The Need for New Search Strategies for Fourth Generation Quarks at the LHC

Michael Geller\textsuperscript{1}, Shaouly Bar-Shalom\textsuperscript{1}, and Gad Eilam \textsuperscript{1,2}

\textit{1. Physics Department, Technion-Institute of Technology, Haifa 32000, Israel.}
\textit{2. Physics Department, University of Virginia, Charlottesville VA 22904, USA.}

Most limits on the fourth generation heavy top quark (the $t'$) are based on the assumed dominance of $t' \to Wb$, which is expected to be case in the minimal fourth generation framework with a single Higgs (the so called SM4). Here we show, within a variant of a Two Higgs Doublet Model with four generations of fermions, that, in general, a different $t'$ detection strategy is required if the physics that underlies the new heavy fermionic degrees of freedom goes beyond the "naive" SM4. We find that the recent CMS lower bounds: $m_{t'} < 450 \text{ GeV}$ in the semi-leptonic channel $pp \to t'T \to t'\nu b\bar{b}$ and $m_{t'} < 557 \text{ GeV}$ in the dilepton channel $pp \to t'T \to \ell^+\ell^-\nu\bar{\nu}b\bar{b}$, that were obtained using the customary (SM4-driven) detection strategies, do not apply.

In particular, we demonstrate that if the decay $t' \to ht$ dominates, then applying the "standard" CMS search tools leads to a considerably relaxed lower bound: $m_{t'} \gtrsim 350 \text{ GeV}$. We, therefore, suggest an alternative search strategy that is more sensitive to beyond SM4 dynamics of the fourth generation fermions.

I. INTRODUCTION

The origin of the ElectroWeak Symmetry Breaking (EWSB), one of the most studied theoretical puzzles in particle physics, is yet unknown. It is anticipated that the LHC, that has already begun its historic mission of collecting data at 7 TeV collision energy, will provide insight into the heart of this problem. In the Standard Model (SM), EWSB is triggered by the Higgs mechanism with a single fundamental scalar, leaving its mass unprotected from radiative corrections at the cutoff scale. This is known as the hierarchy problem: the tension between the cutoff scale (Planck scale or GUT scale) and the seemingly unnatural ElectroWeak (EW) scale, which is usually interpreted as evidence for new physics at the TeV scale.

Hints for the existence of a Higgs boson have been recently reported at the LHC: ATLAS finds an excess of events around $M_h = 126 \text{ GeV}$ with a local significance of 3.5 standard deviations [1], while CMS sees a smaller excess with a local significance of 3.1 sigma around $M_h = 124 \text{ GeV}$ [2]. Later, it was shown at the Tevatron [3] that there is a possibility of an excess of events corresponding to $M_h$ between 115 GeV and 135 GeV. It is, therefore, tempting to consider a Higgs mass in the range $(120 - 130) \text{ GeV}$. The LHC with higher integrated luminosity and 8 TeV c.m. energy should unveil more precise details about the Higgs boson.

In the past few years, several tensions between experiments and the SM predictions were observed in flavor physics [4]–[12]. It was noted, that adding a fourth sequential generation (for reviews see [13]–[16]) to the SM (the so called SM4 which in some instances will be referred to as the "naive" four generations model) might not only be able to resolve these tensions [17]–[29], but may also address the hierarchy problem [30]–[32]. In particular, when the new fermion states are heavy enough, their Yukawa couplings are driven to a Landau pole or a fixed point, possibly at the TeV scale, where some form of strong dynamics might be responsible for dynamically breaking the EW symmetry. Besides, the additional CP-violating phases in the presence of a fourth family [33] may be useful for Baryogenesis [34]–[36].

The fact that adding a heavy fourth generation to the SM implies new physics at $O$ (TeV), serves as a strong hint that the SM4 is not an appropriate framework to account for the expected (possibly strong) dynamics in the presence of these new heavy fermionic states. For this purpose and others, models that go beyond the SM4 (BSM4) were introduced as an alternative to the naive SM4 framework [37]–[39].

One way to incorporate the low-energy effects of such possible TeV-scale strong dynamics in an acceptable phenomenological approach is multi-Higgs scenarios, which may naturally account for the composite scalars associated with the new strong dynamics, see e.g., [37]–[44]. An example of a model designed for this purpose, the 4G2HDM (a model with 2 Higgs doublets and a fourth family of fermions), was proposed in [43], and was shown to agree with Electroweak Precision Data (EWPD) in a wide range of its parameter space. An alternative interesting approach to the compositeness picture is warped extra-dimension scenarios, see e.g., [45]–[48]. Other examples of BSM4 scenarios were suggested in [49]–[51].

Recently, assuming the SM4 framework, CMS reported [52]–[53] a 450 GeV lower limit on the $t'$ mass in the semileptonic channel ($pp \to t'T \to [W^+],_{\text{hadronic}} b [W^-],_{\text{leptonic}} b \to \ell v b\bar{b}$), and a 557 GeV lower limit in the dilepton channel ($pp \to t'T \to [W^+],_{\text{leptonic}} b [W^-],_{\text{leptonic}} b \to \ell^+\ell^-\nu\bar{\nu}b\bar{b}$), replacing the weaker Tevatron limits [54]–[55]. The most recent lower bounds on the $b'$ mass are 480 GeV [50] (Atlas) and 611 GeV [57] (CMS). In addition, the single
Higgs of the SM4 is excluded now by CMS in the range $120 - 600$ GeV \cite{58}. We argue, that these seemingly stringent limits apply only to the SM4 framework, which is disfavored for reasons that are outlined above. In particular, in BSM4 frameworks, collider signals of fourth generation fermions and/or Higgs particles may be drastically altered, leading to weaker bounds. An example of this was discussed in \cite{59}, where an extended Higgs sector (with 4th generation fermions) may highly suppress the signal for Higgs production, circumventing the current limits on the SM4 Higgs mass.

