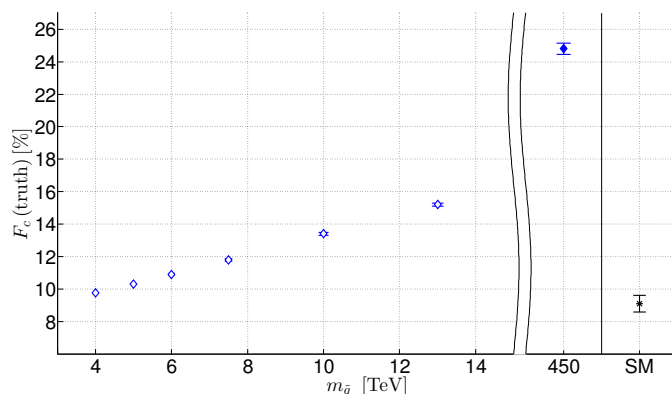
(a) Truth-level charm fraction for jets.
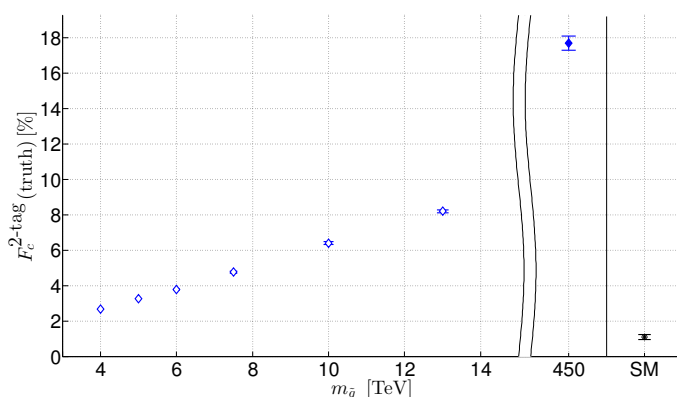
(b) Truth-level charm fraction for double-tagged events.

Figure 5. The truth-level charm fraction for jets (upper) and for double-tagged events (lower) of the signal-only samples as a function of the gluino mass for models with $N_{\tilde{q}} = 8$, $m_{\tilde{q}} = 1.5\,\mathrm{TeV}$, and a massless bino, and that of the SM background sample. Hollow points are excluded by the ATLAS analysis [11]. Error bars are the Monte Carlo uncertainties.

The value for the decoupled gluino is 0.18, which is smaller than the naive expectation of 0.25 because jets from QCD radiation can be harder than charm jets from charm squarks. The SM background is reduced by a larger relative margin as the number of double charm events is smaller for the SM.

