Heavy flavor properties of jets produced in $p\bar{p}$ interactions at $\sqrt{s}=1.8$ TeV

D. Acosta,[11] D. Ambrose,[31] K. Anikeev,[23] J. Antos,[1] G. Apollinari,[10] T. Arisawa,[41] A. Artikov,[8] F. Azfar,[29] P. Azzi-Bacchetta,[30] N. Bacchetta,[30] V. E. Barnes,[33] B. A. Barnett,[18] M. Barone,[12] G. Bauer,[23] F. Bedeschi,[32] S. Behari,[18] S. Belforte,[38] W. H. Bell,[14] G. Bellettini,[32] J. Bellinger,[42] A. Beretvas,[10] A. Bhatti,[35] D. Bisello,[30] C. Blocker,[3] B. Blumenfeld,[18] A. Bocci,[35] G. Bolla,[33] A. Bolshov,[23] D. Bortoletto,[33] J. Budagov,[8] H. S. Budd,[34] K. Burkett,[10] G. Busetto,[30] S. Cabrera,[9] W. Carithers,[21] D. Carlsmith,[42] R. Carosi,[32] A. Castro,[2] D. Cauz,[38] A. Cerri,[21] C. Chen,[31] Y. C. Chen,[1] G. Chiarelli,[32] G. Chlachidze,[8] M. L. Chu,[1] W.-H. Chung,[42] Y. S. Chung,[34] A. G. Clark,[13] M. Coca,[34] M. Convery,[35] M. Cordelli,[12] J. Cranshaw,[37] D. Dagenhart,[3] S. D'Auria,[14] S. Dell'Agnello,[12] P. de Barbaro,[34] S. De Cecco,[36] M. Dell'Orso,[32] L. Demortier,[35] M. Deninno,[2] D. De Pedis,[36] C. Dionisi,[36] S. Donati,[32] M. D'Onofrio,[13] T. Dorigo,[30] R. Eusebi,[34] S. Farrington,[14] J. P. Fernandez,[33] R. D. Field,[11] I. Fiori,[32] A. Foland,[15] L. R. Flores-Castillo,[33] M. Franklin,[15] J. Friedman,[23] I. Furic,[23] M. Gallinaro,[35] A. F. Garfinkel,[33] E. Gerstein,[7] S. Giagu,[32] P. Giannetti,[32] K. Giolo,[33] M. Giordani,[38] P. Giromini,[12] V. Glagolev,[8] G. Gomez,[6] I. Gorelov,[26] A. T. Goshaw,[9] K. Goulianos,[35] A. Gresele,[2] M. Guenther,[33] J. Guimaraes da Costa,[15] E. Halkiadakis,[34] C. Hall,[15] R. Handler,[42] F. Happacher,[12] K. Hara,[39] F. Hartmann,[19] K. Hatakeyama,[35] J. Hauser,[5] J. Heinrich,[31] M. Hennecke,[19] M. Herndon,[18] A. Hocker,[34] S. Hou,[1] B. T. Huffman,[29] G. Introzzi,[32] M. Iori,[36] A. Ivanov,[34] Y. Iwata,[16] B. Iyutin,[23] M. Jones,[33] M. Karagoz