E. Nurse, S.H. Oh, Y.D. Oh, I. Oksuzian, T. Okusawa, R. Orava, K. Osterberg, S. Pagan Griso, C. Pagliaire, E. Palencia, V. Papadimitriou, A. Papaikonomou, A.A. Paramonov, B. Parks, S. Pashapour, J. Patric, G. Pauletta, M. Paulini, C. Paul, D.E. Pellett, A. Penzo, T.J. Phillips, G. Piacentino, J. Piedra, L. Pineri, K. Pitts, C. Plager, L. Pondrom, O. Pountou, F. Prakoslysh, A. Pronko, J. Proudfoot, F. Ptosh, G. Punzi, J. Pursley, J. Rademacker, A. Rahaman, V. Ramakrishnan, N. Ranjan, I. Redondo, B. Reisert, V. Rekovic, P. Renton, M. Rescigno, S. Richter, F. Rimondi, A. Robson, T. Rodrigo, E. Rogers, R. Roser, M. Rossi, R. Rossin, P. Roy, A. Ruiz, J. Russ, V. Rusu, H. Saarikko, A. Safonov, W.K. Sakumoto, G. Salamanna, L. Santi, S. Sarkar, L. Sartori, K. Sato, A. Savoy-Navarro, T. Scheide, P. Schlachber, E.E. Schmidt, M.P. Schmidt, M. Schmitt, T. Schwarz, L. Scodellaro, A.L. Scott, A. Seribano, F. Scuri, A. Sedov, S. Seidel, Y. Seiya, A. Semenov, L. Sexton-Kennedy, A. Sfyrla, S.Z. Shalhout, T. Shears, P.F. Shepard, D. Shermman, M. Shimofjima, M. Shochet, Y. Shon, L. Shreyber, A. Sidoti, A. Sisakyan, J.A. Slaughter, J. Slauhwhite, K. Sliwa, J.R. Smith, F.D. Snider, R. Sunhur, M. Soderberg, A. Soha, S. Somalwar, V. Sorin, J. Spalding, F. Spinella, T. Spiretz, P. Squillacioti, M. Stanitzki, R. St. Denis, B. Stelzer, O. Stelzer-Chilton, D. Stentz, J. Strologas, J.S. Suh, A. Sukhanov, H. Sun, I. Suslov, T. Suzuki, A. Taffard, R. Takashima, Y. Takeuchi, R. Tanaka, M. Tecchio, P.K. Teng, T. Terashi, J. Thom, A.S. Thompson, G.A. Thompson, E. Thomson, P. Tipton, V. Tiwari, S. Tkaczuk, D. Toback, S. Tokar, K. Tollefsen, T. Tomura, D. Tonelli, S. Torre, D. Torretta, S. Tourneur, Y. Tu, N. Turini, F. Ukegawa, S. Uozumi, V. Vallecorsa, N. van Remortel, A. Varganov, E. Vataga, F. Vázquez, G. Velev, C. Vellidis, V. Vespremi, M. Vidal, R. Vidal, T. Vila, R. Vilari, T. Vine, M. Vogel, F. Würthwein, P. Wagner, R.G. Wagner, R.L. Wagner, J. Wagner, W. Wagner, R. Wallny, S.M. Wang, A. Warburton, D. Waters, M. Weinberger, W.C. Wester III, B. Whitehouse, D. White, A. Wicklund, E. Wicklund, G. Williams, H.H. Williams, B.L. Winer, P. Witten, C. Woblers, T. Wright, X. Wu, S.M. Wymne, S. Xie, A. Yagil, K. Yamamoto, J. Yamaoka, T. Yamashita, E. Yama, C. Yang, U.K. Yang, Y.C. Yang, W.M. Yao, G.P. Yeh, J. Yoh, K. Yorita, T. Yoshiida, G.B. Yu, J. Yu, S.S. Yu, J.C. Yun, L. Zanello, A. Zanetti, I. Zaw, X. Zhang, Y. Zheng, and S. Zucchelli

(CDF Collaboration)