The purpose of this work is, therefore, to illustrate that, in general, the limits on the $t'$ mass can be considerably weaker with respect to the SM4 case. We demonstrate that in the 4G2HDM framework of \cite{43}, where new decay patterns of $t'$ emerge through the scalar sector: e.g., $t' \to H^0 b$. We thus propose an alternative search method, more general and, therefore, more suitable to searches for BSM4 scenarios.

II. $t'$ DETECTION AT THE LHC- BEYOND SM4

As mentioned above, CMS has recently reported \cite{52} a 557 GeV lower bound on the $t'$ mass, replacing the previous limits by CDF \cite{54} (358 GeV) and D0 \cite{55} (285 GeV) \cite{69}. A summary on the Tevatron limits can be found at \cite{69}. These searches assume $Br(t' \to W^+ b) \sim O(1)$ (SM4) and focus on either the semileptonic channel ($1\ell + nj + E_T^\text{miss}$) or the dilepton channel ($2\ell + nj + E_T^\text{miss}$) in the $t'$ pair production (for a limit that takes into account the branching ratios in $t'$ decay within the SM4, see \cite{70}): 

$$pp \to t't' \to [W^+_\text{hadronic/leptonic}]_{L} b [W^-_{\text{leptonic}}]_T b \to \ell \nu b q \bar{q} b/\ell^+\ell^- \nu \tau b\bar{b}$$

The limits for the semileptonic channel were extracted using two dimensional fits to $M_{fit}$ and $H_T$ - the reconstructed mass of the $t'$ and the scalar sum of $p_T$ of the visible jets and leptons and missing $E_T$, respectively. In particular, the mass of the $t'$ is reconstructed by fitting each event to the process $pp \to W^+ W^- b\bar{b} \to \ell \nu b q \bar{q} b$, i.e., taking the 4 leading jets (with highest $p_T$) and minimizing the $\chi^2$ for the equations:

$$m(\ell \nu) = M_W$$
$$m(q\bar{q}) = M_W$$
$$m(\ell \nu b) = m(q\bar{q} b) \equiv M_{fit}$$

On the other hand, the analysis for the dilepton channel relies on the fact that $M_{tb}$, which is the invariant mass of a pair of any lepton and a $b$-jet in the event, is much higher in the $t't'$ signal with respect to the leading $t\bar{t}$ background. In particular, in the case of $t\bar{t}$, $M_{tb}$ has an upper bound that corresponds to the mass of the top quark, and therefore in the region above 170 GeV ("signal region") $M_{tb}$ is a clean signal for the $t'$.

Below we will study the relevance of these limits in the general BSM4 case, in the feasible scenario where $t' \to W^+ b$ is not the leading $t'$ decay channel. This may occur in models with an extended Higgs sector, e.g. in the 4G2HDM suggested in \cite{43}. In such models, the new scalars play a crucial role in the $t'$ phenomenology, and new channels like $t' \to H^0 t$, $t' \to H^+ b$ (see also \cite{58}) or even the SM4 forbidden $t' \to W^+ b'$ may dominate. In \cite{43} it was shown that all these channels are allowed by EWPD, and some of them dominate within large areas in the (EWPD allowed) parameter space.
Thus, in this scenario, $pp \rightarrow \ell \nu b q \bar{b}$ are no longer the leading signatures for $t'$ pair production at the LHC; the new channels can dominate, giving rise to several possible new signatures for $t'$ at the LHC (see section 4). We argue that for these cases, the CMS fit to $pp \rightarrow \ell \nu b q \bar{b}$ and $pp \rightarrow \ell \nu b q \bar{b}$ should fail, as it attempts to impose these specific SM4-motivated dynamics on processes that have a completely different topology. In particular, for the semileptonic channel, taking only 4 leading jets out of all the jets in the event will miss a large part of the energy of the event and will, therefore, result in $M_{fit}$ being substantially lower - exactly where the main $t'$ background is.

An example of that can be seen in Figure 1 for the 6W+2b signature which occurs in the 4G2HDM (see section 4). For the dilepton channel, the analysis will fail for signatures with more than 2 leptons or $b$-jets, as the combinatorial background will lower the $M_{fit}$ (see section 6), resulting in much less events in the signal region.

In the semileptonic channel the kinematics of each event is completely determined, thus we suggest using a reconstruction that does not assume a specific decay chain, and instead reconstructs any event of the type $1\ell + nj + E_T$ under the sole assumption that the lepton and the missing energy are the result of the leptonic decay of a W boson, similarly to \cite{60}. This will result in the correct reconstruction of $t\bar{t}$ events (see section 5), regardless of the specific decay signature.

The specifics of the such a reconstruction strategy are the following: we select all the jets with transverse momentum that is high enough, we then reconstruct the leptonic W and choose the correct partition of the event (which jet is $\ell$ and which one from $\bar{\nu}b$) by minimizing the $\chi^2$ of the following equations

\begin{equation}
\begin{align*}
m(\nu) &= M_W \\
m(Left\ Side) &= m(Right\ Side) \equiv M_{gen}
\end{align*}
\end{equation}

Another important point to notice is that for all the new signatures, the number of jets in the final state is higher than the SM background, and for some of the signatures there is also an excess of $b$-jets, as will be shown in section 4. We will use cuts based on these properties in section 5, and, as we will further show, this will result in higher detection sensitivity for these signatures.

