CDF note 7114

Search for Anomalous Kinematics in $t\bar{t}$ Dilepton Events at CDF

The CDF Collaboration
URL http://www-cdf.fnal.gov
(Dated: July 13, 2004)

We report on a search for anomalous kinematics of $t\bar{t}$ dilepton events in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV using 193 pb^{-1} of data collected by the CDF II detector. We developed and employed a new *a priori* technique designed to isolate the subset in a data sample revealing the largest dicrepancy from the Standard Model expectations and to quantify the significance of this departure. In the four-variable phase space considered, we find that the dilepton sample has a 1.0–4.5% probability of consistency with the Standard Model.

Preliminary Results for Summer 2004 Conferences

I. INTRODUCTION

The discovery of the top quark during Run I of Fermilab's TeVatron collider [1] began an experimental program to characterize its production and decay properties in all possible decay channels. Within the SM the top quark decays almost exclusively to a W boson and a bottom quark; the "dilepton" decay channel denotes the cases where the two W bosons both decay into a final state containing an electron or muon and accounts for about 7% of all SM top decays. These events are characterized by two energetic leptons, two jets from the hadronization of the bottom quarks, and large missing transverse energy ($\not{E}_T$) from the unobserved neutrinos. The CDF and D0 collaborations' measurements of the $t\bar{t}$ production cross section in the dilepton channel in Run I showed a slight excess over SM predictions [2]. Perhaps more interestingly, however, several of the events observed in the Run I data had $\not{E}_T$ and lepton E_T's large enough to call into question their compatibility with SM top decay kinematics. In fact, it was proposed that the kinematics of these events could be better described by the cascade decays of heavy squarks [3].

Motivated by these curiosities in the top dilepton sample, we devised a search for new physics based on the comparison of kinematic features of observed events with those expected from SM. The search is designed to be sensitive to any physics that gives rise to events with kinematics different from those expected from SM top and backgrounds, especially physics that results in kinematics similar to the aforementioned Run I events. The method seeks to isolate the most unlikely subset of events in a data sample and assign an *a priori* probability that quantifies its departure from the SM. We present the results of this search for anomalous kinematics in top dilepton events observed at the upgraded Collider Detector at Fermilab (CDF II) using 193 pb^{-1} of TeVatron Run II $p\bar{p}$ collision data, a sample nearly twice as large as that used in Run I.

The CDF detector is described in detail in [4].

II. DATA SAMPLE & EVENT SELECTION

The event selection is identical to the one used in the CDF II $t\bar{t}$ cross section measurement in the dilepton channel and referred as "DIL" in [5]. Two high-P_T ($P_T > 20$ GeV/c) oppositely charged leptons (e or μ) are required with at least one of them well isolated from calorimeter activity. Events with an electron identified as a conversion and muon events containing a cosmic ray are removed. The missing energy transverse to the beam direction corrected for escaping muon momenta is required to be $\not{E}_T > 25$ GeV. To ensure that the $\not{E}_T$ is not due to mismeasurements of the energies of the leptons or jets, we impose $\not{E}_T > 50$ GeV cut, if the azimuthal angle between the direction of the $\not{E}_T$ vector and the nearest lepton/jet is less than 20°. Jets are reconstructed in the calorimeter using a fixed-size cone algorithm, with a cone size of $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2} = 0.4$. At least two jets with $E_T > 15$ GeV within $|\eta| < 2.5$ are required. To reduce Drell-Yan background in the dilepton (ee, $\mu\mu$) mass range within Z mass peak (76-106 GeV) a cut on the ratio of $\not{E}_T$ to the sum of the jet E_T's projected along the $\not{E}_T$ vector is applied. The $t\bar{t}$ signal purity is further enhanced by requiring that H_T, the scalar sum of the lepton P_T, jet E_T and $\not{E}_T$, be greater than 200 GeV. These selection criteria result in a high purity of $t\bar{t}$ signal over SM background, $S/B \approx 3.0$. When applied to the data, a total of 13 events are observed whereas 10.9 ± 1.4 are expected.