1Institute of Physics, Academia Sinica, Taipei, Taiwan 11529, Republic of China
2Argonne National Laboratory, Argonne, Illinois 60439
3Institut de Física d’Altes Energies, Universitat Autonoma de Barcelona, E-08193, Bellaterra (Barcelona), Spain
4Baylor University, Waco, Texas 76798
5Instituto Nazionale di Fisica Nucleare, University of Bologna, I-40127 Bologna, Italy
6Brandeis University, Waltham, Massachusetts 02254
7University of California, Davis, California 95616
8University of California, Los Angeles, Los Angeles, California 90024
9University of California, San Diego, La Jolla, California 92093
10University of California, Santa Barbara, Santa Barbara, California 93106
11Instituto de Fisica de Cantabria, CSIC-University of Cantabria, 39005 Santander, Spain
12Carnegie Mellon University, Pittsburgh, PA 15213
13Enrico Fermi Institute, University of Chicago, Chicago, Illinois 60637
14Comenius University, 842 48 Bratislava, Slovakia; Institute of Experimental Physics, 040 01 Kosice, Slovakia
15Joint Institute for Nuclear Research, RU-141980 Dubna, Russia
16Duke University, Durham, North Carolina 27708
17Fermi National Accelerator Laboratory, Batavia, Illinois 60510
18University of Florida, Gainesville, Florida 32611
19Laboratori Nazionali di Frascati, Istituto Nazionale di Fisica Nucleare, I-00044 Frascati, Italy
20University of Geneva, CH-1211 Geneva 4, Switzerland
21Glasgow University, Glasgow G12 8QQ, United Kingdom
22Harvard University, Cambridge, Massachusetts 02138
23Division of High Energy Physics, Department of Physics, University of Helsinki and Helsinki Institute of Physics, FIN-00014, Helsinki, Finland
24University of Illinois, Urbana, Illinois 61801
25The Johns Hopkins University, Baltimore, Maryland 21218
26Institut für Experimentelle Kernphysik, Universität Karlsruhe, 76128 Karlsruhe, Germany
27Center for High Energy Physics, Kyungpook National University, Taegu 702-701, Korea; Seoul National University, Seoul 151-742,
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. A model-independent approach (VISTA) considers the gross features of the data, and is sensitive to new large cross section physics. A quasi-model-independent approach (SLEUTH) searches for a significant excess of events with large summed transverse momentum, and is particularly sensitive to new electroweak scale physics that appears predominantly in one final state. This global search for new physics in over three hundred exclusive final states in 927 pb$^{-1}$ of $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV reveals no such significant indication of physics beyond the standard model.

*With visitors from a University of Athens, 15784 Athens, Greece, bChinese Academy of Sciences, Beijing 100864, China, cUniversity of Bristol, Bristol BS8 1TL, United Kingdom, dUniversity Libre de Bruxelles, B-1050 Brussels, Belgium, eUniversity of California, Irvine, Irvine, CA 92697, fUniversity of California Santa Cruz, Santa Cruz, CA 95064, gCornell University, Ithaca, NY 14853, hUniversity of Cyprus, Nicosia CY-1678, Cyprus, iUniversity College Dublin, Dublin 4, Ireland, jUniversity of Edinburgh, Edinburgh EH9 3JZ, United Kingdom, kUniversity of Heidelberg, D-69120 Heidelberg, Germany, lUniversidad Iberoamericana, Mexico D.F., Mexico, mUniversity of Manchester, Manchester M13 9PL, nNagasaki Institute of Applied Science, Nagasaki, Japan, oUniversity de Oviedo, E-33007 Oviedo, Spain, pQueen Mary’s College, University of London, London, E1 4NS, England, qTexas Tech University, Lubbock, TX 79409, rIFIC(CSIC-Universitat de Valencia), 46071 Valencia, Spain.
The particle physics standard model (SM) is remarkably successful, but is believed to require expansion beyond the electroweak scale. A variety of possible extensions have been proposed. Many analyses optimized for specific signatures have been performed to search for evidence of these possibilities. Limits have been set on cross sections for postulated processes and on masses of hypothetical particles, but no conclusive indication of physics beyond the standard model has yet been seen [1].

This Letter summarizes a broad search for new physics at the electroweak scale without focusing on any specific proposed scenario. The detailed writeup is provided in Ref. [2]. Events containing one or more particles produced at large transverse momentum collected by the CDF experiment in Run II of the Fermilab Tevatron are analyzed for discrepancies relative to the standard model prediction. A model-independent approach (VISTA) considers gross features of the data, and is sensitive to new large cross section physics. These global algorithms provide a complementary approach to searches optimized for more specific new physics scenarios. Searches in a similar spirit have previously been performed by the D0 Collaboration [3, 4, 5] in Tevatron Run I and by the H1 Collaboration [6] at HERA-I.

