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

T. Aaltonen, J. Adelman, T. Akimoto, M.G. Albrow, B. Álvarez González, S. Ameiro, A. Anidestavc, A. Annovi, J. Antos, G. Apollinari, A. Arpesyan, T. Arisawa, A. Artikov, W. Ashmanskas, A. Aurisano, F. Azzar, F. Azzurri, W. Badgett, V.E. Barnes, B.A. Barnett, V. Bartsch, G. Bauer, H.P.-A. Bechtemin, F. Bedeschi, D. Beecher, S. Behari, G. Belletti, J. Bellinger, D. Benjamin, A. Berätt, A. Bhattacharjee, M. Binkley, D. Bisello, I. Bizjak, D.E. Blair, C. Block, B. Blumenfeld, A. Bocci, V. Boisvert, G. Bolla, D. Bortoletto, J. Boudreau, B. Boveia, B. Braun, A. Bridgeman, L. Brigliadori, C. Bromberg, E. Brukaker, J. Budagov, H.S. Budd, S. Budd, S. Burke, K. Burkett, G. Busetto, P. Bussey, A. Buzatu, K.L. Byrum, S. Cabrera, C. Calancha, M. Campanelli, M. Campbell, F. Canelli, A. Canepa, B. Carls, D. Carlsmith, R. Carosi, S. Carroll, S. Carron, M. Casarsa, A. Castro, P. Catastini, D. Cauz, V. Cavaliere, A. Cerri, L. Cerretti, S.H. Chang, Y.C. Chen, M. Chertok, G. Chiarelli, G. Chlachidze, F. Chlebana, K. Cho, D. Chokheli, J.P. Chou, G. Choudalakis, S.H. Chuang, K. Chung, W.S. Chung, T.H. Chung, Y.S. Chung, T. Chwalek, C.I. Ciobanu, M.A. Ciocci, A. Clark, D. Clark, G. Compostella, M.E. Convery, J. Conway, M. Cordelli, G. Cortiana, C.A. Cox, D.J. Cox, F. Crepaldi, C. Cuenca Almenar, J. Cuevas, R. Cubertson, J.C. Cully, D. Dagenhart, M. Datta, T. Davies, P. de Barbaro, S. De Cecco, A. Deisseroth, M. Dell'Orso, L. Demortier, J. Deng, M. Denino, P.F. Derwent, G.P. Di Giovanni, C. Dionisi, B. Di Ruza, J.R. Dittmann, S. Donati, P. Dong, J. Donini, T. Dorigo, S. Dube, J.F. Eaton, A. Elagin, R. Erbacher, E. Errede, R. Eusebi, H.C. Fang, S. Farrington, W.T. Fedorko, R.G. Feld, M. Feindt, J.P. Fernandez, C. Ferrazza, R. Field, G. Flanagan, R. Forrest, J.M. Frank, M. Franklin, J.C. Freeman, I. Furie, M. Gallinaro, J. Galyardt, F. Garberson, J.E. Garcia, A.F. Garfinkehl, K. Genser, H. Gerberich, D. Gerdes, A. Gessler, S. Giagni, V. Giakoumopoulou, P. Giannetti, K. Gibson, J.L. Gimmell, C.M. Ginsburg, N. Giokaris, M. Giordani, M. Giunta, G. Giorgi, V. Glagolev, D. Glenzinski, M. Gold, N. Goldsmith, A. Golossanov, G. Gomez-Ceballos, M. Goncharov, O. Gonzalez, I. Gorelov, A.T. Goshaw, K. Goulianos, A. Gresele, S. Grinstein, C. Gross-Pilcher, R.C. Group, U. Grundler, J. Guimaraes da Costa, Z. Gunay-Unalan, K. Hahn, S.R. Hahn, E. Halkiadakis, B.Y. Han, J.Y. Han, F. Happacher, K. Hara, D. Hare, M. Hare, S. Harper, R.F. Harr, R.M. Harris, M. Hartz, K. Hatakeyama, C. Hays, M. Heck, A. Heijboer, J. Heinrich, C. Henderson, C. Herndon, J. Heusser, S. Hewamanage, D. Hidas, C.S. Hill, D. Hirschbuehl, A. Hocker, S. Hou, M. Houden, S.-C. Hsu, B.T. Huffman, R.E. Hughes, U. Husemann, J. Huston, J. Incandela, G. Introzzi, M. Iorizzo, A. Ivanov, E. James, B. Jayatilaka, E.J. Jeon, M.K. Jha, L. Jindariani, W. Johnson, M. Jones, K.K. Joo, S.Y. Jun, J.E. Jungr, T.R. Junk, B. Kamon, D. Kar, P.E. Karchin, Y. Kato, R. Kephart, J. Keung, V. Khotilovich, B. Kilminster, D.H. Kim, H.S. Kim, H.W. Kim, J.E. Kim, M.J. Kim, S.B. Kim, S.H. Kim, Y.K. Kim, N. Kinuma, L. Kirsch, S. Klimenko, B. Knutsen, B.R. Ko, K. Kondo, D.J. Kong, J. Konigsberg, A. Korytov, A.V. Kotwal, M. Kreps, J. Kroll, D. Krop, N. Krummack, M. Kruse, V. Krutelyov, T. Kubo, T. Kulkarni, M. Kurata, Y. Kusakabe, S.W. Lee, J.E. Lee, B.-Y. Han, P. McIntyre, T. Maruyama, T. Maki, H. Miyake, N. Moggi, S.H. Oh, S. Pagano Griso, E. Palencia, V. Papadimitriou, A. Papaikonomou, A.A. Paramonov, B. Parks, S. Pashapour, J. Patrick, G. Paulietta, M. Paulini, C. Paus, T. Peiffer