The dominant SM backgrounds to dilepton $t\bar{t}$ are Drell-Yan, diboson production and events with "fake" leptons, where a jet is falsely reconstructed as a lepton candidate. The involved SM processes are simulated with PYTHIA and ALPGEN+HERWIG Monte Carlo simulation and further reconstructed in CDF II detector software. The kinematic distributions of the events with "fake" leptons are determined from the data.

III. KINEMATIC VARIABLES

We expect a possible new physics signal to reveal itself in distortions of kinematic distributions and make only heuristic assumptions about the nature of non-SM events, such as what observable quantities may potentially be sensitive to new physics. Since the Tevatron operates at the energy frontier, a new physics signature at CDF II is likely to be associated with new heavy particles or new phenomena beyond the presently known mass spectrum. Guided by the possible hints from the Run I data sample, we focus our search on events with large lepton p_T and large $\not{E}_T$ resulting from the decay of a heavy unknown particle. In addition, two body decays (*e.g.* $\tilde{\chi}^{\pm} \rightarrow l^{\pm}\tilde{\nu}$) from massive particles result in these quantities being back-to-back, while this is less prominent for the SM $t\bar{t}$ dilepton signature. Thus we chose kinematic variables sensitive to a wide range of new physics: the event $\not{E}_T$, the P_T for the leading lepton, and the angle between the leading lepton and the direction of the $\not{E}_T$ in the plane transverse to beam, Φ_{lm}.

The high purity of the selection criteria allows us to select an additional variable intrinsic to the properties of the $t\bar{t}$ dilepton decay. The initial and intermediate state particles impose constraints on the final state product properties,

$m(l_1\nu_1) = m(l_2\nu_2) = m_W$ and $m(b_1l_1\nu_1) = m(b_2l_2\nu_2) = m_t$, where we assume $m_t = 175$ GeV. The system of kinematic equations is solved for undetected neutrino momenta and a top dilepton event is reconstructed based on the observed information. The sum of the two neutrino transverse momenta should match the observed missing $\vec{\not{E}}_T$. Due to various detector effects such as calorimeter nonuniformity and jet energy resolution, the equality between these two is only approximate. We therefore introduce a variable which defines how well an event topology satisfies the top dilepton kinematics based on the difference between predicted and measured values of $\not{E}_T$:

$$T = \int \exp\left\{-\frac{(\vec{\not{E}}_T^{\ predicted} - \vec{\not{E}}_T^{\ measured})^2}{2\sigma^2_{\not{E}_T}}\right\} d\vec{\not{E}}_T^{\ predicted} \tag{1}$$

where the integral is performed over the phase space of possible solutions of $\vec{\not{E}}_T^{\ predicted}$ enhanced by accounting for detector resolution in the lepton and jet momenta and an ambiguity in the lepton-b-jet pairing. We adopt gaussian resolutions for electrons and muons with $\sigma_E/E = 13.5\%/\sqrt{E} \bigoplus 2\%$ and $\sigma(1/p) = 0.11\%$. Jets are smeared according to the inherent energy resolution of the hadronic calorimeter and due to gluon radiation. The $\vec{\not{E}}_T^{\ predicted}$ value is then recomputed according to the fluctuated jet and lepton energies. For events with more than two jets, only two leading jets are considered. $\sigma_{\not{E}_T} = 4$ GeV represents uncertainty due to mismeasurement of the underlying event. A non-$t\bar{t}$ dilepton event has on average a smaller value of T.

IV. METHOD

Next, we apply the following algorithm to the data sample. Each event is assigned a weight (a measure of interestingness of an event) as follows:

$$W = (w_{\not{E}_T} \cdot w_{P_{Tl}} \cdot w_{\Phi_{lm}} \cdot w_T)^{1/4}, \tag{2}$$

where $w_{\not{E}_T}, w_{P_{Tl}}, w_{\Phi_{lm}}$ represent probabilities (assuming the SM) for an event to have a $\not{E}_T$, P_{Tl}, Φ_{lm} larger than that observed and w_T is a probability to observe a T smaller than that observed, respectively. Next, for each $K = 1, \ldots, 13$, subsets of K events (K-subsets) are formed, starting with the event with the lowest weight W. By selecting these "unlikely" subsets (where "unlikely" is here defined as having large p_T^ℓ, $\not{E}_T$, Φ_{lm}, and/or small T) we increase the sensitivity of our method to new physics; the "dilution" of a possible signal from inclusion of SM events is minimized.

