Global Search for New Physics with 2.0 fb$^{-1}$ at CDF

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 the standard model prediction. A model-independent approach (VISTA) considers gross features of the data, and is sensitive to new large cross-section physics. Further sensitivity to new physics is provided by two additional algorithms: a Bump Hunter searches invariant mass distributions for "bumps" that could indicate resonant production of new particles; and the search for new physics in 2.0 fb^{-1} of pp collisions at \sqrt{s} = 1.96 TeV reveals no indication of physics beyond the standard model.
The standard model (SM) of particle physics has been remarkably successful in describing observed phenomena, but is generally believed to require expansion. Using data corresponding to an integrated luminosity of 2.0 fb$^{-1}$ of $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV collected by the CDF II detector at the Fermilab Tevatron, we present a broad search for physics beyond the standard model without focusing on any specific proposed scenario. A similar search has previously been performed by the CDF collaboration with 927 pb$^{-1}$ of data [1].

Events containing one or more particles with large transverse momentum ($p_T$) are analyzed for discrepancies relative to the SM prediction. A model-independent approach (VISTA) considers gross features of the data, and is sensitive to new large cross-section physics. Further sensitivity to beyond-SM physics is provided by two additional algorithms: a Bump Hunter searches invariant mass distributions for “bumps” that could indicate resonant production of new particles; and the SLEUTH procedure scans for data excesses at large summed transverse momentum. These global algorithms provide a complementary approach to searches optimized for more specific new physics scenarios.

CDF II [2] is a general-purpose detector for high-energy $p\bar{p}$ collisions. Tracking for charged particles is provided by silicon strip detectors and a gas drift chamber inside a 1.4 T magnetic field. The tracking system is surrounded by electromagnetic and hadronic calorimeters and enclosed by muon detectors.

The VISTA procedure is extensively described in [1]. A standard set of object identification criteria is used to identify isolated and energetic objects produced in the hard collision, including electrons ($e^\pm$), muons ($\mu^\pm$), taus ($\tau^\pm$), photons ($\gamma$), jets ($j$), jets originating from a bottom quark ($b$), and missing transverse momentum ($\not{p}_T$) [7]. All objects are required to have $p_T \geq 17$ GeV/c. With all event selections applied, over $4 \times 10^6$ high-$p_T$ events are analyzed in this global search. The standard model prediction is based on Monte Carlo event generators and a simulation of the response of the CDF detector. Data and Monte Carlo events are partitioned into exclusive final states labeled according to the number and type of objects ($e^\pm$, $\mu^\pm$, $\tau^\pm$, $\gamma$, $j$, $b$, $\not{p}_T$) identified in each event.

To obtain an accurate standard model prediction, a correction model is used to improve systematic deficiencies in the Monte Carlo theoretical prediction and the simulation of the detector response – this information can only be obtained from the data themselves. The details of this correction model are motivated by individual discrepancies noted in a global comparison of CDF high-$p_T$ data to the SM prediction; however, the correction model is intentionally kept as simple as possible in order to avoid over-tuning. The correction model includes specific correction factors for the integrated luminosity of the sample, the ratio ($k$-factor) of the actual cross section for an SM process and the leading order approximation given by event generators, object identification efficiencies, object misidentification rates, and trigger efficiencies. Values for the correction factors are determined from a global fit to the data: a global $\chi^2$ is formed by comparison to the SM prediction, and minimized as a function of the correction factors. External information (such as higher-order cross-section calculations) is used to constrain 26 of the 43 total correction factors. A number of minor improvements have been made to the correction model since [1]; these changes are described in detail in [3].

The first stage in the VISTA global comparison is to study the populations of the exclusive final states, comprehensively analyzed in this global search. The standard model prediction for populations of exclusive final states, measured in units of standard deviation ($\sigma$), is represented in Fig. 1. The black line represents the theoretical expectation assuming no new physics.

FIG. 1: Distribution of observed discrepancy between data and SM prediction for populations of final states, measured in units of standard deviation ($\sigma$). The black line represents the theoretical expectation assuming no new physics.

FIG. 2: Distribution of observed discrepancy between data and SM prediction for shapes of kinematic distributions, measured in units of standard deviation ($\sigma$). The black line represents the theoretical expectation assuming no new physics.

