

# Signature-Based Global Searches at CDF

Andy Hocker \*

Fermi National Accelerator - Technical Division  
PO Box 500, Batavia, IL 60510 - USA

Data collected in Run II of the Fermilab Tevatron are searched for indications of new electroweak scale physics. Rather than focusing on particular new physics scenarios, CDF data are analyzed for discrepancies with respect to the Standard Model prediction. Gross features of the data, mass bumps, and significant excesses of events with large summed transverse momentum are examined in a model-independent and quasi-model-independent approach. This global search for new physics in over three hundred exclusive final states in  $2 \text{ fb}^{-1}$  of  $p\bar{p}$  collisions at  $\sqrt{s} = 1.96 \text{ TeV}$  reveals no significant indication of physics beyond the Standard Model.

## 1 Introduction

The Standard Model as presently established by experiment is both enormously successful and manifestly incomplete. At the very least, the origin of electroweak symmetry breaking (EWSB) remains an open question with various possible extensions of the Standard Model offering a solution. Even the “minimal” resolution of EWSB via the Higgs mechanism invites a host of new physics models seeking to eliminate the need for fine tuning corrections to the Higgs mass. Within each class of models are a large number of submodels with large numbers of free parameters, and the possible “model space” in which one could conceive of searching for new physics quickly becomes intractably large (with the additional unfortunate possibility that no currently proposed model is actually correct).

With this in mind, CDF has developed a global search for new physics that is not beholden to any particular model [1]. The overall idea is to systematically search the entirety of the CDF high- $p_T$  dataset for any discrepancies with respect to the Standard Model [2]. This approach essentially casts as wide a net as possible in the search for new phenomena and complements searches that are targeted at the predictions of specific models.

A two-tiered approach is taken for the global search. The first tier is called VISTA and provides a “panoramic view” of the CDF high- $p_T$  dataset. The goal is to determine whether the Standard Model and the CDF detector simulation can describe gross features of the data, namely the populations of exclusive final states and basic event kinematics. The second tier (“Bump Hunter” and “SLEUTH”) then performs a more focused examination of the data. The Bump Hunter [3] goes through each final state, clustering objects and searching for resonances in their invariant mass distributions. SLEUTH searches for new EWSB-scale physics appearing predominantly in one final state by looking for excesses in the high- $p_T$  tails of the summed scalar  $p_T$  distribution ( $\Sigma p_T$ ) for a final state.

## 2 VISTA

The VISTA algorithm begins by using standard CDF criteria to identify objects (electrons, muons, taus, photons, jets,  $b$ -jets, and missing transverse energy) with transverse momenta in excess of  $17 \text{ GeV}/c$ . The data are then filtered to retain events of interest by requiring,

---

\*on behalf of the CDF Collaboration

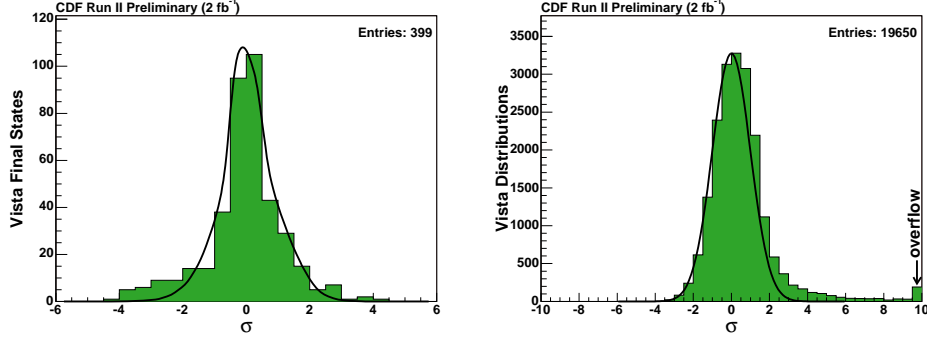


Figure 1: Deviation of the data with respect to the SM prediction expressed in units of  $\sigma$  (histograms) for the VISTA final state populations (left) and the VISTA kinematic distributions (right). The overlaid curves show the distributions expected assuming the null hypothesis (*i.e.* no new physics) and an accurate correction model.

for example, an electron with  $E_T > 25$  GeV or a  $b$ -jet with  $E_T > 25$  GeV accompanied by a 40 GeV photon (for a complete list, see [3]). The events are then partitioned into exclusive final states according to the number and type of objects they contain. For the  $2 \text{ fb}^{-1}$  dataset analyzed here, this partitioning results in 399 final states.