Unel,[27] K. Karr,[40] Y. Kato,[28] B. Kilminster,[34] D. H. Kim,[20] M. J. Kim,[7] S. B. Kim,[20] S. H. Kim,[39] T. H. Kim,[23] M. Kirby,[9] L. Kirsch,[3] S. Klimenko,[11] K. Kondo,[41] J. Konigsberg,[11] A. Korn,[23] A. Korytov,[11] K. Kotelnikov,[25] J. Kroll,[31] M. Kruse,[9] A. T. Laasanen,[33] S. Lami,[35] S. Lammel,[10] J. Lancaster,[9] G. Latino,[26] Y. Le,[18] J. Lee,[34] N. Leonardo,[23] S. Leone,[32] M. Lindgren,[5] N. S. Lockyer,[31] A. Loginov,[25] M. Loreti,[30] D. Lucchesi,[30] S. Lusin,[42] L. Lyons,[29] R. Madrak,[15] P. Maksimovic,[18] L. Malferrari,[2] M. Mariotti,[30] M. Martin,[18] V. Martin,[27] M. Martínez,[10] P. Mazzanti,[2] M. Menguzzato,[30] A. Menzione,[32] C. Mesropian,[35] A. Meyer,[10] S. Miscetti,[12] G. Mitselmakher,[11] Y. Miyazaki,[28] N. Moggi,[2] M. Mulhearn,[23] T. Muller,[19] A. Munar,[31] P. Murat,[10] I. Nakano,[16] R. Napora,[18] S. H. Oh,[9] Y. D. Oh,[20] T. Ohsugi,[16] T. Okusawa,[28] C. Pagliarone,[32] F. Palmonari,[32] R. Paoletti,[32] V. Papadimitriou,[37] A. Parri,[12] G. Pauletta,[38] T. Pauly,[29] C. Paus,[23] A. Penzo,[38] T. J. Phillips,[9] G. Piacentino,[32] J. Piedra,[6] K. T. Pitts,[17] A. Pompoš,[33] L. Pondrom,[42] T. Pratt,[29] F. Prokoshin,[8] F. Ptohos,[12] O. Poukhov,[8] G. Punzi,[32] J. Rademacker,[29] A. Rakitine,[23] H. Ray,[24] A. Reichold,[29] P. Renton,[29] M. Rescigno,[36] F. Rimondi,[2] L. Ristori,[32] W. J. Robertson,[9] T. Rodrigo,[6] S. Rolli,[40] L. Rosenson,[23] R. Rossin,[30] C. Rott,[33] A. Roy,[33] A. Ruiz,[6] D. Ryan,[40] A. Safonov,[4] W. K. Sakumoto,[34] D. Saltzberg,[5] L. Santi,[38] S. Sarkar,[36] A. Savoy-Navarro,[10] P. Schlabach,[10] M. Schmitt,[27] L. Scodellaro,[30] A. Scribano,[32] A. Sedov,[33] S. Seidel,[26] Y. Seiya,[39] A. Semenov,[8] F. Semeria,[2] T. Shibayama,[39] M. Shimojima,[39] A. Sidoti,[32] A. Sill,[37] K. Sliwa,[40] R. Snihur,[22] M. Spezziga,[37] F. Spinella,[32] M. Spiropulu,[15] R. St. Denis,[14] A. Stefanini,[32] A. Sukhanov,[11] K. Sumorok,[26] T. Suzuki,[39] R. Takashima,[16] K. Takikawa,[39] P. K. Teng,[1] K. Terashi,[35] S. Tether,[26] A. S. Thompson,[14] D. Tonelli,[32] J. Tseng,[26] D. Tsybychev,[11] N. Turini,[32] F. Ukegawa,[39] T. Unverhau,[14] E. Vataga,[32] G. Velev,[10] I. Vila,[6] R. Vilar,[6] M. von der Mey,[5] W. Wagner,[19] C. Wang,[9] M. J. Wang,[1] S. M. Wang,[11] B. Ward,[14] S. Waschke,[14] B. Whitehouse,[40] H. H. Williams,[31] M. Wolter,[40] X. Wu,[13] K. Yi,[18] T. Yoshida,[28] I. Yu,[20] S. Yu,[31] L. Zanello,[36] A. Zanetti,[38] and S. Zucchelli[2]