### III. THE 4G2HDM

Let us recapitulate the underlying properties of the 4G2HDM, the full account of which appears in \cite{43}. The 4G2HDM is a two-Higgs doublet model with 4 generations of chiral fermions, where the Yukawa terms are defined as follows:

\begin{equation}
L_Y = -\mathcal{Q}_L \left( \Phi_q F \left( I - I_d^{\alpha \beta q} \right) + \Phi_h F \cdot I_d^{\alpha \beta q} \right) d_R - \mathcal{Q}_L \left( \tilde{\Phi}_l G \left( I - I_u^{\alpha \beta u} \right) + \tilde{\Phi}_h G \cdot I_u^{\alpha \beta u} \right) u_R + h.c.
\end{equation}

\begin{equation}
\text{(4)}
\end{equation}

such that $Q_L$ is the left-handed SU(2) quark doublet, $f_{L(R)}$ are the left (right)- handed fermion fields and $F,G$ are arbitrary 4x4 Yukawa matrices in flavor space. Also, $I$ is the identity matrix, $I_d^{\alpha \beta q}$ are diagonal matrices defined as

\begin{equation}
I_d^{\alpha \beta q} = \text{diag}(0,0,\alpha_q,\beta_q)
\end{equation}

\begin{equation}
\text{(4)}
\end{equation}

and $\Phi_q, \Phi_h$ are the two Higgs doublets.

The different choices of $\alpha_q, \beta_q$ will result in different Yukawa textures and, therfore, different models. In this work we assume $(\alpha_d, \beta_d, \alpha_u, \beta_u) = (0,1,0,1)$ for which $\Phi_q$ couples only to the $4^{th}$ generations quarks, while $\Phi_h$ couples to the first 3 generations. This setup was defined as 4G2HDM of type I in \cite{43} and is partly motivated by the absence of (tree-level) FCNC in the first 3 generations. In particular, in this case, FCNC is only allowed between the third and fourth generations and it plays a crucial role in the phenomenology of the 4th generation fermions.

The physical Higgs scalars that emerge after EWSB are: $H^{\pm}$ (charged scalar), $h,H$ (lighter and heavier CP-even neutral scalars) and $A$ (CP-odd neutral scalar). Their couplings to the quarks are given by:

\begin{equation}
\begin{align*}
L(hq,qj) &= \frac{g}{2M_W} \tilde{q}_i \left( m_q \frac{c_\alpha}{c_\beta} \delta_{ij} - \frac{\frac{c_\alpha}{s_\beta}}{s_\beta} \right) \left( m_q \Sigma^a_{ij} R + m_q \Sigma^a_{j} L \right) q_j h, \\
L(Hq,qj) &= \frac{g}{2M_W} \tilde{q}_i \left( -m_q \frac{c_\alpha}{c_\beta} \delta_{ij} - \frac{\frac{c_\alpha}{s_\beta}}{s_\beta} \right) \left( m_q \Sigma^a_{ij} R + m_q \Sigma^a_{j} L \right) q_j H, \\
L(Aq,qj) &= -i \frac{g}{2M_W} \tilde{q}_i \left( m_q \tan \beta \gamma_5 \delta_{ij} - (\tan \beta + \cot \beta) \right) \left( m_q \Sigma^a_{ij} R - m_q \Sigma^a_{j} L \right) q_j A, \\
L(H^+ q_j q_j) &= \frac{g}{\sqrt{2}M_W} \tilde{u}_i \left[ m_d \tan \beta V_{ui,dj} - m_d \tan \beta \right] V_{ik} \Sigma^a_{k} R^+, \\
&\quad \left[ -m_u \tan \beta V_{ui,dj} + m_u \tan \beta \Sigma^a_{k} V_{jk} \right] L d_j H^+,
\end{align*}
\end{equation}

where $q = u, d$, $I_u, I_d = \pm \frac{1}{2}, -\frac{1}{2}$, $V$ is 4x4 CKM matrix, $\alpha$ is the mixing angle in the CP even neutral Higgs sector and $\tan \beta$ is the ratio between the VEV of $\Phi_h$ and the VEV of $\Phi_q$, i.e. $\tan \beta = \frac{v_h}{v_q}$. Also, $\Sigma^a$ are new mixing matrices in the
FIG. 2: $R_{gg} = \frac{\Gamma(h \rightarrow gg)_{\text{4G2HDM}}}{\Gamma(h \rightarrow gg)_{\text{SM}}}$ as a function of $\alpha$ for $m_{t'} = m_{b'} = 450$ GeV and $\tan \beta = 1$.

up and the down sectors accordingly, which, for our choice of $\alpha_q, \beta_q$, can be parametrized in the following way [43]:

$$
\Sigma^d \simeq \begin{pmatrix}
0 & 0 & 0 & \epsilon_b^* \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & \epsilon_b & 1 - |\epsilon_b|^2 & 0
\end{pmatrix}
$$

$$
\Sigma^u = \begin{pmatrix}
0 & 0 & 0 & \epsilon_t^* \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & \epsilon_t & 1 - |\epsilon_t|^2 & 0
\end{pmatrix}
$$

(6)

where $\epsilon_b$ and $\epsilon_t$ effectively parametrize the mixing between the third and fourth generation quarks in the down and up sectors, respectively. In the following we will set $\epsilon_b = \frac{m_b}{m_{b'}}$ and $\epsilon_t = \frac{m_t}{m_{t'}}$, as natural values representative for these quantities (see [43]).