The results of figures 5(a) and 5(b) are based on truth-level parton flavor; however, realistically, we must consider charm-tagged jets. In figure 6, we show the results for various model points in the N_{ev}–F_c plane, assuming tagging efficiencies of $(\epsilon_c, \epsilon_b, \epsilon_l) = (0.5, 0.2, 0.005)$. We focus on a narrow range of squark masses, for which the models may potentially be probed by the HL-LHC, and which are not yet excluded for at least some gluino masses. We then vary the number of squarks produced, over $N_{\tilde{q}} = 2, 4, 8$, and consider both heavy (300 GeV) and massless binos. The squark and bino masses are chosen such that all the models yield similar kinematics, and cannot be distinguished based on m_{T2} [7]. Each shape-color combination maps to a particular choice of $(N_{\tilde{q}}, m_{\tilde{q}}, m_{\tilde{\chi}_1^0})$: the
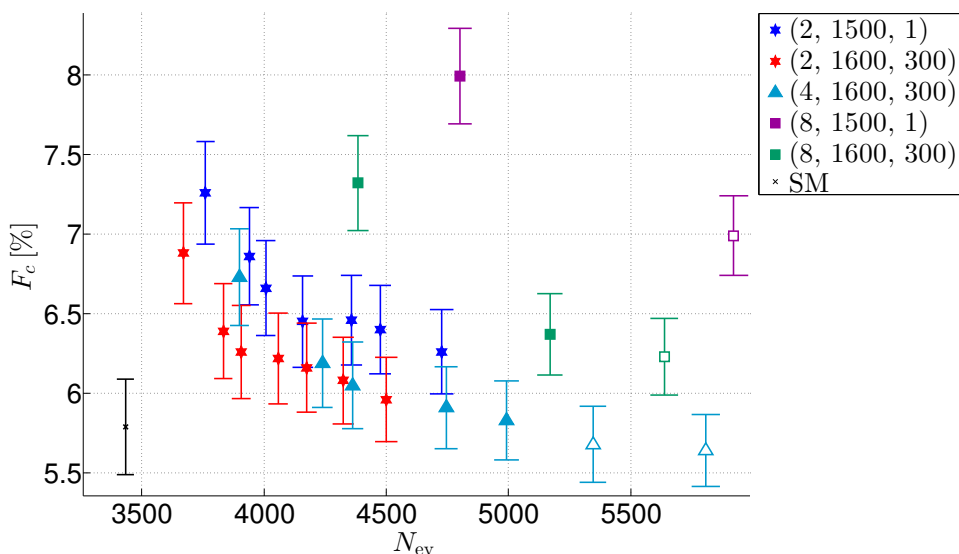
Figure 6. The number of events N_{ev} and the charm fraction F_c in the SM+supersymmetry samples, as well as the SM-only sample, in the Meff-2j-3100 SR, expected at the HL-LHC with $\sqrt{s} = 14\,\text{TeV}$ and $\int \mathcal{L} = 3000\,\text{fb}^{-1}$. The bars are the statistical uncertainty on F_c. Hollow points are excluded by the ATLAS analysis [11]. The tagging efficiencies are $(\epsilon_c, \epsilon_b, \epsilon_l) = (0.5, 0.2, 0.005)$. The theoretical m_{T2} endpoints for these models are in the range $1500 \pm 50\,\text{GeV}$. The numbers in the legend correspond to $(N_{\tilde{q}}, m_{\tilde{q}}/\text{GeV}, m_{\tilde{\chi}_1^0}/\text{GeV})$.

shape of the central value marker indicates $N_{\tilde{q}}$, and the color designates pairs of squark and bino masses $(m_{\tilde{q}}, m_{\tilde{\chi}_1^0})$. Points with the same shape and color correspond to the different gluino masses of table 1: in a sequence increasing in N_{ev}, the values of $m_{\tilde{g}}$ decrease, beginning with $m_{\tilde{g}} = 450\,\text{TeV}$. Points with hollow central values are already excluded by the Meff-2j-2000 analysis of [11]. Only statistical uncertainties on F_c (assuming $3000\,\text{fb}^{-1}$ of integrated luminosity) are shown.

For the largest gluino masses, discovery based on N_{ev} alone would be challenging. In figure 6, we have not displayed horizontal error bars on N_{ev} since we cannot reliably estimate the dominant systematic uncertainties. Still, using current LHC analyses as a guide, it is reasonable to expect the systematic uncertainty on N_{ev} to be around 10% [10]. The SM prediction is then $N_{\text{ev}}^{\text{SM}} = 3433 \pm 59_{\text{stat}} \pm 343_{\text{syst}} = 3433 \pm 348$. Roughly, most of the model points of figure 6 lead to excesses in N_{ev} below 3σ (corresponding to $N_{\text{ev}} \lesssim 4500$). Furthermore, for fixed values of $(N_{\tilde{q}}, m_{\tilde{q}}/\text{GeV}, m_{\tilde{\chi}_1^0}/\text{GeV})$, only a limited range of gluino masses remains for which a 5σ discovery, requiring $N_{\text{ev}} \gtrsim 5200$, would be possible. Recall that hollow points denote models which are excluded by ref. [11].

On the other hand, for large gluino masses, the charm content of supersymmetry events is large, so charm tagging can be used to increase the sensitivity to these models. Since it is down by the fraction of charm squarks produced and the charm tagging efficiency, the number of charm-tagged events is prone to larger statistical uncertainties compared to N_{ev}; however, many systematic uncertainties cancel out in this ratio, including the uncertainty

on the jet energy scale, which affects the determination of both the missing energy and m_{eff}. We expect the dominant remaining sources of uncertainties to be the charm tagging efficiencies and the PDF's. Note that the latter do not completely cancel in the ratio, as the charm fraction is sensitive to the relative sizes of the PDF's of the valence quarks, gluons, and sea quarks.[2] As seen in figure 6, for models with a heavy gluino, the charm fraction displays the largest deviation from the SM background.