(CDF Collaboration)

[1]*Institute of Physics, Academia Sinica, Taipei, Taiwan 11529, Republic of China*
[2]*Istituto Nazionale di Fisica Nucleare, University of Bologna, I-40127 Bologna, Italy*
[3]*Brandeis University, Waltham, Massachusetts 02254, USA*
[4]*University of California at Davis, Davis, California 95616, USA*
[5]*University of California at Los Angeles, Los Angeles, California 90024, USA*
[6]*Instituto de Fisica de Cantabria, CSIC-University of Cantabria, 39005 Santander, Spain*
[7]*Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA*
[8]*Joint Institute for Nuclear Research, RU-141980 Dubna, Russia*
[9]*Duke University, Durham, North Carolina 27708, USA*
[10]*Fermi National Accelerator Laboratory, Batavia, Illinois 60510, USA*
[11]*University of Florida, Gainesville, Florida 32611, USA*
[12]*Laboratori Nazionali di Frascati, Istituto Nazionale di Fisica Nucleare, I-00044 Frascati, Italy*
[13]*University of Geneva, CH-1211 Geneva 4, Switzerland*
[14]*Glasgow University, Glasgow G12 8QQ, United Kingdom*
[15]*Harvard University, Cambridge, Massachusetts 02138, USA*
[16]*Hiroshima University, Higashi-Hiroshima 724, Japan*
[17]*University of Illinois, Urbana, Illinois 61801, USA*
[18]*The Johns Hopkins University, Baltimore, Maryland 21218, USA*
[19]*Institut für Experimentelle Kernphysik, Universität Karlsruhe, 76128 Karlsruhe, Germany*
[20]*Center for High Energy Physics: Kyungpook National University, Taegu 702-701; Seoul National University, Seoul 151-742 and SungKyunKwan University, Suwon 440-746; Korea*
[21]*Ernest Orlando Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA*

[22]*University College London, London WC1E 6BT, United Kingdom*
[23]*Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*
[24]*University of Michigan, Ann Arbor, Michigan 48109, USA*
[25]*Institution for Theoretical and Experimental Physics, ITEP, Moscow 117259, Russia*
[26]*University of New Mexico, Albuquerque, New Mexico 87131, USA*
[27]*Northwestern University, Evanston, Illinois 60208, USA*
[28]*Osaka City University, Osaka 588, Japan*
[29]*University of Oxford, Oxford OX1 3RH, United Kingdom*
[30]*Universita di Padova, Istituto Nazionale di Fisica Nucleare, Sezione di Padova, I-35131 Padova, Italy*
[31]*University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA*
[32]*Istituto Nazionale di Fisica Nucleare, University and Scuola Normale Superiore of Pisa, I-56100 Pisa, Italy*
[33]*Purdue University, West Lafayette, Indiana 47907, USA*
[34]*University of Rochester, Rochester, New York 14627, USA*
[35]*Rockefeller University, New York, New York 10021, USA*
[36]*Istituto Nazionale de Fisica Nucleare, Sezione di Roma, University di Roma I, "La Sapienza," I-00185 Roma, Italy*
[37]*Texas Tech University, Lubbock, Texas 79409, USA*
[38]*Istituto Nazionale di Fisica Nucleare, University of Trieste/Udine, Italy*
[39]*University of Tsukuba, Tsukuba, Ibaraki 305, Japan*
[40]*Tufts University, Medford, Massachusetts 02155, USA*
[41]*Waseda University, Tokyo 169, Japan*
[42]*University of Wisconsin, Madison, Wisconsin 53706, USA*
(Received 2 December 2003; published 30 April 2004)

We present a detailed examination of the heavy flavor properties of jets produced at the Fermilab Tevatron collider. The data set, collected with the Collider Detector at Fermilab, consists of events with two or more jets with transverse energy $E_T \geq 15$ GeV and pseudorapidity $|\eta| \leq 1.5$. The heavy flavor content of the data set is enriched by requiring that at least one of the jets (lepton-jet) contains a lepton with a transverse momentum larger than 8 GeV/c. Jets containing hadrons with heavy flavor are selected via the identification of secondary vertices. The parton-level cross sections predicted by the HERWIG Monte Carlo generator program are tuned within theoretical and experimental uncertainties to reproduce the secondary-vertex rates in the data. The tuned simulation provides new information on the origin of the discrepancy between the $b\bar{b}$ cross section measurements at the Tevatron and the next-to-leading order QCD prediction. We also compare the rate of away-jets (jets recoiling against the lepton-jet) containing a soft lepton ($p_T \geq 2$ GeV/c) in the data to that in the tuned simulation. We find that this rate is larger than what is expected for the conventional production and semileptonic decay of pairs of hadrons with heavy flavor.