IV. BEYOND SM4 DECAY PATTERNS OF $t'$ AND $b'$; A 4G2HDM CASE

As mentioned above, in the 4G2HDM the 4th generation quarks have a more complex decay pattern that stems from their interactions with the extended Higgs sector. In addition, the SM4 forbidden channels, $t' \rightarrow Wb'$ and $b' \rightarrow W't'$, are no longer in contradiction with the EWPD, and may be kinematically open as well. For simplification, we will assume that $h$ is the only neutral scalar light enough to take part in the decays of $t'$ and $b'$. We will thus consider the following decay modes:

1. $t' \rightarrow ht$ ($b' \rightarrow hb$).
2. $t' \rightarrow H^+ b$ ($b' \rightarrow H^{-} t$).
3. $t' \rightarrow Wb$ ($b' \rightarrow Wt$).
4. $t' \rightarrow Wb'$ ($b' \rightarrow W't'$).

We assume the decay mode $t' \rightarrow H^+ b'$ to be kinematically closed taking $M_{H^+}$ to be larger than the mass gap in the fourth generation doublet.

To complete the picture of $t'$ and $b'$ decays, we need to fully account for the light neutral and charged Higgs decays. For $h$ we have:

1. $h \rightarrow \bar{b}b$. This decay mode is prominent for $M_h < 130$ GeV.
2. $h \rightarrow W^+W^-, ZZ$. These modes can be suppressed in this model (as in any 2HDM model), due to the $\sin(\alpha - \beta)$ factor in the $hWW$ and $hZZ$ vertices.
3. $h \rightarrow gg/\gamma\gamma$. Both channels are loop induced. In the SM4, $h \rightarrow gg$ is the leading decay channel for $M_h < 150$ GeV and is about 10 times larger than the SM, while $h \rightarrow \gamma\gamma$ is about 10 times smaller than the SM value [61]. On the other hand, in the 4G2HDM, $h \rightarrow gg$ could be suppressed in some areas of the parameter space, as a result of a destructive interference between the heavy fermion loops due to extra factors in the corresponding Yukawa terms which depend on the angles $\alpha$ and $\beta$ (see Eq. [3]). This would lead to either a suppression or enhancement of light Higgs production by gluon fusion as a function of $\alpha$, as is demonstrated in Figure 2. It is obvious that, even for the “popular” choice of $\alpha = \beta$, the enhancement of Higgs production is only by a factor of 2, which is almost 5 times smaller than the corresponding enhancement in the SM4. This is an example of how an extended Higgs sector may suppress the Higgs production signal (see also [59, 71]), providing a possible explanation of the absence of an observed signal [58]. Similarly, $h \rightarrow \gamma\gamma$ could be enhanced or suppressed in the 4G2HDM with respect to the SM4 or the SM [71, 72]. These issues are beyond the scope of this work and therefore will not be discussed here.

4. $h \rightarrow \nu_\tau\nu_\tau'/\tau'^\tau/\bar{t}t$. For the purposes of this work, we assume that $M_h < 2m_{\nu_{\tau'}}/2m_{\tau'}/2m_t$ allowing us to ignore these channels.

The $H^+$ decays are the following:

1. $H^+ \rightarrow tb$
2. $H^+ \rightarrow t'b$
3. $H^+ \rightarrow hW^+$
4. $H^+ \rightarrow \tau\nu_\tau$ and $H^+ \rightarrow \tau'^\tau/\nu_\tau'$.

The decay modes for $H^+$ that involve the first two generations are much smaller than the above ones, due their light masses. Combining all these channels to a cascade decay, we are left with several topologies for the $t'$ signatures, with different decay chains for each possibility:

1. $t' \rightarrow W + b$ (SM4-like)
2. $t' \rightarrow W + 3b$
   - $t' \rightarrow ht \rightarrow [\bar{b}b]_h [W^+b]_t$
   - $t' \rightarrow H^+b \rightarrow [\bar{b}b]_{H^+} b \rightarrow [(W^+b)_t]_{H^+} b$
   - $t' \rightarrow W^+b' \rightarrow W^+ [hb]_{\nu'} \rightarrow W^+ [(\bar{b}b)_{H^+}]_{\nu'}$
3. $t' \rightarrow 3W + b$
   - $t' \rightarrow ht \rightarrow [W^+W^-]_h [W^+b]_t$
   - $t' \rightarrow H^+b \rightarrow [W^+W^-]_{H^+} b \rightarrow [(W^+W^-)_h W^+]_{H^+} b$
   - $t' \rightarrow W^+b' \rightarrow W^+ [hb]_{\nu'} \rightarrow W^+ [(W^+W^-)_{H^+} b]_{\nu'}$

While for $b'$ decays we have:

1. $b' \rightarrow 3b$
   - $b' \rightarrow \bar{b}b \rightarrow [\bar{b}b]_h b$
2. $b' \rightarrow 2W + b$ (SM4-like)
   - $b' \rightarrow \bar{b}b \rightarrow [W^+W^-]_h b$
   - $b' \rightarrow W^-t \rightarrow [W^+b]_t W^-$ (SM4- Like)
   - $b' \rightarrow W^-t' \rightarrow [W^+b]_{t'} W^-$
3. $b' \rightarrow 4W + b$
   - $b' \rightarrow H^-t \rightarrow [W^-W^-]_{H^-} [W^+W^-]_t \rightarrow [(W^+W^-)_{H^-} W^-]_{H^-} [W^+b]_t$
   - $b' \rightarrow t'W^- \rightarrow [ht]_{t'} W^- \rightarrow [(W^+W^-)h (W^+b)_{t'}]_{t'} W^-$
   - $b' \rightarrow W^- \rightarrow [H^+b]_{t'} W^- \rightarrow [(W^+W^-)_{H^+} b]_{t'} W^- \rightarrow [(W^+W^-)_{H^+} b]_{t'} W^-$
4. $b' \rightarrow 2W + 3b$