This search for new physics is designed with the intention of maximizing the chance for discovery, rather than excluding model parameter space if no discrepancy is found. Discrepancies between data and a complete standard model background estimate are identified in a global sample of high transverse momentum (high-\(p_T\)) collision events. Three statistics are employed to identify and quantify disagreement: populations of exclusive final states defined by the objects the events contain, shapes of kinematic distributions, and excesses on the tail of summed scalar transverse momentum distributions. These statistics identify discrepancies worthy of further study.

A discovery claim can be made to the extent that a highlighted discrepancy can be demonstrated to be not due to a statistical fluctuation, a mismodeling of the detector response, or an inadequate implementation of the standard model prediction, and must therefore be due to some new underlying physics. Any observed discrepancy is subject to scrutiny, and explanations are sought in terms of the above points.

The VISTA and SLEUTH algorithms provide a means for making the above three arguments, with a high threshold placed on the statistical significance of a discrepancy in order to minimize the chance of a false discovery claim. As described later, this threshold is the requirement that the false discovery rate is less than 0.001, after taking into account the total number of final states, distributions, or regions being examined.

The traditional notions of signal and control regions are modified. Removing prejudice as to where new physics may appear, all regions of the data are treated as both signal and control. This analysis is not blind, but rather seeks to identify and understand discrepancies between data and the standard model prediction. With the goal of discovery, emphasis is placed on examining discrepancies, focusing on outliers rather than global goodness of fit. Individual discrepancies that are not statistically significant are generally not pursued.

VISTA and SLEUTH are employed simultaneously, rather than sequentially. An effect highlighted by SLEUTH prompts additional investigation of the discrepancy, usually resulting in a specific hypothesis explaining the discrepancy in terms of a detector effect or adjustment to the standard model prediction that is then fed back and tested for global consistency using VISTA.

Forming hypotheses for the cause of specific discrepancies, implementing those hypotheses to assess their wider consequences, and testing global agreement after the implementation are emphasized as the crucial activities for the investigator throughout the process of data analysis [1]. This process is constrained by the requirement that all adjustments be physically motivated.

This search for new physics terminates when one of two conditions are satisfied: either a compelling case for new physics is made, or there remain no statistically significant discrepancies on which a new physics case can be made. In the former case, to quantitatively assess the significance of the potential discovery, a full treatment of systematic uncertainties must be implemented. In the latter case, it is sufficient to demonstrate that all observed effects are not in significant disagreement with an appropriate global standard model description.

This analysis uses data corresponding to an integrated luminosity of 927 \(pb^{-1}\) of \(pp\) collisions at \(\sqrt{s} = 1.96\) TeV recorded by the CDF II detector [7]. CDF II consists of a charged particle tracking system composed of silicon strip detectors and a gas drift chamber inside a 1.4 T magnetic field, surrounded by electromagnetic and hadronic calorimeters and enclosed by muon detectors.

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 momentum (\(\not p\)). Monte Carlo event generators are used to determine the standard model prediction. VISTA partitions data and Monte Carlo events into exclusive final states labeled according to the objects (\(e^\pm, \mu^\pm, \tau^\pm, \gamma, j, b, \not p\)) identified in each event. Each event belongs to one and only one exclusive final state [12].

A correction model is developed to improve systematic deficiencies in the standard model theoretical prediction and the simulation of the detector response. Achieving this on the entire high-\(p_T\) dataset requires a framework...
TABLE I: A subset of the VISTA comparison between Tevatron Run II data and the standard model prediction, showing the final states with greatest discrepancies in population. Final states are labeled in this table according to the number and types of objects present, and are ordered according to decreasing discrepancy between the total number of events expected and the total number observed in the data. Only statistical uncertainties on the standard model prediction are shown; systematics are incorporated by allowing their values to float in the overall fit. A total of 344 populated exclusive final states are considered.

| Final State Data | SM prediction | Final State Data | SM prediction |
|------------------|---------------|------------------|---------------|
| $e^+\gamma$      | 1661          | $e^+\gamma$      | 28656         |
| $\tau^+$         | 71            | $\tau^+$         | 636           |
| $\mu$            | 233           | $\mu$            | 131           |
| $j_{\gamma}$     | 6             | $j_{\gamma}$     | 50            |
| $b\tau$          | 2207          | $b\tau$          | 74            |
| $b\gamma$        | 35436         | $b\gamma$        | 10            |
| $e^+\gamma\gamma$| 1954          | $e^+\gamma\gamma$| 286           |
| $\mu\gamma$      | 798           | $\mu\gamma$      | 20            |
| $\mu\tau$        | 811           | $\mu\tau$        | 96502         |
| $e^+\mu$         | 26            | $e^+\mu$         | 356           |