To quantify the departure of the K-subsets from the SM predictions we do a shape comparison using the Kolmogorov-Smirnov (KS) statistic [6], which represents a maximum deviation between the cumulative functions of two distributions in question. For each of the 4 variables $i = 1, \ldots, 4$ the KS statistic $\Delta_{K,i}$ between the SM cumulative function and the cumulative function of the K-subset is computed. To assess a probability of a deviation we generate a large number of pseudoexperiments by randomly drawing events from large Monte Carlo samples of $t\bar{t}$ and respective SM backgrounds. The number of events corresponding to each SM process is a Poisson smeared estimate based on its kinematic and geometrical acceptance. Only pseudoexperiments with a total of 13 events are accepted. Further, in each pseudoexperiment, K-subsets are formed and the respective $\Delta_{K,i}$ for each are calculated. By doing this we build probability distribution functions $\mathcal{W}_{K,i}(\Delta_{K,i})$. The probability of SM consistency of each K-subset in the variable i is given by

$$p_{K,i} = \int_{\Delta_{K,i}}^{\infty} \mathcal{W}_{K,i}(\Delta) d\Delta \tag{3}$$

Next the information from each variable is combined into

$$\Pi_K = (p_{K,1} \cdot p_{K,2} \cdot p_{K,3} \cdot p_{K,4})^{1/4} \tag{4}$$

The values of Π_K are calculated for each pseudoexperiment and the probability distribution functions $\mathcal{M}_K(\Pi_K)$ are formed. The quantity

$$P_K = \int_0^{\Pi_K} \mathcal{M}_K(\Pi) d\Pi \tag{5}$$

determines how well each K-subset stands in agreement with the theoretical expectation based on the combined information from 4 variables.

We define $K = Q$, where P_K achieves a minimum among the subsets. The quantity

$$\mathcal{P} = P_Q = \min_K P_K \tag{6}$$

is adopted as our test statistic, used to quantify the discrepancy of the data with the SM. The probability distribution function $\mathcal{L}(\mathcal{P})$ is built by generating yet another set of pseudoexperiments from SM Monte Carlo. For each pseudoexperiment the above procedure is repeated and the value of $\mathcal{P}$ is determined. Following the same *a priori* defined steps for the data sample we determine the significance of the departure from the SM of the unlikely Q-subset of events as follows:

$$\alpha = \int_0^{\mathcal{P}_{data}} \mathcal{L}(\mathcal{P})\ d\mathcal{P} \tag{7}$$

Sufficiently low values of α would indicate the presence of new physics in the data sample, and the Q events that comprise the most unlikely subset would exhibit the largest deviation from SM in our four kinematic variables. We test the robustness of the method by simulating a SUSY sample with parameters suggested in [3] as a possible explanation of the anomalous events observed in the Tevatron Run I. We adopt the SM kinematic distributions as a null-hypothesis $H0$ and a 50%:50% mixture of the SM and the SUSY as an alternative hypothesis $H1$. We test the sensitivity of the method by comparing the respective probability distribution functions $\mathcal{L}_{H0}(P)$ and $\mathcal{L}_{H1}(\mathcal{P})$, which are built based on pseudoexperiments drawn from the $H0/H1$ hypotheses, respectively, whereas the KS test in both cases is performed with the $H0$ distributions. We observe that with 13-event $\mathrm{t\bar{t}}$ dilepton samples, we would observe less than 1% consistency with the SM 60% of the time.