DOI: 10.1103/PhysRevD.69.072004 PACS number(s): 13.85.Qk, 13.20.Fc, 13.20.He

I. INTRODUCTION

This paper presents a study of semileptonic decays in jets containing heavy flavor and is motivated by several anomalies that have been previously reported. The Collider Detector at Fermilab (CDF) has found the rate of jets with both a secondary vertex and a soft lepton (superjets) to be larger than expected in the $W+2,3$ jet sample. The kinematical properties of the events with a superjet are difficult to reconcile with the standard model (SM) expectation [1].

The discrepancy between the single bottom quark cross section measurements at the Fermilab Tevatron and the next-to-leading order (NLO) QCD prediction [2] can be explained either in terms of new physics [3] or by the lack of robustness of the NLO prediction [4]. However, at the Tevatron there are two additional discrepancies between the measured and predicted value of the $b\bar{b}$ cross section that are more difficult to accommodate within the theoretical uncertainty. In Ref. [5], the correlated $\mu+\bar{b}$ jet cross section is measured to be 1.5 times larger than $\sigma_{b\bar{b}} \times \mathrm{BR}$, where BR is the average semileptonic branching ratio of b hadrons produced at the Tevatron and $\sigma_{b\bar{b}}$ is the NLO prediction of the cross section for producing pairs of b and $\bar{b}$ quarks. A further discrepancy is found by both CDF and DØ experiments [6,7] when comparing the cross section for producing dimuons from b-hadron semileptonic decays to $\sigma_{b\bar{b}} \times \mathrm{BR}^2$. The value of $\sigma_{b\bar{b}} \times \mathrm{BR}^2$ is found to be approximately 2.2 times larger than the NLO prediction.[1] There are possible conventional explanations presented in the literature for these anomalies [9,10].

However, all these discrepancies could also be mitigated by postulating the existence of a light strong-interacting object with a 100% semileptonic branching ratio. Since there

[1]In both measurements, $\sigma_{b\bar{b}}$ is the cross section for producing two central bottom quarks, both with transverse momentum approximately larger than 10 GeV/c. In this case, the LO and NLO predictions are equal within a few percent, and the NLO prediction changes by no more than 15% when changing the renormalization and factorization scales by a factor of two [8].

are no limits to the existence of a charge-1/3 scalar quark with mass smaller than 7.4 GeV/c^2 [11–15], the supersymmetric partner of the bottom quark is a potential candidate. This paper presents an analysis of multijet data intended to search for evidence either supporting or disfavoring this hypothesis.

The strategy of this search is outlined in Sec. II. Section III describes the detector systems relevant to this analysis, while the sample selection and the tagging algorithms (SECVTX and JPB) used to select heavy flavors are discussed in Sec. IV. Section V describes the data sample composition and the heavy flavor simulation. The data set consists of events with two or more jets with transverse energy $E_T \geqslant 15$ GeV and contained in the silicon microvertex detector (SVX) acceptance. The sample is enriched in heavy flavor by requiring that at least one of the jets contains a lepton with $p_T \geqslant 8$ GeV/c. We use measured rates of SECVTX and JPB tags to determine the bottom and charmed content of the data; we then tune the simulation to match the heavy-flavor content of the data. The evaluation of the number of SECVTX and JPB tags due to heavy flavor in the data and the simulation is described in Secs. VI and VII, respectively. The tuning of the heavy flavor production cross sections in the simulation is described in Sec. VIII. In Sec. IX we measure the yields of jets containing soft leptons ($p_T \geqslant 2$ GeV/c), and compare them to the prediction of the tuned simulation. Section X contains cross-checks and a discussion of the systematic uncertainties. Our conclusions are presented in Sec. XI.