10 Jan 2009
D.E. Pellett, A. Penzo, T.J. Phillips, E. Piacentino, E. Pianini, L. Pinera, K. Pitts, C. Plager, L. Pondrom, O. Poukhov, N. Pounder, F. Prakoshlyan, A. Pranko, J. Proudfoot, F. Pthohs, E. Pueschel, G. Punzi, J. Pursey, J. Rademacker, A. Rahaman, V. Ramakrishnan, N. Ranjan, I. Redondo, V. Rekovic, P. Renton, M. Renz, M. Rescigno, S. Richter, F. Rimondi, L. Ristori, A. Robson, T. Rodrigo, T. Rodriguez, E. Rogers, R. Roser, M. Rossi, R. Roy, A. Ruiz, V. Rusu, A. Safonov, W.K. Sakumoto, L. Santi, S. Sarkar, L. Sartori, K. Sato, A. Savoy-Navarro, P. Schlabach, A. Schmidt, E.E. Schmidt, M.A. Schmidt, M.P. Schmidt, M. Schmitt, T. Schwarz, L. Scodellaro, A. Scrici, F. Scru, A. Sedov, S. Seidel, Y. Seiya, E. Pueschel, I. Suslov, E. Pueschel, I. Redondo, P. Totaro, M. Vidal, D.E. Pellett, S. Tkaczyk, M. Schmitt, A. Ruiz, S. Wilbur, 5

Institut de Fisica d’Altes Energies, Universitat Autonoma de Barcelona, E-08193, Bellaterra (Barcelona), Spain

Comenius University, 842 48 Bratislava, Slovakia; Institute of Experimental Physics, 040 01 Kosice, Slovakia

A. Savoy-Navarro, T. Suzuki, U.K. Yang

20 Laboratori Nazionali di Frascati, Istituto Nazionale di Fisica Nucleare, I-00044 Frascati, Italy

University of Geneva, CH-1211 Geneva 4, Switzerland

University of Geneva, CH-1211 Geneva 4, Switzerland

University of Geneva, CH-1211 Geneva 4, Switzerland

University of Geneva, CH-1211 Geneva 4, Switzerland

University of Geneva, CH-1211 Geneva 4, Switzerland

University of Geneva, CH-1211 Geneva 4, Switzerland

University of Geneva, CH-1211 Geneva 4, Switzerland

University of Geneva, CH-1211 Geneva 4, Switzerland

University of Geneva, CH-1211 Geneva 4, Switzerland
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 Sleuth procedure scans for data excesses at large summed transverse momentum. This combined global search for new physics in 2.0 fb$^{-1}$ of $p\bar{p}$ 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 \(\text{[1]}\).

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 \(\text{[2]}\) 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 \(\text{[1]}\). 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 (\(\not{p}_T\)) \(\text{[7]}\). 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 ratios (\(k\)-factors) of the actual cross sections for SM processes to the leading order approximations 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 \(\text{[1]}\); these changes are described in detail in \(\text{[3]}\).

The first stage in the VISTA global comparison is to study the populations of the exclusive final states, compared to the SM expectation. Figure \(\text{[1]}\) 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 \(\text{[1]}\). After accounting for the trials factor associated with considering many final
mass variables that can be constructed from combinations of standard deviations.

The Bump Hunter algorithm is designed to identify mass resonances of resonant production of a new particle. The bump variable would be the most direct evidence of these shape discrepancies. The nature of these shape discrepancies is treated as statistically significant population discrepancies.