V. VALIDATION

We test our procedure, as well as our ability to correctly model events with leptons, jets, and $\not\!\!E_T$, in a high statistics control sample of W + 3 jets data events. We select events with one and only one central isolated lepton, three or more jets, and $\not\!\!E_T > 25$ GeV. To reduce QCD background the azimuthal angle, for events with $\not\!\!E_T < 35$ GeV, $\Delta\phi$ between the most energetic jet and $\vec{\not\!\!E}_T$ is required to be within [0.5,2.5] radians. The total sample size corresponds to the same integrated luminosty as the dilepton sample. We observe 973 events. The dominant contributing process are W bosons produced in with associated jet activity which we model with W + 3 parton ALPGEN+HERWIG Monte Carlo. $\mathrm{t\bar{t}} \rightarrow \mathrm{b}l\nu\mathrm{\bar{b}q\bar{q}}$ accounts for $11.2\% \pm 0.6\%$ of the events based on the theoretical $\mathrm{t\bar{t}}$ cross section of $6.7pb^{-1}$, modeled with PYTHIA. The QCD background constitutes $8.8\% \pm 6.6\%$ of the sample events and its shape is estimated from data.

We compare data with a Monte Carlo simulation of $\not\!\!E_T$, P_{Tl} and Φ_{lm} using these SM processes added in the relative amounts estimated above. We observe a good fit to the data (see Figure 1). We apply a 3-dimensional version of this technique and observe that the data indicates good consistency with the SM ($\alpha = 35.1\%$), We test the modelling of T in the conrol sample of W + 4 jets events using a similar event selection as given above. W + 4parton Monte Carlo simulation is used. We attempt to reconstruct these events as if they were $\mathrm{t\bar{t}}$ dilepton events, by treating the leading jet as a second lepton, the next highest E_T jet as a second neutrino, and two other jets in the event as b-quark jets. Applying this reconstruction to both Monte Carlo and data events, we observe very good consistency between the respective T distributions (K-S probability of 0.97).

VI. RESULTS

The kinematic distributions of the selected variables for $\mathrm{t\bar{t}}$ dilepton events are presented in Figure 2. Applying the outlined technique to the $\mathrm{t\bar{t}}$ dilepton sample we find the most unlikely subset of events to be the entire data set, with a p-value = 1.6%. This result is entirely driven by the excess of leptons at low p_T; since the method effectively orders the subsets from high p_T to low p_T, the p-value decreases as more of the low-p_T excess is included, reaching a minimum when the entire data sample is considered.

In order to better understand the nature of the low-p_T excess, we have investigated these events in more detail. Since top quark events are characterized by the presence of long-lived B hadron decays in the final state, we identify events with jets containing a displaced secondary vertex as likely top events. We present the bottom quark content of the sample as well as the distribution of events in the (P_T, T) plane in Figure 3. Six of the nine low-p_T events contain at least one identified b-jet, and about half are quite consistent with the $\mathrm{t\bar{t}}$ kinematic hypothesis as evidenced by their large values of T. In addition, the events do not cluster near the 20 GeV p_T threshold. All of this indicates that the low-p_T events are more likely to be top than background; details of the thirteen events can be found elsewhere [7].