II. PROBING THE PRODUCTION OF LIGHT SCALAR QUARKS WITH A LARGE SEMILEPTONIC BRANCHING RATIO

In previous publications [1,16] we have compared the b- and c-quark content of several samples of generic-jet data to the QCD prediction of the standard model using the HERWIG generator program [17]. We identify (tag) jets produced by heavy quarks using the CDF silicon microvertex detector (SVX) to locate secondary vertices produced by the decay of b and c hadrons inside a jet. These vertices (SECVTX tags) are separated from the primary event vertex as a result of the long b and c lifetime. We also use track impact parameters to select jets with a small probability of originating from the primary vertex of the event (JPB tags) [18].

In Ref. [16] we have compared rates of SECVTX and JPB tags in generic-jet data and their simulation first to calibrate the efficiency of the tagging algorithms in the simulation, and then to tune the heavy flavor cross sections evaluated with the HERWIG parton-shower Monte Carlo program. In the simulation, jets with heavy flavor are produced by heavy quarks in the initial or final state of the hard scattering (flavor excitation and direct production, respectively) or from gluons branching into $b\bar{b}$ or $c\bar{c}$ pairs (gluon splitting). The fraction of generic-jet data containing $b\bar{b}$ or $c\bar{c}$ pairs calculated by HERWIG models correctly the observed rate of tags after minor adjustments within the theoretical and experimental uncertainties. In Refs. [1,16], we have extended this comparison to $W+$ jet events. We find again good agreement between the observed rates of SECVTX and JPB tags and the SM prediction, which includes single and pair production of top quarks.

We also identify heavy flavors by searching jets for leptons (e or μ) produced in the decay of b and c hadrons [1,16]; we refer to these as soft lepton tags (SLT). As shown in Refs. [1,16], rates of SLT tags in generic-jet data and in $W+$ jet events are generally well modeled by the simulation. An exception is the rate of SECVTX+SLT tags in the same jet (called supertags in Ref. [1]) that, in $W+2,3$ jet events, is larger than in the simulation, whereas in generic-jet samples is slightly overpredicted by the same simulation.

This analysis uses two data samples, referred to as the signal or inclusive lepton sample and the control or generic-jet sample. The signal sample consists of events with two or more jets that have been acquired with the trigger request that events contain a lepton with $p_T \geqslant 8$ GeV/c. The request of a jet containing a lepton (lepton-jet) enriches the heavy flavor content of the sample with respect to generic jets. The control or generic-jet sample is the same sample studied in Refs. [1,16], and consists of events with one or more jets acquired with three trigger thresholds of 20, 50 and 100 GeV, respectively.

In the signal sample, we study jets recoiling against the lepton-jet (away-jets) and we perform three measurements: we count the number of away-jets that contain a lepton (SLT tag); that contain an SLT tag and a SECVTX tag; that contain an SLT tag and a JPB tag. The latter two are referred to as supertags. We compare the three measurements to a Monte Carlo simulation which is tuned and normalized to the data by equalizing numbers of SECVTX and JPB tags. The normalization and tuning procedure serves two purposes: it removes the dependence on the efficiency for finding the trigger lepton and ensures that the simulation reproduces the heavy-flavor content of the data, respectively. To calibrate the efficiency for finding SLT tags or supertags in the simulation, we use rates of SLT tags and supertags in generic-jet data (control sample). In Ref. [1], we have compared these measurements to a Monte Carlo simulation that was also tuned and normalized to generic-jet data by equalizing numbers of SECVTX and JPB tags. These three comparisons are used to verify the simulated efficiency for finding SLT tags, and to empirically calibrate the efficiency for finding supertags in the simulation.