The SM expectation for each final state is determined from Monte Carlo samples generated for each contributing physical process and passed through the CDF detector simulation. Correction factors must be applied to these Monte Carlo/simulation-based predictions to account for deficiencies in the modelling of object identification and misidentification rates, as well as to appropriately scale the typically leading-order MC cross sections calculated. These correction factors are determined from a global fit to the high- $p_T$  data, in which a  $\chi^2$  based on the observed and predicted number of events in each final state is minimized. External constraints on the correction factors, for example from available NLO cross section calculations, are incorporated into the  $\chi^2$ .

After determining the correction factors, the observed final state populations are compared to the corrected Monte Carlo predictions in order to search for remaining significant discrepancies. In addition, the expected distributions of relevant event-level kinematic quantities are compared to those observed in the data; their level of agreement is quantified using the Kolmogorov-Smirnov statistic. The VISTA results are summarized in Figure 1. No significant ( $\sim 5\sigma$ ) outliers are observed in the final state populations, however there are a number of discrepant kinematic shapes. Closer examinations of these distributions [2] reveal that they largely stem from the same underlying problem, that of modeling the angle between lower- $p_T$  jets in multijet events (see Figure 2). As the resolution of this problem more likely lies in the details of parton shower modeling and not in new physics, these discrepancies are not pursued further.

### 3 Bump Hunter

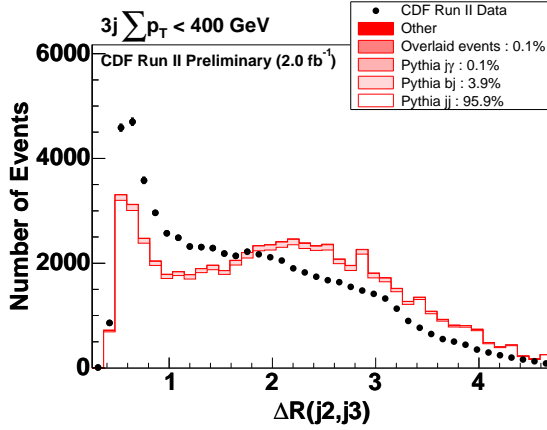


Figure 2: The angle between the second- and third-highest  $E_T$  jets in the VISTA 3-jet final state. The histogram is the prediction based on SM Monte Carlo and detector simulation; the points are the CDF data.

consequence of the jet-angle modeling problem discussed in the previous section and shown in Figure 2; therefore this bump does not motivate any new physics claim.

#### 4 SLEUTH

The SLEUTH algorithm sharpens the focus of the new physics search by making three assumptions, none of them particularly restrictive, on the form that new physics will take, namely that it will appear as an excess in the tails of the  $\Sigma p_T$  distribution, in predominantly one final state. VISTA final states of similar type are collapsed together (for example, electrons and muons are treated generically as “leptons”); in each of these SLEUTH final states the tail of the  $\Sigma p_T$  distribution is found that maximizes the excess of data with respect to Standard Model expectation. An ensemble of SM pseudoexperiments is then constructed to determine the probability (denoted by  $\mathcal{P}$ ) of the SM producing a tail as discrepant as that observed. The final state with the smallest  $\mathcal{P}$  ( $\mathcal{P}_{min}$ ) is the potential new physics signal, but first one must take into account the “trials factor:” the more final states that are searched, the more likely it is that one will exhibit a seemingly significant discrepancy. Therefore the quantity of interest is the probability that the SM alone could produce a final state (*any*

If a new massive particle with narrow width is produced and it decays to objects visible in the CDF detector, searching for a localized enhancement in an invariant mass spectrum is potentially more sensitive than a KS test on an entire distribution. To that end, all possible combinations of objects in each VISTA final state are clustered and the resulting invariant mass distributions are scanned with a sliding window of width  $2\Delta m$ , where  $\Delta m$  is the mass resolution determined from the expected detector energy resolutions for the  $E_T$ ’s and types of objects being clustered. The number of events observed and predicted within the window are determined and the significance of any excess is quantified. The most significant excess is shown in Figure 3; this is the invariant mass of all jets in the VISTA 4-jet final state. Closer study shows that the low- $p_T$  mass bump observed here is a direct