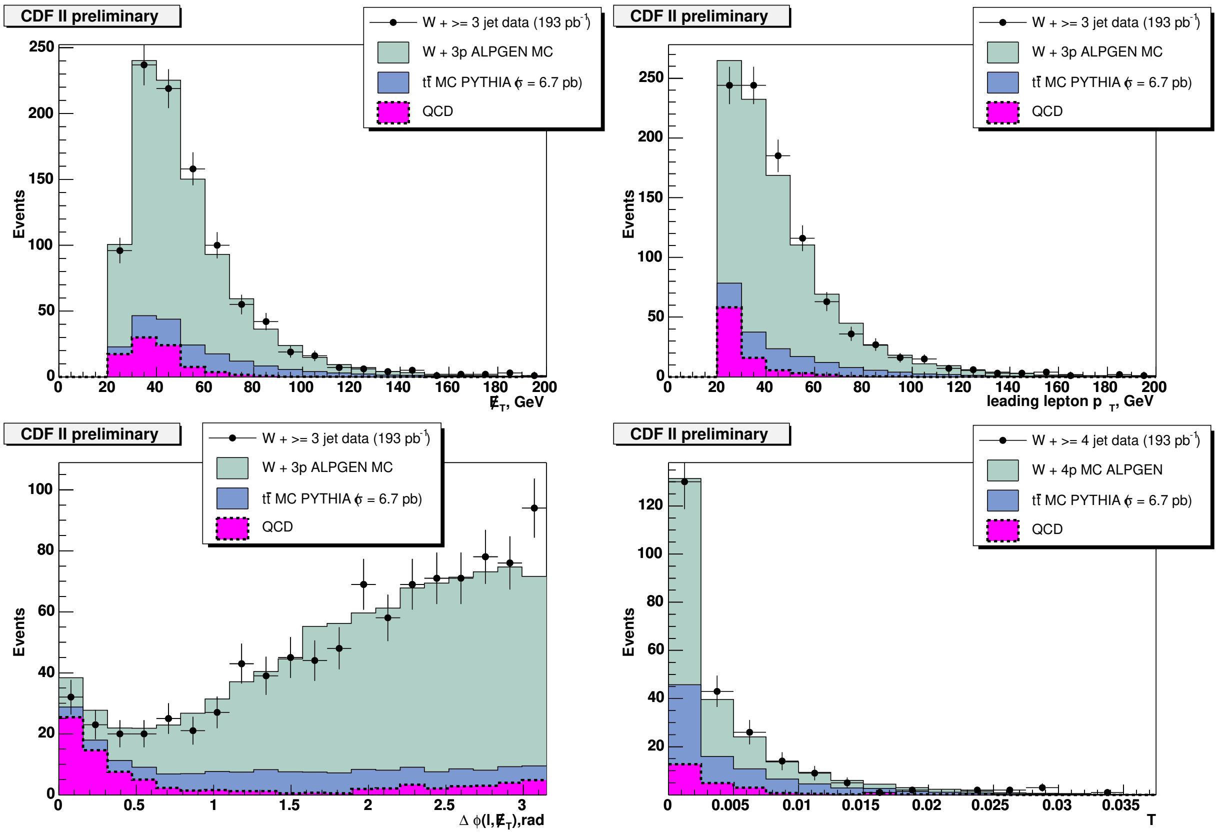

FIG. 1: $\not\!\!E_T$, leading lepton P_T, Φ_{lm}, and T distributions in the W + jets samples.

VII. SYSTEMATIC UNCERTAINTIES

Due to systematic uncertainties, the shapes of the kinematic distributions are not predicted to infinite precision. This leads to an uncertainty in the probability distribution function $\mathcal{L}(\mathcal{P})$, and respectively to uncertainties in the significance level of our mesurement. To quantify this uncertainty we model deviations in the $\mathcal{L}(\mathcal{P})$ shape due to various systematic effects that give distortions in the SM kinematic distributions. We treat each source of systematic uncertainty as an alternative hypothesis $H1$. We generate a large number of $H1$ pseudoexepriments and build a probability distribution function $\mathcal{L}_{H1}(\mathcal{P})$. The respective p-value due to the uncertainty in question is then given by

$$\alpha' = \int_0^{\mathcal{P}_{data}} \mathcal{L}_{H1}(\mathcal{P})\; d\mathcal{P} \qquad (8)$$

The systematic uncertanties due to various sources are estimated and given in Table I. The combined systematic uncertainty is estimated based on a comparison of the $H0$ distribution with the $H1$ distribution when all systematic effects deteriorating the result are included, resulting in p-values in the range 1.0–4.5%. The major uncertainties are due to jet energy scale, SM background predictions, top mass and uncertainty due to PDFs.

VIII. CONCLUSIONS

We have assessed the $t\bar{t}$ dilepton sample's consistency with the SM in the four-variable space described and find a probability of 1.0–4.5%. The $\not\!\!E_T$, Φ_{lm}, and T distributions of the data are quite consistent with SM expectations; the lepton p_T distribution exhibits a mild excess at low p_T that is consistent with a statistical fluctuation of SM top. No anomalies are seen in the kinematic regions expected to be populated by events containing new heavy particles. It can be concluded that new physics scenarios invoked to describe the high-p_T^{ℓ}/high-$\not\!\!E_T$ events observed in Run I are not favored by the Run II data.