- $b' \rightarrow H^- t \rightarrow [hW^-]_{H^-} [W^+ b]_t \rightarrow [(tb)_{H^-} W^-]_{H^-} [W^+ b]_t$
- $b' \rightarrow t'W^- \rightarrow [ht]_{t'} W^- \rightarrow [(tb)_{W^+} b]_{t'} W^-
- $b' \rightarrow t'W^- \rightarrow [hW^+]_{t'} W^- \rightarrow [(hW^+)_{H^+} b]_{t'} W^- \rightarrow [(tb)_{H^+} b]_{t'} W^-
- $b' \rightarrow t'W^- \rightarrow [hW^+]_{t'} W^- \rightarrow [(hW^+)_{W^+} b]_{t'} W^- \rightarrow [(tb)_{W^+} b]_{t'} W^-$

The decay chains not listed here, e.g. decay chains with Z bosons or fourth generation leptons, may further complicate the phenomenology of $t'$ and $b'$ and will not be considered in this work.

The above $t'$ and $b'$ decay channels give rise to a set of new signatures for $t'$ and $b'$ pair production at the LHC which we will generically denote by $pp \rightarrow t't'/b'b' \rightarrow n_W W + n_b b$, with $n_W$ and $n_b$ being the number of $W$ and $b$ jets in the event, respectively. In particular, the $t't'$ signatures are $2W+2b$ (SM-like), $2W+6b$, $6W+2b$, $2W+4b$, $4W+2b$ and $4W+4b$. As was mentioned earlier, for these signatures there is an excess of jets, leptons and $b$-jets with respect to the SM and the SM4 signals, as is demonstrated in Figure 3.

We stress again that, in general, these signatures are not exclusive to the 4G2HDM; they can appear in any model consisting of new (neutral or charged) scalars in the reasonable scenario that they interact mostly with the $3^{\text{rd}}$ and $4^{\text{th}}$ generation fermions.

V. RESULTS: SEMILEPTONIC CHANNEL

For the semileptonic channel ($pp \rightarrow t't' \rightarrow 1\ell + n_j + E_T$), we suggest a general reconstruction method that is based on [63] and test the extent to which the new signatures are already excluded by the "old" CMS analysis (which gives $m_{t'} > 450$ GeV for the SM4) [62]. We focus our study bellow on the new $pp \rightarrow t't' \rightarrow 2W + 6b$ or $6W + 2b$ signals, and show that, in those cases, $t'$ can be detected at the LHC using the excess of jets and $b$-jets mentioned above. The analysis is performed for two choices of parameter space in the 4G2HDM for which $B_r (t' \rightarrow th) \sim \mathcal{O} (1)$, but arguably, due to the fact that no care is given to the specifics of each decay chain, our conclusions hold for any other model or parameter space that results in the same signature (in the previous section, we showed that for each signature there are several decay chains in the 4G2HDM with different particles in the intermediate states but with the same final state).

We use Madgraph/MadEvent [62] to generate the signal and two $\int L dt = 1 \text{ fb}^{-1}$ sets of background events: $W$+jets and $t\bar{t}$+jets, with $K$-factors of 1.5 for the signal and the $t\bar{t}$+jets background [63], and 1.3 for $W$+jets background [64]. We use MLM parton-jet matching method [64] for the background with $p_T^{\min} = 100$ GeV for the $t\bar{t}$+jets and $p_T^{\min} = 150$ GeV for $W$+jets (see [64]). We use BRIDGE [65] for the decay of the new particles and Pythia [66] for the decay of the SM particles, shower, fragmentation and hadronizations. We use PGS [67] with the LHC card for the detector simulation.

We look for $1\ell + n_j + E_T$ signatures, requiring exactly one muon or electron with $p_T > 20$ GeV, $E_T > 20$ GeV, and at least four jets with $p_T > 120, 90, 35, 35$ GeV and $|\eta| < 2.5$

For the signal, we choose $m_{t'} = 350, 400, 450$ GeV, $\tan \beta = 1$, $V_{34} = 0.1$, and two different values for the Higgs mass, $M_h = 130, 170$ GeV [1], while the other scalars are assumed to be heavier and thus play no role in the decays. For both values of $M_h$ the main decay mode is $t' \rightarrow th$ with a branching ratio of $\sim 0.9$. For $M_h = 170$ GeV, $h$ decays mainly to $W^+W^-$, while $h \rightarrow b\bar{b}$ for $M_h = 130$ GeV; this leaves us with the two aforementioned signatures: $6W + 2b$ and $2W + 6b$, respectively.

Firstly, we demonstrate in Figure 3 the expected excess of jets and $b$-jets for the $6b+2W$ signature. We would not use the jet and $b$-jet distribution as a signal, as that might underestimate the background, especially the high multiplicity one ($t\bar{t} + n_j$ with $n > 2$). Instead, we use the demonstrated excess as a basis for one of the cuts in the selection of the signal. When enough low multiplicity background events survive the cut, the addition of the high multiplicity background events will not play a dramatic role in the $S/B$ ratio and will not change our conclusions.

For the same signature, we compare $M_{fit}$ (the reconstructed mass with the SM4 hypothesis) and $M_{gen}$ (the general reconstruction mentioned earlier, see Eqn. [3] in Figure 4) while in Figure 5 we plot the distribution of $M_{gen}$ for the signal ($2W+6b$) and background events. We can see in Figure 4 that the peak of the distribution of $M_{gen}$ is roughly

[1] Although $M_h \approx 125$ GeV seems more realistic in view of the recent hints from the LHC, we fix $M_h = 130, 170$ GeV for demonstration purposes.
FIG. 3: Number of jets (left) and number of $b$-jets (right) histograms for the $pp \rightarrow t't' \rightarrow 2W + 6b$ signature (red) with $m_{t'} = 450$ GeV. Also shown are the backgrounds from $W$+jets (green) and $tt$+jets (blue) for a set of 7 TeV LHC events with $\int L dt = 1$ fb$^{-1}$, in the semileptonic channel ($1\ell + n_j + \not{E}_T$).