Let us assume a 10% uncertainty on F_c to get a rough estimate of the discriminating power of this variable. The SM then predicts $F_c = (5.8 \pm 0.3_{\text{stat}} \pm 0.6_{\text{syst}})\% = (5.8 \pm 0.6)\%$. The deviation of F_c from the SM prediction is then at the level of 1.4–3.2σ for decoupled gluino models, and when combined with N_{ev}, may allow for discovering these models. Obviously, the F_c values shown in figure 6 are weighted averages of the SM and supersymmetry samples, and are particularly skewed towards the smaller SM value when the squark production cross section is small. Thus for example, for the $N_{\tilde{q}} = 2$ models with $m_{\tilde{q}} = 1.5\,\text{TeV}$, a vanishing bino mass, and gluino masses of 6 TeV and above, the excesses in N_{ev} vary between 0.5–2.7σ, while the deviations of the charm fraction from the SM prediction vary between 2.3σ (for the decoupled gluino) and 1.0σ (for the 6 TeV gluino). The combination of these two variables increases the sensitivity for these challenging scenarios.

If an excess in N_{ev} is observed, attention will be focused on the properties of the new particles produced and on whether additional new particles exist. As explained above, the model points shown here are chosen such that the end points of their various m_{T2} distributions lie in the range $1500 \pm 50\,\text{GeV}$.[3] Thus, it will be difficult to distinguish between them based on their missing energy signatures. The charm fraction can clearly break the degeneracy between different underlying models. While a definitive statement cannot be made given that we cannot reliably estimate the systematic uncertainties, the results suggest that models with gluino masses around or below 10 TeV can be discriminated from decoupled gluino scenarios. Thus for example, F_c can easily discriminate between the $(8, 1500, 1)$ model with a decoupled gluino, which gives $N_{\text{ev}} \simeq 4800$ and $F_c = (7.9 \pm 0.3_{\text{stat}})\%$, and the other models which give a similar value of N_{ev}, but with gluino masses between 4–7.5 TeV and $F_c = (5.9\text{–}6.3)\%$; assuming a 10% systematic uncertainty on F_c, these are 2–3σ away from each other.

With a high efficiency to tag charm jets, it is sensible to also consider the fraction of double-charm-tagged events, $F_c^{2\text{-tag}}$, which we plot for the same set of models in figure 7. Compared to F_c, $F_c^{2\text{-tag}}$ suffers from QCD radiation effects and larger statistical uncertainties due to a further reduction by approximately ϵ_c. If the systematic uncertainty in the charm fraction is dominated by uncertainties on tagging efficiencies, $F_c^{2\text{-tag}}$ will also be subject to a systematic uncertainty approximately twice that of F_c. However, because the SM prediction for $F_c^{2\text{-tag}}$ is small, a deviation from the SM value will be more significant. For example, assuming a 20% systematic uncertainty on $F_c^{2\text{-tag}}$, the decoupled gluino scenarios with $N_{\tilde{q}} = 2$ will have $F_c^{2\text{-tag}} = (1.2 \pm 0.3)\%$ and $(0.96 \pm 0.25)\%$, which are 4–5σ away

---

[2]The PDF uncertainties are expected to shrink by the end of the HL-LHC program (see, e.g., ref. [23]).