Virginia, Charlottesville, VA 22904, On leave from J. Stefan Institute, Ljubljana, Slovenia,
pared to the SM expectation. Fig. 1 summarizes the population discrepancies in all 399 final states, and the ten final states with the largest deviation from the SM expectation are listed in Table I. After accounting for the trials factor associated with considering many final states, we find that no final state exhibits a statistically significant population discrepancy.

The VISTA global comparison also considers the shapes of kinematic distributions. The Kolmogorov-Smirnov test is used to assess the agreement between data and the SM prediction for 19 650 distributions. The results are summarized in Fig. 2 which shows the degree of discrepancy, which is defined as being greater than 3σ, for each final state. This observation has been discussed in more detail in [1]. The nature of these shape discrepancies does not warrant treating any of them as indicative of potential new physics.

A statistically significant local excess of data in an invariant mass variable would be the most direct evidence of resonant production of a new particle. The Bump Hunter algorithm is designed to identify mass resonances with narrow natural width that would appear as Gaussian “bumps” on top of the SM background, with width equal to the detector resolution. The Bump Hunter searches in all exclusive final states, and examines all mass variables that can be constructed from combinations of the final state objects. If there is a bump, transverse mass variables are also considered. The SM background is obtained from the VISTA procedure.

The method is described in detail in [3]. Each mass variable is scanned with a sliding window of width equal to twice the typical detector resolution for the component objects. Windows are only considered which contain at least 5 data events. The p-value for the window is defined as the Poisson probability that the expected SM background would fluctuate up to or above the number of data events observed. To ensure the window really represents a bump of the correct resolution-based width and not some broader excess, the “side-bands” of equal width on either side of the central window are required to meet certain minimal criteria regarding consistency with the SM expectation.

In each mass variable, the bump candidate with the smallest p-value is selected. The significance of this bump is given by $P_a$, the fraction of pseudo-experiments which would have produced a more interesting bump in this mass variable purely by random fluctuations of the SM background. $P_a$ incorporates the trials factor associated with examining multiple overlapping windows within the mass variable. For computational reasons, it is prohibitive to determine $P_a$ by pseudo-experiments for all mass variables, so instead an analytic approximation is used. If the analytic estimation returns a value of $P_a$ with a significance of $\geq 4.5\sigma$, then pseudo-experiments are performed for accurate determination.

Each mass variable is further assigned a probability $P_b$, which

<table>
<thead>
<tr>
<th>Final State</th>
<th>Data</th>
<th>Background</th>
<th>$\sigma$</th>
<th>$\sigma_t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b\ell_{b}'$</td>
<td>690</td>
<td>817.7 ± 9.2</td>
<td>-4.3</td>
<td>-2.7</td>
</tr>
<tr>
<td>$\gamma T^\pm$</td>
<td>1371</td>
<td>1217.6 ± 13.3</td>
<td>+4.0</td>
<td>+2.2</td>
</tr>
<tr>
<td>$\mu^\pm T^\pm$</td>
<td>63</td>
<td>35.2 ± 2.8</td>
<td>+3.7</td>
<td>+1.7</td>
</tr>
<tr>
<td>$b2j_{pT}$ ($\Sigma p_T &gt; 400$ GeV)</td>
<td>255</td>
<td>327.2 ± 8.9</td>
<td>-3.7</td>
<td>-1.7</td>
</tr>
<tr>
<td>$2j_T^\pm$ ($\Sigma p_T &lt; 400$ GeV)</td>
<td>574</td>
<td>670.3 ± 8.6</td>
<td>-3.6</td>
<td>-1.5</td>
</tr>
<tr>
<td>$3j_T^\pm$ ($\Sigma p_T &lt; 400$ GeV)</td>
<td>148</td>
<td>199.8 ± 5.2</td>
<td>-3.5</td>
<td>-1.4</td>
</tr>
<tr>
<td>$e^\pm T^\pm$ $p_T$</td>
<td>36</td>
<td>17.2 ± 1.7</td>
<td>+3.5</td>
<td>+1.4</td>
</tr>
<tr>
<td>$2j_T^\pm T^\pm$</td>
<td>33</td>
<td>62.1 ± 4.3</td>
<td>-3.5</td>
<td>-1.3</td>
</tr>
<tr>
<td>$e^\pm j$</td>
<td>741710</td>
<td>764832 ± 6447.2</td>
<td>-3.5</td>
<td>-1.3</td>
</tr>
<tr>
<td>$j2\ell^\pm$</td>
<td>105</td>
<td>150.8 ± 6.3</td>
<td>-3.4</td>
<td>-1.2</td>
</tr>
</tbody>
</table>