We find that 555 distributions have a significant discrepancy, which is defined as being greater than 3σ after accounting for the trials factor associated with the number of distributions considered. The discrepant distributions fall into three categories. Residual “crude” in the correction model, primarily from using simplified \( p_T \) dependences for fake rate correction functions, accounts for 3%. Another 16% are attributed to an inadequate modeling of the transverse boost of the colliding system. The remaining 81% most likely arise from incorrect modeling of soft QCD parton showering. This is best exemplified by Fig. 2, which shows \( \Delta R = \sqrt{\Delta \phi^2 + \Delta \eta^2} \) between the second and third highest \( p_T \) jets in the \textit{VISTA} 3-jet final state. This observation has been discussed in more detail in [3]. 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 a 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 \( p_T \), transverse mass variables are also considered. The SM background is obtained from the \textit{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. Only windows that contain at least 5 data events are considered. The p-value for the window is defined as the Poisson probability that the expected SM events are considered. The p-value for the window is defined as the Poisson probability that the expected SM events are considered.

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 pseudoexperiments 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 pseudoexperiments 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 pseudoexperiments are performed for accurate determination.

Each mass variable is further assigned a probability \( P_b \), 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
The Bump Hunter examines 5036 mass variables, of which 2316 are found to have at least one local excess satisfying our bump definition (the other variables are mainly from small-population final states which fail 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 the SM prediction. This reflects the fact that the Bump Hunter algorithm is quite sensitive to local features in 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, 5, 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 $\tilde{P}$, the fraction of pseudoexperiments 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 $\tilde{P}$, and calculates the overall significance, $\tilde{P}$, which accounts for the number of final states considered. With an accurate correction model and in the absence of new physics, the distribution of $\tilde{P}$ is uniform between zero and unity; in the presence of new physics, a small value of $\tilde{P}$ is expected. The threshold for pursuit of a possible discovery case is taken to be $\tilde{P} < 0.001$. 

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{Fig4}
\caption{Significance of the most interesting bump in each mass variable ($P_a$, in units of standard deviations) considered by the Bump Hunter. The black line represents the theoretical expectation assuming no new physics.}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{Fig5}
\caption{Distribution of the invariant mass of all four jets in the $4j \sum p_T < 400$ GeV final state. This variable contains the most significant bump found by the Bump Hunter, indicated by the dashed (blue) lines.}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{Fig6}
\caption{The distribution of $\tilde{P}$ in the data, with one entry for each final state considered by SLEUTH. The black line represents the theoretical expectation assuming no new physics.}
\end{figure}
The distribution of $P$ for the final states considered by SLEUTH in the data is shown in Fig. 4. The concavity of this distribution reflects the crudeness (i.e. under-tuning) of our correction model. A crude correction model results in more outliers than expected, which, when converted into values of $P$, produces excesses at the extremes of both low and high probability.

The $\sum p_T$ distributions of the four most interesting final states found by SLEUTH are shown in Fig. 7. These are: $e^\pm\mu^\pm$, $e^\pm\mu^\pm jj$, $e^\pm\mu^\pm t\bar{t}$, and $e^\pm e^\pm\mu^\pm p_T + \mu^\pm e^\pm p_T$. It is intriguing to note that all four contain 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 $t\bar{t}$ 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 $W/Z+$jets production, where a primary quark or gluon jet is misidentified 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 the 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 $pp$ 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 the 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.

[1] T. Aaltonen et al. (CDF Collaboration), Phys. Rev. D 78, 012002 (2008).
[2] A. Abulencia et al. (CDF Collaboration), J. Phys. G 34, 2457 (2007).
[3] G. Choudalakis, Ph.D. thesis, Massachusetts Institute of Technology (2008), arXiv:0805.3954.
[4] B. Abbott et al. (D0 Collaboration), Phys. Rev. Lett. 86, 3712 (2001).
[5] B. Abbott et al. (D0 Collaboration), Phys. Rev. D 62, 092004 (2000).
[6] V. M. Abazov et al. (D0 Collaboration), Phys. Rev. D 64, 012004 (2001).
[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 pseudo-rapidity is defined as $\eta = -\ln(\tan(\theta/2))$. The transverse momentum $p_T$ is defined as $p_T = p\sin\theta$. The missing transverse momentum $p_T$ is defined as the magnitude of the transverse component of the negative vector sum of the four-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.
FIG. 7: The $\sum p_T$ distributions of the four most interesting final states found by SLEUTH. Data are shown as filled (black) circles, with the expected contribution from standard model processes shown as the shaded (red) histograms and identified in the legend. The category “Other” represents the sum of all remaining relevant SM processes, each of which individually is smaller than the smallest itemized contribution. The label in the top left corner of each plot lists the objects in the final state, where $l$ is a lepton ($\ell$ 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 the 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.