

Sleuth at CDF

A quasi-model-independent search for new electroweak scale physics

Georgios Choudalakis^a

Massachusetts Institute of Technology, for the CDF collaboration

Abstract. These proceedings describe SLEUTH, a quasi-model-independent search strategy targeting new electroweak scale physics, and its application to 927 pb^{-1} of CDF II data. Exclusive final states are analyzed for an excess of data beyond the Standard Model prediction at large summed scalar transverse momentum ($\sum p_T$). This analysis of high- $\sum p_T$ data represents one of the most encompassing searches so far conducted for new physics at the energy frontier.

PACS. 12.60.-i Models beyond the standard model

1 Introduction

Searching a large subset of collision events recorded at the energy frontier for evidence of new physics requires understanding the Standard Model prediction and detector response. To facilitate understanding, debris from collision events are reduced to 4-vectors of reconstructed physics objects: electrons (e), muons (μ), taus (τ), photons (γ), jets (j), and jets tagged as containing a bottom quark (b). Existing event generators are used to construct the Standard Model prediction, and the response of the CDF detector to these events is estimated using a GEANT-based simulation of the detector. A correction model is developed to correct possible inadequacies in the Standard Model prediction and modeling of the detector response. These procedures form the basis of VISTA, described in a companion proceedings in this conference [1]. Having obtained a global Standard Model background estimate, SLEUTH is used to target new physics likely to appear as an excess of data beyond Standard Model prediction at large summed scalar transverse momentum. Similar search strategies have been developed previously at DØ [2, 3, 4] and H1 [5].

2 The Method

SLEUTH is slightly less model-independent search strategy than VISTA, as it invokes three additional assumptions. The three additional assumptions motivating SLEUTH are that new physics will likely appear

1. as an excess of data over the Standard Model expectation,
2. at large $\sum p_T$ with respect to the Standard Model background, and

3. predominantly in one SLEUTH final state.

SLEUTH is expected to be sensitive to any particular incarnation of new physics to the extent that the above criteria are satisfied.

Following these assumptions, SLEUTH searches for statistically significant excesses of data in the high- $\sum p_T$ tails of all high- p_T final states with non-zero Standard Model expectation, or where data are observed. Final states are defined so as to concentrate more of the signal of possible new physics. Final states equivalent under a global switch of the two lesser generations (such as e^+e^- and $\mu^+\mu^-$) are merged (into $\ell^+\ell^-$). Final states equivalent under global charge conjugation (such as $\ell^+\ell^+$ and $\ell^-\ell^-$) are also merged. Since new physics is expected to produce light quarks (j) and bottom quarks (b) in pairs, each event is said to contain n jet pairs if the number of reconstructed jets is $2n$ or $2n + 1$, and is said to contain m b -jet pairs if the number of reconstructed b -jets is $2m$, or $2m - 1$ and there is another jet in the event, for integer m and n .

SLEUTH examines the variable

$$\sum p_T = \left(\sum_{i \in \text{objects}} |\mathbf{p}_{Ti}| \right) + |\mathbf{p}_{T\text{uncl}}| + |\not{p}_T|, \quad (1)$$

where the first term is the scalar sum of the transverse momentum of all reconstructed objects, the second term is the transverse momentum of energy visible in the detector but not clustered into reconstructed objects, and the third term is the transverse component of any missing energy in the event. By definition, these quantities satisfy

$$\sum_i \mathbf{p}_{Ti} + \not{p}_T + \mathbf{p}_{T\text{uncl}} = \mathbf{0}, \quad (2)$$

a two-component vector equation in the spatial coordinates transverse to the axis of the colliding beams.

^a <http://www.mit.edu/~gchouda/>, gchouda@mit.edu

In addition to the fact that many new physics scenarios involve new massive resonances decaying into Standard Model objects resulting in events with large $\sum p_T$, the variable $\sum p_T$ is useful in practice since it is easily defined for any final state, it is fairly insensitive to soft physics, including the details of final state parton showering, and its expected distribution can be easily populated on the high tails with additional Monte Carlo events by running event generators with a (slightly lower) threshold on generated $\sum p_T$.

In each final state, SLEUTH considers regions stretching from each data point up to infinity in the distribution of $\sum p_T$. For each region, the Poisson probability (p -value) that the Standard Model background in that region would fluctuate up to or above the number of data events observed in that region is calculated. The most interesting region in each final state is the one with the smallest p -value. SLEUTH next produces ensembles of pseudo data events drawn from the $\sum p_T$ distribution of the Standard Model background, and for each set of pseudo data events finds the most interesting region and corresponding p -value. The fraction (\mathcal{P}) of this ensemble of pseudo experiments in which the most interesting region corresponds to a p -value smaller than that of the most interesting region observed in the actual data is determined. The value \mathcal{P} , a number ranging between zero (very interesting) and unity (not interesting), quantifies the probability that a random fluctuation of the Standard Model background would create any region more interesting than the most interesting region observed in the data in that particular final state. In each final state considered individually, \mathcal{P} is the relevant measure of the statistical significance of the most interesting region on the tail of the $\sum p_T$ distribution. This process is repeated for all SLEUTH final states populated by three or more data events.