TABLE I: The ten most discrepant VISTA final states, showing the number of data events observed and the number of background events expected. Only Monte Carlo statistical uncertainties on the background prediction are included. $\sigma$ and $\sigma_t$ represent the level of discrepancy, before and after accounting for the trials factor.
defined as the probability under the null hypothesis that any mass variable would appear more significant than this. Assuming no correlations, \( P_b = 1 - (1 - P_a)^N \), where \( N \) is the total number of mass variables examined. If \( P_b \) corresponds to a significance of \( \geq 3\sigma \), that effect is then considered as potentially due to new physics.

The Bump Hunter examines 5036 mass variables, of which 2316 are found to have at least one bump satisfying the above criteria (where the difference is mainly due to small-population final states failing to satisfy the criterion of 5 data events in a mass window). The expected and observed distributions of \( P_a \), converted to units of \( \sigma \), are shown in Fig. 4. The distribution of \( P_a \) in the data is seen to be shifted towards positive \( \sigma \) relative to the expectation, indicating disagreement between data and SM prediction. This reflects the fact that the Bump Hunter algorithm is quite sensitive to local features in the mass variables which can arise since the Monte Carlo-based SM background prediction does not perfectly describe the data. The sharp drop seen in the data at \( P_a = 4.5\sigma \) results from the transition between analytic estimation and accurate determination of \( P_a \).

The only mass variable with a bump which exceeds the discovery threshold is the invariant mass of all four jets in the 4j final state, shown in Fig. 5. This mass variable has \( P_a \) corresponding to 5.7\( \sigma \), and \( P_b \) to 4.1\( \sigma \). However, this bump is attributed to the aforementioned difficulty modeling soft QCD jets and is not thought to indicate new physics.

The final component of this global search for physics beyond the standard model is a procedure called SLEUTH [4–6]. SLEUTH is a quasi-model-independent search technique, based on the assumption that new electroweak-scale physics will manifest itself as a high-\( p_T \) excess of data over the SM expectation in a particular final state. Tests have shown SLEUTH to have sensitivity comparable to targeted searches for phenomena that satisfy SLEUTH’s basic assumptions.

The procedure is identical to that used in [1]. The algorithm considers a single variable, the summed scalar transverse momentum (\( \sum p_T \)) of all objects in the event. The SM prediction for the distribution of \( \sum p_T \) is determined as part of the VISTA procedure. The exclusive final states examined by SLEUTH are created by merging VISTA final states according to certain rules described in [1]. For each final state, SLEUTH determines the region (defined as an interval in \( \sum p_T \) extending from a data-point up to infinity) which has the smallest probability that the SM prediction would fluctuate up to or above the number of observed data events. The algorithm then finds \( \mathcal{P} \), the fraction of pseudo-experiments drawn from the SM \( \sum p_T \) distribution which produce any region more interesting than the region found in the data. SLEUTH selects the final state with the smallest value of \( \mathcal{P} \), and calculates the overall significance, \( \mathcal{P} \), which accounts for the number of final states considered. With an accurate correction model and in the absence of new physics, the distribution of \( \mathcal{P} \) is uniform between zero and unity; in the presence of new physics, a small value of \( \mathcal{P} \) is expected. The threshold for pursuit of a possible discovery case is taken to be \( \mathcal{P} < 0.001 \).

The distribution of \( \mathcal{P} \) for the final states considered by SLEUTH in the data is shown in Fig. 6. The concavity of this distribution reflects the degree to which the correction model has been tuned: a crude correction model tends to produce a distribution that curves upwards, as seen in this figure, while an excessively tuned correction model would produce a distribution that curves downwards, with more final states than expected having \( \mathcal{P} \) near the midpoint of the unit interval.