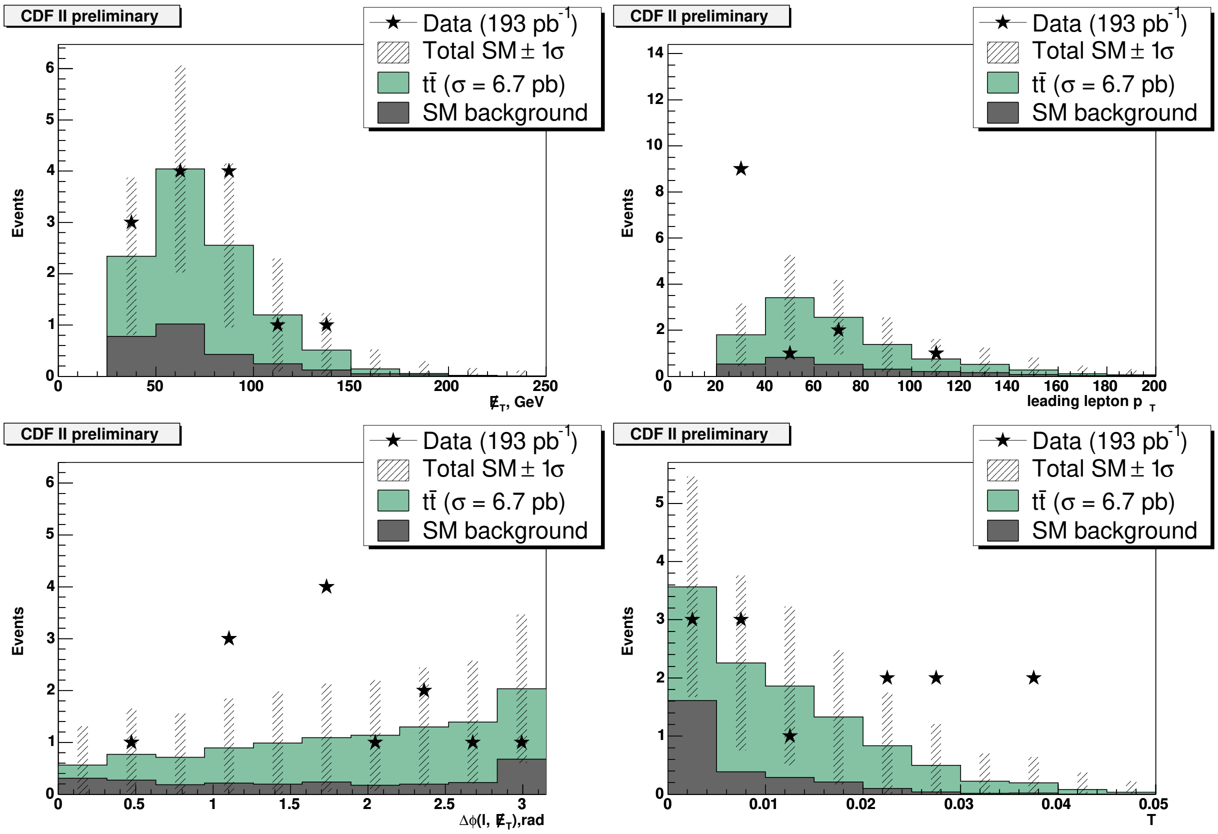

FIG. 2: $\not\!E_T$, leading lepton P_T, Φ_{lm}, and T distributions in the top dilepton sample.