The most interesting final state is identified as the final state with smallest \mathcal{P} , denoted by \mathcal{P}_{\min} . An ensemble of pseudo experiments involving all final states is then used to find the fraction ($\tilde{\mathcal{P}}$) of these pseudo experiments that would produce one or more final states more interesting than the most interesting final state in the data. The number $\tilde{\mathcal{P}}$, which ranges between zero (very interesting) and unity (not interesting), quantifies the interest of the final state with the most interesting region, accounting for the trials factor associated with having examined many final states.

SLEUTH's discovery threshold is $\tilde{\mathcal{P}} \lesssim 0.001$. The trials factor is such that $\tilde{\mathcal{P}}$ at this threshold typically corresponds to a region with p -value of about 10^{-7} , motivating the default criterion of 5σ typically employed in our field.

3 Sensitivity

Sensitivity studies are performed to estimate SLEUTH's sensitivity to possible new physics in 927 pb^{-1} of CDF data. Monte Carlo events corresponding to particular models of new physics, including specific parameter

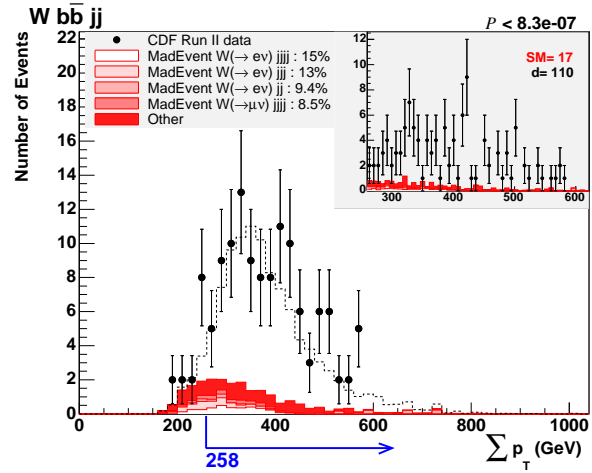


Fig. 1. The SLEUTH $Wb\bar{b}jj$ final state is dominated by Standard Model $t\bar{t}$ production. The total Standard Model prediction including $t\bar{t}$ production (dashed line) is in agreement with observed CDF data (black points). If the top quark were unknown, the Standard Model prediction *sans* $t\bar{t}$ production (red shaded histogram) significantly underestimates the observed data. A global fit of the VISTA correction factors [1] has been unable to coerce the Standard Model prediction without $t\bar{t}$ to match the data in this final state. The (blue) arrow indicates the most interesting region, identified by SLEUTH. The interest of this region, considering this final state alone, is found to be $\mathcal{P} < 8.3 \times 10^{-7}$ (upper right), corresponding to a value of $\tilde{\mathcal{P}}$ easily satisfying SLEUTH's discovery threshold of 0.001.

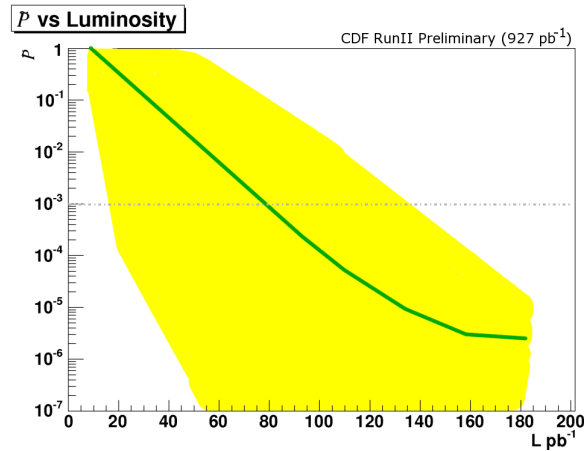


Fig. 2. SLEUTH's $\tilde{\mathcal{P}}$ as a function of assumed integrated luminosity, where top quark pair production is removed from the Standard Model background estimate. The horizontal axis shows integrated luminosity, in units of pb^{-1} . The shaded (yellow) band shows the range of values of $\tilde{\mathcal{P}}$ obtained when the integrated luminosity of the CDF data sample is reduced, with the width of the band resulting from individual CDF top quark candidate events being randomly retained or discarded. The median $\tilde{\mathcal{P}}$ obtained is shown as the solid (green) line. The SLEUTH discovery threshold $\tilde{\mathcal{P}} \lesssim 0.001$ is shown as the horizontal dashed (gray) line.

points within the MSSM and several flavors of Z' , are partitioned into SLEUTH final states according to the objects reconstructed in these events, and are added to pseudo data drawn from the Standard Model prediction. SLEUTH is run “blind” on the resulting mix of pseudo data and trace amounts of pseudo signal, not knowing anything about the pseudo signal that has been added. The amount of pseudo signal is gradually increased until SLEUTH identifies a region of interest corresponding to $\tilde{\mathcal{P}} < 0.001$.