FIG. 4: Comparison between $M_{\text{fit}} = m(l\nu b) = m(q\bar{q}b)$ (the reconstructed $t'$ mass using the CMS method, see Eq. 2) in blue and $M_{\text{gen}} = m(\text{Left Side}) = m(\text{Right Side})$ (the reconstructed $t'$ mass using our method, see Eq. 3) in red, for the $pp \rightarrow t't' \rightarrow 2W + 6b$ signature with $m_{t'} = 450$ GeV at the LHC (7 TeV) with $\int L dt = 1$ fb$^{-1}$, in the semileptonic channel ($1\ell + n_j + \not{E}_T$). See also text.

At $m_{t'}$, while for $M_{\text{fit}}$ the peak is misplaced, being substantially lower. We also note that the distribution of $M_{\text{gen}}$ in the new signatures is much broader around the peak than the distribution of $M_{\text{fit}}$; this is due to the massive combinatorial background coming from the high number of jets in the final state. In Figure 6 we plot the distribution of $H_T$ for the signal and the background events for the same signature.

Based on the above, we test the sensitivity of detection in our method for the two signatures ($6W+2b$ and $2b+6W$) for the $3 t'$ mass values, $m_{t'} = 350, 400, 450$ GeV, using the following cuts:

1. $M_{\text{gen}} > 300$ GeV
2. $H_T > 600$ GeV
3. $n_{j+b_j} = n_{\text{jets}} + n_{b-jets} > 6$

The results are summarized in Table I. We see that the leading background is $tt$+jets, as $W$+jets almost vanishes when imposing $n_{\text{jets}} + n_{b-jets} > 6$. For $m_{t'} = 350$ GeV we get a statistical significance of $\frac{S}{\sqrt{B}} \sim 4.7 - 5.7$, depending on the signature. For higher values of $m_{t'}$ the statistical significance drops, e.g., to the level of $\sim 2.3 - 2.9$ for the
two signatures when $m_{t'} = 450 \text{ GeV}$. In contrast, applying the "standard" CMS method (i.e. using $M_{fit}$ instead of $M_{gen}$ and requiring that at least one of the jets is tagged as a $b$-jet instead of the $n_{j+bj}$ cut), we get $\frac{S}{\sqrt{B}} \sim 2.6$ for $m_{t'} = 350 \text{ GeV}$ and $\frac{S}{\sqrt{B}} \sim 1.7$ for $m_{t'} = 450 \text{ GeV}$.

We proceed to test the consistency of the CMS exclusions results in the semileptonic channel (which exclude $t' \rightarrow Wb$ as is the case in the SM4) with the existence of a BSM4 signal in the current data. We recall that the CMS analysis is based on a 2d ($M_{fit}$ and $H_T$) likelihood fit to the data under the hypothesis of signal (SM4) + background and the hypothesis of background only, and a CLs method [73, 74] for the computation of the exclusion limits. They found that $m_{t'} > 450 \text{ GeV}$ at 95% CL (for the SM4 case) [52].

We perform a simplified computation of what would the CMS exclusion limit be for $m_{t'}$ in the case of background only and in the case of background+BSM4. To do so, we generate 1000 LHC events ($\int Ldt = 1 \text{ fb}^{-1}$) for each case. We then use the ratio $L = L_b/L_{s+b}$ of the likelihoods to the signal (SM4) + background hypothesis ($L_{s+b}$) and to the background hypothesis ($L_b$), as the test statistic. We naively assign a 20% systematic uncertainty on the background cross-section, minimizing the likelihood with respect to this parameter in each case. We calculate the likelihood ratio ($L$) for each simulated signature, and determine the probability to observe a higher value of $L$ for the background only hypothesis ($CL_b$) and the signal+background hypothesis($CL_{s+b}(\sigma)$). The 95% CL limit on the signal cross-section is then the value of the signal cross-section $\sigma$, for which

$$CL_s = \frac{CL_{s+b}(\sigma)}{CL_b} = 0.05.$$ (7)
Process | without cuts $n_{j+bj} \ M_{gen} \ H_T$ | all cuts combined $\frac{S}{\sqrt{B}}$
--- | --- | ---
$t\ell+jets$ | 1206 | 179 | 536 | 678 | 93
$W+jets$ | 626 | 9 | 353 | 443 | 5
$2W+6b \ m_{t'} = 350 \ GeV$ | 172 | 99 | 93 | 138 | 56
$2W+6b \ m_{t'} = 400 \ GeV$ | 119 | 74 | 69 | 104 | 45
$2W+6b \ m_{t'} = 450 \ GeV$ | 70 | 46 | 43 | 65 | 29
$6W+2b \ m_{t'} = 350 \ GeV$ | 168 | 80 | 87 | 135 | 46
$6W+2b \ m_{t'} = 400 \ GeV$ | 113 | 58 | 60 | 99 | 34
$6W+2b \ m_{t'} = 450 \ GeV$ | 77 | 43 | 40 | 74 | 23

TABLE I: Number of events after the selection (in the second column), after each of the three cuts: $n_{j+bj} \equiv n_{jets} + n_{b-jets} > 6$, $M_{gen} > 300 \ GeV$, $H_T > 600 \ GeV$ (3rd - 5th columns) and all of them combined (6th column). $\frac{S}{\sqrt{B}}$ for each of the new signatures at the LHC (7 Tev) with $\int L dt = 1 \ fb^{-1}$ is given in the seventh column.