The \( \sum p_T \) distributions of the four most interesting final states found by SLEUTH are shown in Fig. 7. These are: \( e^{\pm}\mu^{\pm} \), \( e^{\pm}\mu^{\pm}jj \), \( e^{\pm}\mu^{\pm}p_T \), and \( e^{\pm}e^{\pm}\mu^{\pm}p_T + \mu^{\pm}\mu^{\mp}e^{\pm}p_T \). It is intriguing to note that all four con-
tain the rare signature of a same-sign electron-muon pair. Such a signature can arise in a number of ways. SM processes that produce real electrons and muons with the same charge include $WZ$ production with leptonic decays, where one of the leptons is not reconstructed in the detector. There are also processes which produce real electrons in the forward region of the CDF II detector, where the reduced tracking coverage means the electron charge sign has a higher probability of being falsely reconstructed; such processes include $tf$ production, and $Z \rightarrow \tau^+\tau^-$ where both taus decay leptonically. In addition, there are processes with a real muon and a fake electron. These are largely $W/Z+\text{jets}$ production, where a primary quark or gluon jet is mis-identified as an electron in the detector, and $W\gamma/Z\gamma$, where the photon undergoes conversion to produce an electron. Also relevant is the case when both the electron and muon are fakes, predominantly from dijet events. The relative proportion of these potential backgrounds varies for each final state, depending on the presence of $p_T$ and the number of jets. Since all of these processes and detector effects also contribute to other more highly-populated final states where good agreement is seen, their rates are quite well constrained by this global analysis.

However, while it is noteworthy that the top four final states all contain the same rare signature, this is an a posteriori observation and its significance is therefore difficult to estimate. SLEUTH’s a priori procedure is to calculate the significance of only the single most discrepant final state. We find that $P = 0.08$, i.e. that 8% of pseudo-experiments drawn from the VISTA SM implementation would have produced a more significant excess in a single final state purely by chance fluctuations. This is far from the threshold of $P < 0.001$, and therefore we do not pursue this as a potential discovery.

In summary, CDF has performed a model-independent global search for new high-$p_T$ physics in 2.0 fb$^{-1}$ of $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV. The populations of 399 exclusive final states are compared to a standard model prediction, but no significant discrepancy is found after accounting for the trials factor associated with looking in many places. The shapes of 19 650 kinematic distributions are also studied, and although 555 show a significant discrepancy, most of these are attributed to inadequate modeling of soft QCD jet emission in the underlying Monte Carlo prediction, rather than a sign of new physics. A Bump Hunter algorithm scans invariant mass distributions for narrow bumps that could indicate resonant production of new particles: only one significant bump is found, and it is attributed to the same underlying problem as above. The SLEUTH algorithm searches the $\sum p_T$ spectrum of each final state, but finds no significant excesses of data over SM prediction in the tails of any single distribution. This CDF global search has not discovered new physics in 2.0 fb$^{-1}$.

We thank the Fermilab staff and the technical staffs of the participating institutions for their vital contributions. This work was supported by the U.S. Department of Energy and National Science Foundation; the Italian Istituto Nazionale di Fisica Nucleare; the Ministry of Education, Culture, Sports, Science and Technology of Japan; the Natural Sciences and Engineering Research Council of Canada; the National Science Council of the Republic of China; the Swiss National Science Foundation; the A.P. Sloan Foundation; the Bundesministerium für Bildung und Forschung, Germany; the Korean Science and Engineering Foundation and the Korean Research Foundation; the Science and Technology Facilities Council and the Royal Society, UK; the Institut National de Physique Nucléaire et Physique des Particules/CNRS; the Russian Foundation for Basic Research; the Ministerio de Educación y Ciencia and Programa Consolider-Ingenio 2010, Spain; the Slovak R&D Agency; and the Academy of Finland.

[7] We use a cylindrical coordinate system where the $z$-axis is in the direction of the proton beam, and $\theta$ and $\phi$ are respectively the polar and azimuthal angles. The pseudorapidity is defined as $\eta = -\ln(\tan(\theta/2))$. The transverse
FIG. 7: The $\sum p_T$ distributions of the four most interesting final states found by Sleuth. The label in the top left corner of each plot lists the objects in the final state, where $l^\pm$ is a lepton ($e$ or $\mu$), $l'$ is an additional lepton of different flavor, $j$ denotes a jet, and $p_T$ represents missing transverse momentum. Global charge conjugation is implied, so that a final state labeled $l^+ l'^+$ also includes $l^- l'^-$. The region with the most significant excess of data over SM expectation is indicated by the arrow below the $x$-axis, and displayed in the inset with the number of events expected (SM) and observed (d). The significance of the excess is shown by the value of $P$ in the top right corner.

momentum $p_T$ is defined as $p_T = p \sin \theta$. Missing transverse momentum $p_T$ is defined as the magnitude of the transverse component of the negative vector sum of the 4-vectors of all identified objects and unclustered momentum in an event, where unclustered momentum is visible in the detector but not clustered into an identified object.