[3]The central value of 1500 GeV roughly corresponds to the limit for discovery of squarks at the HL-LHC in simplified models with a decoupled gluino, $N_{\tilde{q}} = 8$, and a massless bino LSP [10].
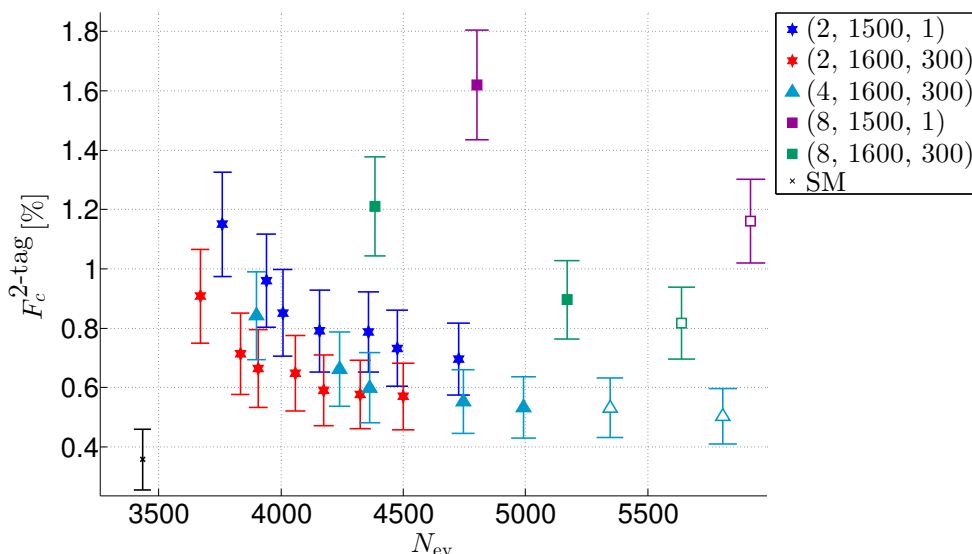
Figure 7. Same as figure 6, but with the fraction of double-tagged events $F_c^{\text{2-tag}}$ as the vertical axis.

from the SM expectation. As for discriminating between different supersymmetry models, comparing figures 6 and 7, we find that F_c and $F_c^{\text{2-tag}}$ have approximately the same analyzing power. Considering the aforementioned model points with $N_{\text{ev}} \sim 4800$, the $(8, 1500, 1)$ model with a decoupled gluino gives $F_c^{\text{2-tag}} = (1.58 \pm 0.18_{\text{stat}} \pm 0.32_{\text{syst}})\% = (1.58 \pm 0.36)\%$, while the other models give $(0.60 \pm 0.16)\%$ and $(0.72 \pm 0.19)\%$, which are 2–3σ away from each other.

Figure 6 assumes charm tagging efficiencies of $(\epsilon_c, \epsilon_b, \epsilon_l) = (0.5, 0.2, 0.005)$. This is far better than those currently published [1, 2]. Furthermore, these numbers are expected to deteriorate as the jet p_T increases; however, since we do not know the high-p_T and η dependence of the tagger at the HL-LHC, we take them to be constant over the entire range. For comparison, we also show results for the same model points, but with more conservative efficiencies, in figures 8 and 9.

3.1 Discussion

The charm fractions displayed in figures 6–9 contain both SM and supersymmetric contributions. As noted in section 2.4, our analysis overestimates the number of charm jets from gluon splitting compared to realistic detectors, since with `Delphes`, a jet containing a $c\bar{c}$ pair is labelled as a charm jet. This occurs in both supersymmetry and SM events, but since gluon splitting is more important in the SM, the effect on the SM charm fraction is more pronounced.