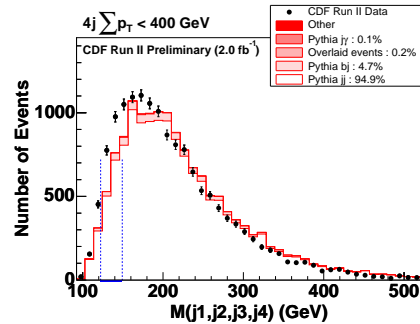


Figure 3: 4-jet invariant mass distribution in the VISTA 4-jet final state. The histogram is the prediction based on SM Monte Carlo and detector simulation; the points are the CDF data. The dotted lines show the mass window with the most significant excess.

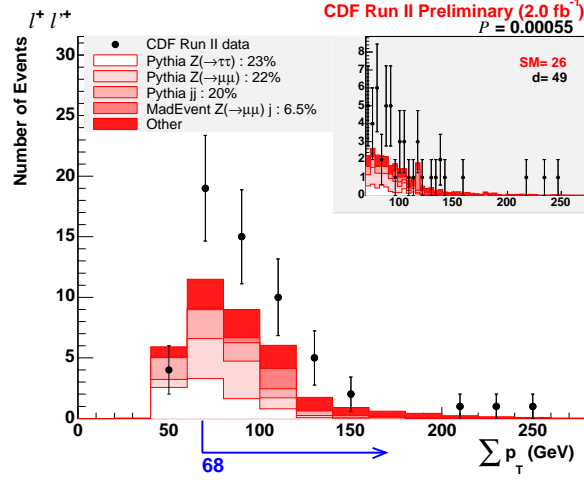


Figure 4:  $\Sigma p_T$  distribution for the SLEUTH  $e^\pm\mu^\pm$  final state. The histogram is the prediction based on SM Monte Carlo and detector simulation; the points are the CDF data. The arrow shows the region of interest selected by SLEUTH and the inset shows a zoom of this region.

final state) with  $\mathcal{P} < \mathcal{P}_{min}$ . This probability is denoted by  $\tilde{\mathcal{P}}$ ; a value of  $\tilde{\mathcal{P}} < 10^{-3}$  serves as the threshold for which a discovery claim can be made.

The application of SLEUTH to  $2 \text{ fb}^{-1}$  of high- $p_T$  CDF data results in  $\tilde{\mathcal{P}} = 0.08$ . As this is  $\gg 10^{-3}$ , no discovery claim can be made. The most discrepant final state is that containing an electron and a muon of the same electric charge; see Figure 4. Although the excess in this final state is not statistically significant, scientific curiosity leads one to examine it more closely. SM processes containing real like-signed dileptons contribute almost negligibly to this final state, leading one to consider the possibility of incorrect charge assignment. A large fraction of the excess contains electrons in the forward regions of the detector, where charge misassignment rates can be large. However, the modeling of this misassignment rate is well-constrained by the like-signed  $ee$  final state, which is dominated by the high-rate  $Z \rightarrow e^+e^-$  process.

## 5 Conclusions

A global search for new physics based not on any particular model but rather on a systematic search for discrepancies with respect to the SM in  $2 \text{ fb}^{-1}$  of CDF high- $p_T$  data has been performed. 399 exclusive final states and almost 20000 kinematic distributions have been searched; a number of interesting effects have been observed, but this search reveals no indication of physics beyond the Standard Model.

## References

- [1] Slides:  
<http://indico.cern.ch/contributionDisplay.py?contribId=107&sessionId=15&confId=24657>
- [2] T. Aaltonen *et al.* (CDF Collaboration), arXiv:hep-ex/0712.1311, accepted by Phys. Rev. D (2007).
- [3] G. Choudalakis, Ph.D. thesis, Massachusetts Institute of Technology, arXiv:hep-ex/0805.3954 (2008).