FIG. 7: The 95% CL exclusion plot distribution on the $t'$ mass assuming the SM4 signature in the semileptonic channel ($t\ell + nj + H_T$). For the case of background only, the red dotted line is the median and the yellow and green bands are the $\pm 1$ and $\pm 2$ standard deviations accordingly. The black line is the median for background + $t't' \rightarrow 2b + 6W$ with $m_{t'} = 450 \ GeV$ and the blue line is the median for background + $t't' \rightarrow 6b + 2W$ with $m_{t'} = 350 \ GeV$. The results for the rest of the signatures with $m_{t'} = 350 - 450 \ GeV$ lie between these two lines.

Finally, we calculate the distribution of the 95% CL limit on the signal cross-section for background only and for background+BSM4 hypothesis, and plot in Figure 7 the exclusion limits for background only, and for background with two chosen BSM4 signatures for $t'$ pair production: $t't' \rightarrow 2b + 6W$ with $m_{t'} = 450 \ GeV$ and $t't' \rightarrow 6b + 2W$ with $m_{t'} = 350 \ GeV$. The exclusion limits for all the other BSM4 signatures that we’ve checked ($6b+2W,2W+6b$ with $m_{t'} = 350,400,450 \ GeV$) are between these two limits for the chosen signatures. We observe that for $m_{t'} = 350 \ GeV$, the median of the exclusion limit distribution for the BSM4 signatures is within 2 standard deviations of the expected limit for the background only case, and for $m_{t'} = 450 \ GeV$ it is fully within 1 standard deviation of the expected limit for the background only case. Thus, the discrepancy between the CMS observed exclusion limit, which is close to the expected one, see [52] (the expected limit in our naive calculation is denoted by the red dotted line in Figure 7), and the aforementioned BSM4 scenarios is less than 2$\sigma$ for $m_{t'} \gtrsim 350 \ GeV$, and 1-2$\sigma$ for $m_{t'} = 350 \ GeV$. We, therefore, go on to conclude that BSM4 scenarios with $m_{t'} > 350 \ GeV$, can be consistent with the current semileptonic data and therefore the limit on the $t'$ mass for the BSM4 case should be around 350 GeV in the semileptonic channel.
The most stringent bounds to this date on the $t'$ mass ($557 \text{ GeV}$) are set by the CMS, analyzing the dilepton channel ($pp \rightarrow t't' \rightarrow 2W + 2b$) for $t'$ pair production and decay at the LHC [53]. The signal in this channel is relatively rare (about 11% of the $t't'$ events) but clean due to the presence of the two leptons that suppresses the SM background. The suppression of the background is further promoted by requiring that exactly two of the jets in each selected event are tagged as $b$-jets. The kinematic variable that is used to discriminate the signal from the huge $t\bar{t}$ background is the invariant mass of one of the leptons and one of the $b$-jets, $M_{\text{min}}^{lb}$, where the configuration with the lowest invariant mass is chosen. In the case of $t\bar{t}$ production this invariant mass is bounded from above by the top mass ($170 \text{ GeV}$), while for the $t't'$ signal, there should be a large portion of events giving $M_{\text{min}}^{lb} > 170$. This leaves the region above this mass (named the "Signal Region") almost background-free (except mostly misidentification background) and completely signal-dominated. The number of events in the Signal Region is therefore used for the limit calculation.

When the new textures of $t'$ decay with higher multiplicities of leptons and $b$-jets are considered, the dominance of the signal for $M_{\text{min}}^{lb} > 170 \text{ GeV}$ may no longer be realized, due to the combinatorial background and the altered kinematic shape of the events. We try to address this issue using a simulation similar to the one in the previous section, with different selection rules. We generate $t't'$ events with $5 \text{ fb}^{-1}$ for the SM4-like $2W+2b$ signature and for the new $2W+6b$ signature, choosing $m_{t'} = 350 \text{ GeV}$ and selecting events with at least two leptons and exactly two $b$-jets, with the kinematic cuts that were used in the CMS analysis [53]. Due to the fact that the $b$-tagging efficiency in PGS [67] is not well calibrated with the efficiency of the TCHEM method of $b$-tagging [75] that was used in the CMS analysis, we get a lower number of events for the SM4-like signature than the number of events reported in [53]. We thus use a higher efficiency to roughly simulate the correct number of events in the Signal Region. Thus, our results below should be considered only as an illustration of the effect of BSM4.

The results are presented in Figure 8. We first note, that the total number of selected events is lower for the $2W+6b$ than for the SM4 case. This is due to the fact that there is a higher number of $b$ jets (6) in $2W+6b$ events, thus, for high enough $b$-tagging efficiency (above 0.5) there is a lower chance of tagging exactly 2 of them. In the signal region, the number of the events for the $2W+6b$ signature is negligible- about 1/10 of the events in SM4 case. In fact, for $m_{t'} > 350 \text{ GeV}$, the number of $2W+6b$ events in the Signal Region is even lower due to the lower production cross-section. We therefore conclude that the current stringent limit on the $t'$ mass ($557 \text{ GeV}$) in the dilepton channel is completely irrelevant for some BSM4 scenarios, for $m_{t'} > 350 \text{ GeV}$ (we recall that $m_{t'} = 350 \text{ GeV}$ is the lowest mass considered in [53]).
we have explored new detection strategies for the $t'$ that hold in the general case, where new decay patterns (beyond the SM4) for $t'$ emerge, e.g. $t' \to th$. In such cases, new $t'$ signals, such as $pp \to t\bar{t}' \to 6W + 2b$ and $pp \to \bar{t}t' \to 2W + 6b$ that we have studied here may dominate. As a representative framework for BSM4 we chose the 4G2HDM of [33], which is a variation of a 2HDM framework with 4 generations. We then studied its implications for $t'$ phenomenology at the LHC, keeping the analysis as general as possible and, therefore, model-independent to a large extent.