The charm fraction of supersymmetric models can in principle be lower than in SM events. This requires a low gluino mass, which is not of much interest to us, since discovery in this case occurs based on the total number of events. In the pure supersymmetry sample, the charm fractions of $N_{\tilde{q}} = 8$ and $N_{\tilde{q}} = 4$ are identical (assuming the other three
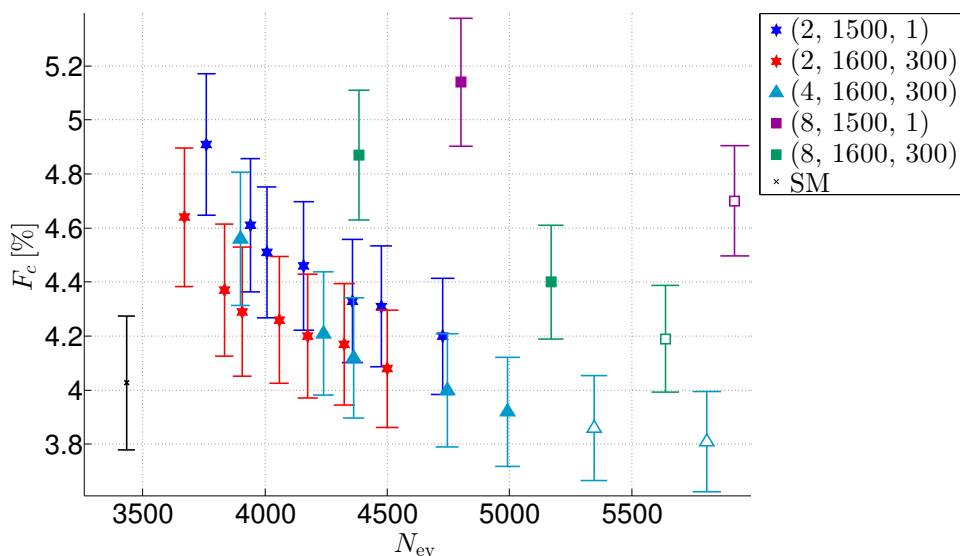
Figure 8. Same as figure 6, but with tagging efficiencies $(\epsilon_c, \epsilon_b, \epsilon_l) = (0.3, 0.2, 0.005)$.
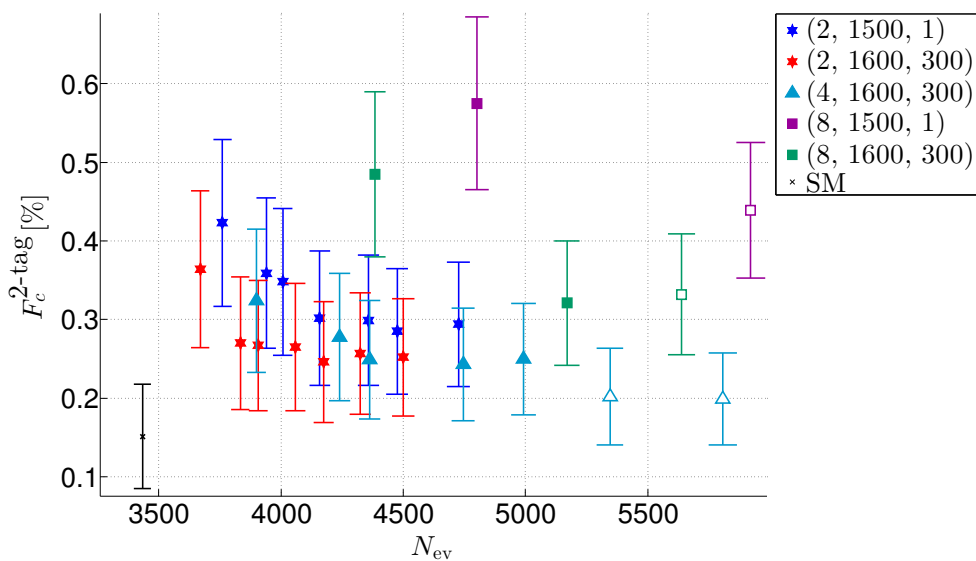

Figure 9. Same as figure 8, but for double tagged events.

parameters are equal). Note that, while bino-mediated t-channel processes give an $\mathcal{O}(1\%)$ modification of the cross section and are thus negligible, processes involving winos may give an $\mathcal{O}(10\%)$ contribution even if the winos are as heavy as the squarks. Therefore, in the presence of left-handed squarks, our results assume an extremely large wino mass. Still, our analysis can be straightforwardly generalized to include winos, with little qualitative changes. Charm squarks would mainly decay to strange quarks in this case and vice versa.

As mentioned above, estimating the number of charm jets from SM processes is nontrivial. Beyond this theoretical difficulty, in standard experimental analyses, collinear charm pairs from gluon splitting are typically merged into a single jet, but jets containing two heavy quarks are subsequently discarded [24]. New approaches for heavy flavor tagging were proposed recently to address this problem [24]. The intrinsic charm fraction of the proton is another potential source of charm quarks that is hard to estimate [25].