The models tested require a cross section of roughly a few picobarns to reach SLEUTH’s discovery threshold in 927 pb^{-1} of data. SLEUTH’s sensitivity is found to be comparable to that of targeted searches for models that sufficiently satisfy the three assumptions on which SLEUTH is based. SLEUTH’s sensitivity to unspecified new physics is in general significantly greater than any specific targeted search, due to SLEUTH’s noticeably broader scope.

Specific Standard Model processes are also used to examine SLEUTH’s sensitivity. If the top quark were unknown, its production at the Tevatron would have been caught using SLEUTH. Observation of $t\bar{t}$ production in 927 pb^{-1} is shown in Fig. 1. An integrated luminosity of roughly 80 pb^{-1} would be required to reach the SLEUTH discovery threshold¹, as shown in Fig. 2. Standard Model WW production is also easily observed. SLEUTH has more difficulty with single top quark and Higgs boson production, which only partially satisfy SLEUTH’s assumption that new physics will appear at large $\sum p_T$ relative to (other) Standard Model processes.

4 Results

In 927 pb^{-1} of CDF Run II data, the most interesting region is observed in the $b\bar{b}$ final state, shown in Fig. 3. The fraction of pseudo experiments in this final state alone that would produce a region more interesting than the one observed is $\mathcal{P} = 0.0055$. Taking into account the many final states considered, the fraction of hypothetical similar CDF experiments that would observe a final state with $\mathcal{P} < 0.0055$ is $\tilde{\mathcal{P}} = 0.46$.

Unfortunately, with $\tilde{\mathcal{P}} = 0.46 \gg 0.001$, SLEUTH has not identified any final state containing a region at large $\sum p_T$ with a significant excess of data above Standard Model prediction.

5 Conclusion

These proceedings have described SLEUTH, a quasi-model-independent search algorithm designed to identify an excess of data above Standard Model prediction at large summed scalar transverse momentum in

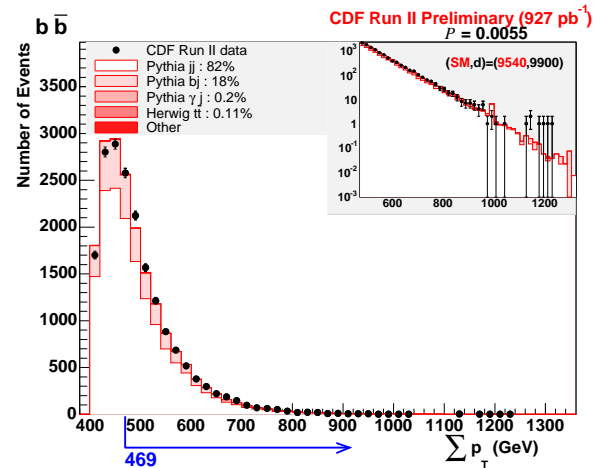


Fig. 3. The most interesting SLEUTH final state ($b\bar{b}$). Filled circles (black) show CDF Run II data; the (red) histogram shows the standard model prediction, with jj standing for two non- b generated partons, and bj for at least one generated b -quark. The horizontal axis shows the summed scalar transverse momentum of all objects in each event in the histogram, with the number of events per 20 GeV bin shown on the vertical axis. The region SLEUTH determines to be most interesting is shown with the (blue) arrow to be the region corresponding to $\sum p_T > 469 \text{ GeV}$. A magnified view of this region in logarithmic scale is shown in the inset, together with the number of events predicted (SM) and the number of events observed (d). Taking into account the trials factor associated with looking at many different regions, SLEUTH determines the fraction of hypothetical similar experiments that would produce something in this final state as interesting as the region shown to be $\mathcal{P} = 0.0055$ (number at upper right).

any final state. A specific definition of regions considered allows SLEUTH to rigorously calculate the trials factor associated with looking in many different places. Sensitivity tests performed indicate SLEUTH is broadly comparable to targeted searches for specific models adequately satisfying the three assumptions on which it is based. SLEUTH is expected to be significantly more sensitive to unspecified new physics than any particular targeted search, due to SLEUTH’s noticeably broader scope. Application of SLEUTH in association with VISTA [1] to the first 927 pb^{-1} of CDF Run II data has unfortunately revealed no indication of new physics. This analysis represents one of the most encompassing searches so far conducted for new physics at the energy frontier.

This result does not prove that no new physics is hiding in Tevatron data; merely that this particular global analysis strategy has not yet revealed a discrepancy on which a discovery claim can be based in the 927 pb^{-1} so far analyzed. A global analysis of more recently collected CDF data is in progress.

References

1. G. Choudalakis, companion (SUSY 07) proceedings.
2. D0 collaboration, Phys. Rev. Lett. 86 (2001) 3712.

¹ When comparing to the 67 pb^{-1} needed for discovery in CDF Run I [6], differences such as the improved CDF II detector should be kept in mind.

3. D0 collaboration, Phys. Rev. D62 (2000) 092004.
4. D0 collaboration, Phys. Rev. D64 (2001) 012004.
5. H1 collaboration, Phys. Lett. B602 (2004) 14.
6. CDF collaboration, Phys. Rev. Lett. 74 (1995) 2626.