We have shown that if the new $t'$ decay channels lead to $pp \to \bar{t}t' \to 2W + 6b$ or to $pp \to \bar{t}t' \to 6W + 2b$, then the current limits on the $t'$ mass (450/557 GeV for the semileptonic and the dilepton channel in the SM4 case) do not apply and a different detection strategy is required. For the semileptonic channel, we estimate the current limit on the $t'$ mass in the BSM4 case to be around 350 GeV, demonstrating that a general reconstruction method (without the assumption of a specific topology of each event) is successful in pushing the limit upwards.

In the dilepton channel, we find the CMS limit to be completely irrelevant for the mass range in the analysis (350-600 GeV) for several BSM4 scenarios. This includes the case that the mass of the lightest Higgs is around 130 GeV, which may be consistent with what ATLAS and CMS see in the Higgs searches. We leave for a future study the constraints on new scenarios resulting from the $bt$ searches at the LHC [54, 57].

Acknowledgments: The authors would like to thank Amarjit Soni for useful discussions. GE thanks P.Q. Hung for helpful discussions and SBS acknowledges research support from the Technion. MG thanks Itah Galon for his help with the numerical analysis.

After completing our work we received a preprint by K. Rao and D. Whiteson, [arXiv:1204.4504 [hep-ph]) in which ideas similar to ours are expressed.

[1] G. Aad et al. [ATLAS Collaboration], Phys. Lett. B 710, 49 (2012).
[2] S. Chatrchyan et al. [CMS Collaboration], [arXiv:1202.1485 [hep-ex]].
[3] [TEVNPH (Tevatron New Phenomena and Higgs Working Group) and CDF and D0 Collaborations], arXiv:1203.3774 [hep-ex].
[4] E. Lunghi and A. Soni, Phys. Lett. B666, 162 (2008).
[5] E. Lunghi and A. Soni, JHEP 0908, 051 (2009).
[6] A. Lenz et al. [CKMfitter Group], Phys. Rev. D 83, 036004 (2011).
[7] M. Bonan et al. [UTfit Collaboration], Phys. Lett. B 687, 61 (2010) and A. J. Bevan et al. [UTfit Collaboration], PoS ICHEP 2010, 270 (2010).
[8] E. Lunghi and A. Soni, Phys. Rev. Lett. 104, 251802 (2010).
[9] E. Lunghi and A. Soni, Phys. Lett. B 697, 323 (2011).
[10] D. Asner et al. [Heavy Flavor Averaging Group Collaboration], arXiv:1010.1589 [hep-ex].
[11] V. M. Abazov et al. [D0 Collaboration], Phys. Rev. Lett. 105, 081801 (2010).
[12] Frederic Teubert et al [CMS and LHCb Collaborations], LHCb-CONF-2011-055; CERN-LHCb-CONF-2011-055, EPS-HEP 2011, July 21-27, 2011; Grenoble, Rhone-Alpes France.
[13] P.H. Frampton, P.Q. Hung and M. Sher, Phys. Rept. 330, 263 (2000).
[14] B. Holdom, W.S. Hou, T. Hurth, M.L. Mangano, S. Sultansoy, G. Unel, talk presented at Beyond the 3rd SM generation at the LHC era workshop, Geneva, Switzerland, Sep 2008, [arXiv:0904.4698 [hep-ph]], published in PMC Phys. A3, 4 (2009); S. A. Cetin, G. W. S. Hou, V. E. Ozcan, A. N. Rozanov and S. Sultansoy, Status of the Fourth Generation: A Brief Summary of BSM-III Workshop in Four Parts, arXiv:1112. 2907 [hep-ph].
[15] For older literature on the 4th generation SM, see: Proceedings of the First (February 1987) and the Second (February 1989) International Symposia on the fourth family of quarks and leptons, Santa Monica, CA, published in Annals of the New York Academy of Sciences, 517 (1987) & 578 (1989), edited by D. Cline and A. Soni.
[16] For a review of the implications of stable quarks of the 4th on cosmology, astrophysics and accelerator physics, see: K. Belotsky, M. Khlopov and K. Shibaev, arXiv:0806.1067 [astro-ph] and references therein.
[17] A. Soni, A. K. Alok, A. Giri, R. Mohanta and S. Nandi, Phys. Lett. B683, 302 (2010).
[18] A. Soni, A. K. Alok, A. Giri, R. Mohanta and S. Nandi, Phys. Rev. D82, 033009 (2010).
[19] A. J. Buras, B. Duling, T. Feldmann, T. Heidsieck, C. Promberger and S. Recksiegel, JHEP 1009, 106 (2010).
[20] A. J. Buras, B. Duling, T. Feldmann, T. Heidsieck, C. Promberger and S. Recksiegel, JHEP 1007, 094 (2010).
[21] W. S. Hou and C. Y. Ma, Phys. Rev. D82, 036002 (2010).
[22] S. Nandi and A. Soni, Phys. Rev. D 83, 114510 (2011).
[23] M. Bobrowski, A. Lenz, J. Riedl and J. Rohrwild, Phys. Rev. D 79, 113006 (2009).
[24] O. Eberhardt, A. Lenz and J. Rohrwild, Phys. Rev. D 82, 095006 (2010).