for quickly implementing and testing modifications to the correction model, including a quick fit for values of associated correction factors. The specific details of the correction model are intentionally kept as simple as possible in the interest of transparency in the event of a possible new physics claim. The details of this correction model are motivated by individual discrepancies noted in a global comparison of CDF high-$p_T$ data to the standard model prediction. The correction model includes specific correction factors for the integrated luminosity of the sample, the ratio ($k$-factor) of the actual cross section for a standard model process and the usually leading order approximation given by event generators, object identification efficiencies, object misidentification rates, and trigger efficiencies. A total of 44 correction factors are used, of which over twenty are constrained by external information. A global $\chi^2$ is formed by comparison of CDF data to the standard model prediction, and minimized as a function of these correction factors. Correlations to object identification efficiencies are typically less than 10%; fake rates are consistent with an understanding of the underlying physical mechanisms responsible; $k$-factors range from slightly less than unity to greater than two for some with multiple jets.

A global comparison of data to standard model prediction is made in 16,486 kinematic distributions in 344 populated exclusive final states. In each final state, the number of events observed is compared with the standard model prediction, as shown in Table I and the Poisson probability that the number of predicted events would fluctuate up to or above (or down to or below) the observed number of events is calculated and converted into units of standard deviation. In each kinematic distribution, the shape of the data is compared to the shape of the standard model prediction using the Kolmogorov-Smirnov (KS) statistic, which is converted to a probability and then into units of standard deviation.

VISTA highlights final states and kinematic distributions where the statistical significance of any discrepancy corresponds to a probability < 0.001 after accounting for the appropriate number of final states or distributions considered. The algorithm itself cannot determine whether a particular discrepancy constitutes a discovery of new physics. Physics judgement is required to deter-
mine whether the discrepancy can be explained as a deficiency in the modeling of the CDF II detector response or in the calculation of the standard model prediction.

A summary of the VISTA comparison is shown in Fig. 1.

The numbers of events observed are in agreement with the standard model prediction. The narrow core of the histogram of VISTA final states (top of Fig. 1) is due to final states with few data events. The excess at large σ in the histogram of VISTA distributions (bottom of Fig. 1) shows disagreement between data and standard model prediction in some distributions. The number of distributions showing a significant (> 3σ after the trials factor) difference in shape between data and the standard model prediction is 384. Of these, 312 are attributed to modeling of the parton radiation (with 186 of these 312 pointing out that individual jet masses are larger in data than in the prediction), and 59 reflect an inadequate modeling of the overall transverse boost of the system (“intrinsic kT”). The nature of these discrepant distributions makes it difficult to use them to support a new physics claim, since at present these discrepancies appear most probably due to an imperfect implementation of the standard model prediction. Further investigation into obtaining an adequate QCD-based description is continuing. The remaining 13 discrepant distributions arise from the coarseness of the correction model. Additional details are provided in Ref. [2].

SLEUTH is simultaneously used to search for evidence of new physics on the high-\(p_T\) tails. 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 standard model expectation in a particular final state. The strengths and limitations of SLEUTH follow directly from these assumptions.

SLEUTH considers a single variable, the summed scalar transverse momentum (\(\sum p_T\)) of all objects in the event. The standard model prediction for the distribution of \(\sum p_T\) is determined using the correction factors found by VISTA. For each final state, SLEUTH determines the most interesting region on the tail of this distribution. A final state contains as many regions as data points, where the \(d^{th}\) region is defined as the semi-infinite interval with lower bound equal to the \(d^{th}\) largest data \(\sum p_T\). The \(d^{th}\) region contains \(d\) data events; the number of events expected from the standard model is obtained by integrating the predicted standard model \(\sum p_T\) distribution over this semi-infinite region.

For a region containing \(d\) data points, \(p_d\) is defined as the Poisson probability that the standard model prediction would fluctuate up to or above \(d\). The most interesting region \(R\) is defined as the region for which \(p_d\) is smallest. Pseudo experiments are performed by drawing pseudo data from the standard model \(\sum p_T\) distribution, and the most interesting region is found for each pseudo experiment. The fraction \(P\) of these pseudo experiments producing a region more interesting than the region \(R\) found in the data quantifies the interest of this final state.