Fortunately, these theoretical uncertainties can be straightforwardly circumvented by measuring the charm content of the SM background in the data. For the Meff-2j-3100 SR analysis, the background is dominated by Z + jets, and the charm fraction can be measured in the analogous sample with the Z decaying leptonically. In fact, $Z + c$ production with leptonic Z decays has been used by CMS for training some of their charm taggers [26]. Thus, one can extract the numbers of both charm and non-charm jets in the sample of invisible-Z decays, and by subtracting them, obtain the purely supersymmetric charm fraction. While this will be subject to larger experimental uncertainties, the theory systematics will be significantly improved.

4 Conclusions

The exclusion limits on superpartner masses from ATLAS and CMS are fast approaching the discovery reach of the LHC. The fraction of charm quarks in jet plus missing energy events provides a new handle on superpartner production, and may increase the sensitivity of LHC searches to squark-pair production. While we have only studied here squark pair production, the charm fraction in gluino-pair production is of interest too as this process is flavor democratic for degenerate squarks.

We did not address here the production of top and bottom squarks. Because of their relatively large Yukawa couplings, these are likely to be split in mass from the first- and second-generation squarks. Furthermore, because the bottom and top content of the proton is negligible, their production is mainly gluon-mediated, and to a good approximation, independent of the gluino mass. Thus, while their discovery would yield additional information, it is orthogonal to our discussion here. We also neglect winos and higgsinos in this study. The latter have little effect on first- and second-generation squark production. Winos, on the other hand, mediate t-channel squark production, and would alter our results unless they are very heavy.

We have argued that the charm fraction can be used to disentangle different model points with similar kinematics. We note that, while event kinematics are largely governed by the squark and bino mass, they have some sensitivity to the gluino mass as well, since how central or forward the events are depends on the weight of t-channel gluino processes. This suggests that measurements of the charm fraction may be optimized by a judicious choice of kinematic cuts in order to extract the gluino mass.

Here we studied models with mass-degenerate up and charm squarks. While plausible, this is by no means mandatory. With mass splittings between the squarks, the fermion Cabibbo mixing will typically translate into up-charm mixing of the left-handed squarks; for concrete spectra, see, e.g., ref. [27]. Measuring the charm fraction will yield information on the squark flavor composition.

Acknowledgments

References

[1] ATLAS collaboration, *Performance and Calibration of the JetFitterCharm Algorithm for c-Jet Identification*, [ATL-PHYS-PUB-2015-001](#) (2015).

[2] CMS collaboration, *Identification of c-quark jets at the CMS experiment*, [CMS-PAS-BTV-16-001](#) (2016).

[3] R. Mahbubani, M. Papucci, G. Perez, J.T. Ruderman and A. Weiler, *Light Nondegenerate Squarks at the LHC*, [*Phys. Rev. Lett.* **110** (2013) 151804](#) [[arXiv:1212.3328](#)] [[INSPIRE](#)].

[4] C.W. Kalderon, *Searches for Scharms and Gluinos with the ATLAS Detector*, Ph.D. Thesis, Oxford University, Oxford U.K. (2016).

[5] CMS collaboration, *Search for top squarks decaying to a charm quark and a neutralino in events with a jet and missing transverse momentum*, [CMS-PAS-SUS-13-009](#) (2014).

[6] ATLAS collaboration, *Search for Scalar Charm Quark Pair Production in pp Collisions at $\sqrt{s} = 8\,TeV$ with the ATLAS Detector*, [*Phys. Rev. Lett.* **114** (2015) 161801](#) [[arXiv:1501.01325](#)] [[INSPIRE](#)].

[7] C.G. Lester and D.J. Summers, *Measuring masses of semiinvisibly decaying particles pair produced at hadron colliders*, [*Phys. Lett.* **B 463** (1999) 99](#) [[hep-ph/9906349](#)] [[INSPIRE](#)].

[8] T. Cohen et al., *SUSY Simplified Models at 14, 33 and 100 TeV Proton Colliders*, [*JHEP* **04** (2014) 117](#) [[arXiv:1311.6480](#)] [[INSPIRE](#)].