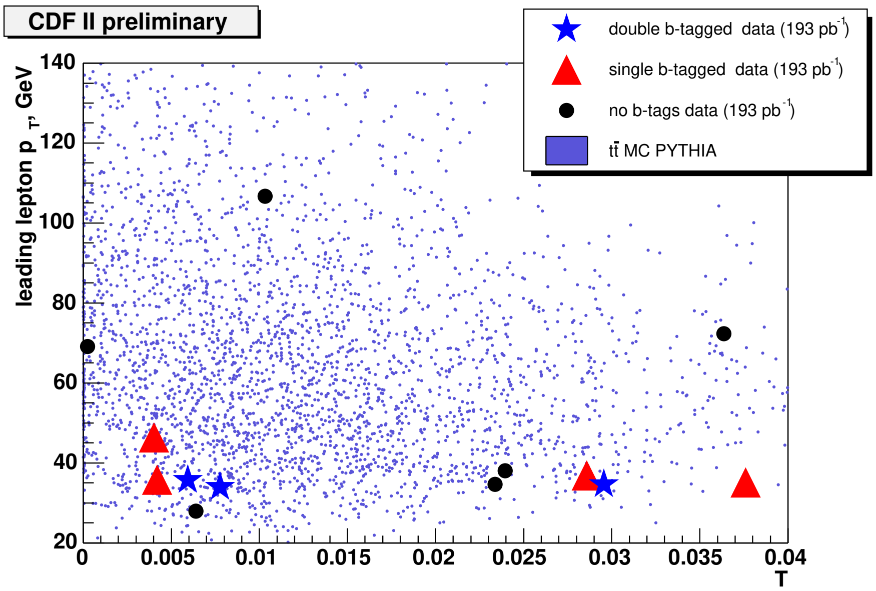

FIG. 3: Top dilepton events in (P_T, T) plane with b-tagging information.

TABLE I: Effect of systematic uncertainties from various sources.

Source of uncertainty	α'
MC generator	0.016
Initial state radiation	0.012
Final state radiation	0.016
Parton distribution functions	0.019
M_{top} = 170 GeV	0.014
M_{top} = 180 GeV	0.021
Jet energy scale, +1 σ	0.021
Jet energy scale, -1 σ	0.026
Background estimates, +1 σ	0.027
Background estimates, -1 σ	0.010
combined, -1 σ	0.010
combined, +1 σ	0.045

Acknowledgments

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 fuer 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 Russian Foundation for Basic Research; the Comision Interministerial de Ciencia y Tecnologia, Spain; in part by the European Community's Human Potential Programme under contract HPRN-CT-20002, Probe for New Physics; by the Research Fund of Istanbul University Project No. 1755/21122001; and by the Research Corporation.

[1] F. Abe, et al.(CDF), Evidence for top quark production in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV, Phys. Rev. **D50**, 2966 (1994);
F. Abe, et al.(CDF), Observation of top quark production in $p\bar{p}$ collisions, Phys. Rev. Lett. **74**, 2626 (1995), hep-ex/9503002;
S.Abachi, et al.(D0), Observation of the top quark, Phys. Rev. Lett. **74**, 2632 (1995), hep-ex/9503003;

[2] F. Abe, et al.(CDF), Measurement of the top quark mass and $t\bar{t}$ production cross section from dilepton events at the Collider Detector at Fermilab, Phys. Rev. Lett.**80**, 2779 (1998), hep-ex/9802017;
S. Abachi, et al.(D0), Measurement of the top quark pair production cross section in $p\bar{p}$ collisions, Phys. Rev. Lett.**79**, 1203 (1997), hep-ex/9704015;

[3] M.Barnett, L.Hall, A multilepton signal for supersymmetric particles in Tevatron data?, Phys. Rev. Lett.**77**, 3506 (1996), hep-ph/9607342;

[4] The CDFII Detector Technical Design Report, Fermilab-Pub-96/390-E

[5] D. Acosta, et al.(CDF), Measurement of the $t\bar{t}$ production cross section in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV using dilepton events, submitted to Phys. Rev. Lett., 2779 (2004), hep-ex/0404036;

[6] A. Kolmogorov, On the empirical determination of a distribution function, Giornale dell'Istituto Italiano degli Attuari (1933);

[7] A. Ivanov, Ph.D. thesis, University of Rochester, in preparation;