Considering all final states, SLEUTH determines the most interesting final state in the CDF high-\(p_T\) data, and calculates \(P\), the fraction of hypothetical similar experiments that would have produced a region more interesting than the region chosen in this final state, and finds \(P_{Wb\bar{b}jj} < 8.3 \times 10^{-7}\), corresponding to a value of \(P\) that easily satisfies SLEUTH’s discovery threshold of \(P < 0.001\).

Figure 2 shows a sensitivity test in which the standard model process \(p\bar{p} \rightarrow t\bar{t}\) is subtracted from the standard model background and observed as an excess in the CDF data. SLEUTH observes the top quark with an integrated Run II luminosity comparable to that accumulated by CDF and D0 in Tevatron Run I when the top quark discovery was announced [9, 11]. Several other sensitivity
tests have been conducted with pseudo signal events injected into pseudo data drawn from the standard model prediction. On these sensitivity tests, SLEUTH performs comparably to targeted searches for phenomena satisfying SLEUTH’s basic assumptions that new physics will appear as an excess of data over the standard model prediction at large summed scalar transverse momentum in one primary final state.

In 927 pb$^{-1}$ of CDF Run II data, SLEUTH finds $\tilde{P} = 0.46$. Assuming any deficiencies in the standard model implementation and detector simulation are accurately resolved by the correction model, the fraction of hypothetical similar CDF experiments that would observe something as interesting as the most interesting region observed in the CDF Run II data is 46%. None of the regions examined surpass SLEUTH’s discovery threshold. Further discussion of the most interesting regions is provided in Ref. [2].

In conclusion, a broad search for new physics (VISTA) has been performed in 927 pb$^{-1}$ of CDF Run II data. A complete standard model background estimate has been obtained and compared with data in 344 populated exclusive final states and 16,486 relevant kinematic distributions. Consideration of exclusive final state populations yields no statistically significant ($> 3\sigma$) discrepancy after the trials factor is accounted for. Quantifying the difference in shape of kinematic distributions using the Kolmogorov-Smirnov statistic, significant discrepancies are observed between data and standard model predictions. These discrepancies are believed to arise from mismeasurement of the parton shower and intrinsic $k_T$, and represent observables for which a QCD-based understanding is highly motivated. None of the shape discrepancies highlighted motivates a new physics claim.

A further systematic search (SLEUTH) for regions of excess on the high-$\sum p_T$ tails of exclusive final states has been performed, representing a quasi-model-independent search for new electroweak scale physics. A measure of interest rigorously accounting for the trials factor associated with looking in many regions is defined, and used to quantify the most interesting region observed in the CDF Run II data. No region of excess on the high-$\sum p_T$ tail of any of the SLEUTH exclusive final states surpasses the discovery threshold.

Although this global analysis of course cannot prove that no new physics is hiding in these data, this broad search of the Tevatron Run II data represents one of the single most encompassing tests of the particle physics standard model at the energy frontier.

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 Particle Physics and Astronomy Research Council and the Royal Society, UK; the Institut National de Physique Nucleaire et Physique des Particules/CNRS; the Russian Foundation for Basic Research; the Comisión Interministerial de Ciencia y Tecnología, Spain; the European Community’s Human Potential Programme; the Slovak R&D Agency; and the Academy of Finland.

[1] Particle Data Group, J. Phys. G33, 119 (2006).
[2] CDF Collaboration (2007), Submitted to Phys. Rev. D. arXiv:0712.1311
[3] D0 Collaboration, Phys. Rev. Lett. 86, 3712 (2001).
[4] D0 Collaboration, Phys. Rev. D62, 092004 (2000).
[5] D0 Collaboration, Phys. Rev. D64, 012004 (2001).
[6] H1 Collaboration, Phys. Lett. B602, 14 (2004).
[7] CDF Collaboration, J. Phys. G34, 2457 (2007).
[8] B. Knuteson, Nucl. Instrum. Meth. A534, 7 (2004).
[9] CDF Collaboration, Phys. Rev. Lett. 74, 2626 (1995).
[10] D0 Collaboration, Phys. Rev. Lett. 74, 2632 (1995).
[11] It is not possible to systematically simulate the process of constructing, implementing, and testing hypotheses motivated by particular discrepancies, since this process is carried out by individuals. The statistical interpretation of this analysis is made bearing this process in mind.
[12] Events that are equivalent under global charge conjugation (such as a $e^+ \bar{p}$ event and a $e^- p$ event) are placed into a single final state (labeled in this case by $e^+ \bar{p}$).
[13] A probability of 0.001 after proper incorporation of the trials factor corresponds roughly to the usual criterion of $5\sigma$ if the trials factor is not accounted for.