[9] T. Golling et al., *Physics at a 100 TeV pp collider: beyond the Standard Model phenomena*, [*CERN Yellow Report* (2017) 441](#) [[arXiv:1606.00947](#)] [[INSPIRE](#)].

[10] ATLAS collaboration, *Search for Supersymmetry at the high luminosity LHC with the ATLAS experiment*, [ATL-PHYS-PUB-2014-010](#) (2014).

[11] ATLAS collaboration, *Further searches for squarks and gluinos in final states with jets and missing transverse momentum at $\sqrt{s} = 13\,TeV$ with the ATLAS detector*, [ATLAS-CONF-2016-078](#) (2016).

[12] A. Banfi, G.P. Salam and G. Zanderighi, *Accurate QCD predictions for heavy-quark jets at the Tevatron and LHC*, *JHEP* **07** (2007) 026 [arXiv:0704.2999] [INSPIRE].

[13] J. Alwall et al., *The automated computation of tree-level and next-to-leading order differential cross sections and their matching to parton shower simulations*, *JHEP* **07** (2014) 079 [arXiv:1405.0301] [INSPIRE].

[14] NNPDF collaboration, R.D. Ball et al., *Parton distributions with QED corrections*, *Nucl. Phys.* **B 877** (2013) 290 [arXiv:1308.0598] [INSPIRE].

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

[16] S. Jadach, Z. Was, R. Decker and J.H. Kuhn, *The tau decay library TAUOLA: Version 2.4*, *Comput. Phys. Commun.* **76** (1993) 361 [INSPIRE].

[17] DELPHES 3 collaboration, J. de Favereau et al., *DELPHES 3, A modular framework for fast simulation of a generic collider experiment*, *JHEP* **02** (2014) 057 [arXiv:1307.6346] [INSPIRE].

[18] M. Cacciari, G.P. Salam and G. Soyez, *The Anti-k(t) jet clustering algorithm*, *JHEP* **04** (2008) 063 [arXiv:0802.1189] [INSPIRE].

[19] M. Cacciari, G.P. Salam and G. Soyez, *FastJet User Manual*, *Eur. Phys. J.* **C 72** (2012) 1896 [arXiv:1111.6097] [INSPIRE].

[20] W. Beenakker, R. Hopker, M. Spira and P.M. Zerwas, *Squark and gluino production at hadron colliders*, *Nucl. Phys.* **B 492** (1997) 51 [hep-ph/9610490] [INSPIRE].

[21] D. Gonçalves-Netto, D. López-Val, K. Mawatari, T. Plehn and I. Wigmore, *Automated Squark and Gluino Production to Next-to-Leading Order*, *Phys. Rev.* **D 87** (2013) 014002 [arXiv:1211.0286] [INSPIRE].

[22] G. Perez, Y. Soreq, E. Stamou and K. Tobioka, *Prospects for measuring the Higgs boson coupling to light quarks*, *Phys. Rev.* **D 93** (2016) 013001 [arXiv:1505.06689] [INSPIRE].

[23] D. d'Enterria, *QCD at FCC*, talk at *1st FCC Physics Workshop*, Geneva Switzerland (2017), https://indico.cern.ch/event/550509/contributions/2413809/attachments/1397831/2134849/dde_qcd_at_fcc_cern_jan17.pdf.

[24] P. Ilten, N.L. Rodd, J. Thaler and M. Williams, *Disentangling Heavy Flavor at Colliders*, arXiv:1702.02947 [INSPIRE].

[25] S.J. Brodsky, A. Kusina, F. Lyonnet, I. Schienbein, H. Spiesberger and R. Vogt, *A review of the intrinsic heavy quark content of the nucleon*, *Adv. High Energy Phys.* **2015** (2015) 231547 [arXiv:1504.06287] [INSPIRE].

[26] CMS collaboration, *Measurement of associated Z + charm production in pp collisions at $\sqrt{s} = 8\,TeV$*, CMS-PAS-SMP-15-009 (2016).

[27] N. Ierushalmi, S. Iwamoto, G. Lee, V. Nepomnyashy and Y. Shadmi, *LHC Benchmarks from Flavored Gauge Mediation*, *JHEP* **07** (2016) 058 [arXiv:1603.02637] [INSPIRE].