This analysis strategy is motivated by the following argument. If low-mass bottom squarks existed, they would be produced copiously at the Tevatron. The NLO calculation of the process $p\bar{p} \rightarrow \tilde{b}\tilde{b}^*$, implemented in the PROSPINO Monte Carlo generator [19], predicts a cross section that is $\simeq 15\%$ of the NLO prediction for the production cross section of quarks with the same mass [8]. In Ref. [16], we have tuned, within the theoretical and experimental uncertainties, the heavy flavor production cross sections calculated by HERWIG to reproduce the rates of SECVTX and JPB tags observed in generic-jet data. If the squark lifetime is similar to that of conventional heavy flavors, we have unfortunately tuned the parton-level cross section evaluated by HERWIG (or the number of simulated SECVTX and JPB tags predicted by the simulation) to explain in terms of conventional processes the squark production. However, if bottom squarks have a

TABLE I. Comparison between $\sigma = \mathrm{BR}_b \times \sigma_{b\bar{b}} + \mathrm{BR}_c \times \sigma_{c\bar{c}} + \mathrm{BR}_{\tilde{b}} \times \sigma_{\tilde{b}\tilde{b}*}$, the total heavy-flavor production cross section (b, c, and $\tilde{b}$) contributing to different hypothetical samples, and $\sigma^{norm} = \mathrm{BR}_b \times \sigma_{b\bar{b}}^{norm} + \mathrm{BR}_c \times \sigma_{c\bar{c}}$, the total heavy-flavor cross section determined with a conventional-QCD simulation under the hypothesis that scalar quarks have the same lifetime of b quarks ($\sigma_{b\bar{b}}^{norm} = \sigma_{b\bar{b}} + \sigma_{\tilde{b}\tilde{b}*}$). In samples containing leptons, each cross section is also multiplied by the appropriate semileptonic branching ratio BR.

Sample	σ (nb)	σ^{norm} (nb)	σ/σ^{norm}
A = generic jets	$869 = 298 + 487 + 84$	$869 = 382 + 487$	1.0
B = A with one lepton	$296 = 0.37 \times 298 + 0.21 \times 487 + 1.0 \times 84$	$244 = 0.37 \times 382 + 0.21 \times 487$	1.2
C = A with two leptons	$146 = 0.37^2 \times 298 + 0.21^2 \times 487 + 1.0 \times 84$	$74 = 0.37^2 \times 382 + 0.21^2 \times 487$	2.0
D = B renormalized	$296 = 110 + 102 + 84$	$296 = 194 + 102$	1.0
E = D with one lepton	$146 = 0.37 \times 110 + 0.21 \times 102 + 1.0 \times 84$	$93 = 0.37 \times 194 + 0.21 \times 102$	1.5

100% semileptonic branching ratio, it is still possible to identify their presence by comparing the observed number of jets containing a lepton to that expected from b and c decays.

We illustrate the procedure used in this paper with a numeric example detailed in Table I. The first column is what there would be in the data in the presence of $\tilde{b}$ quarks with 100% semileptonic BR.[2] The cross sections in the first column of row A represent approximately the different heavy flavor contributions to the generic-jet sample. The second column is what one would predict after having tuned a simulation, in which only b and c quarks are present, to reproduce the number of SECVTX and JPB tags observed in the sample corresponding to the first column of row A, in the assumption that b and $\tilde{b}$ quarks have the same lifetime. In row B, we model the request that a jet contains a lepton by multiplying the heavy flavor cross sections by the respective semileptonic branching ratios BR. A 20% excess is observed. In row C, we mimic the case in which two jets contain a lepton, and the same analysis leads to an excess of a factor of two. Since a discrepancy that depends on the number of leptons could be due to a wrong simulation of the lepton-identification efficiency, row D presents the stratagem of tuning again the conventional heavy flavor cross sections for producing events with one lepton (second column in row B) to model the cross section contributing to events with one lepton (first column in row B).[3] Next, row E shows the result of requiring an additional lepton in sample D: the excess is a factor of 1.5. If one chooses, as we did in previous studies, to use sample B to empirically correct the simulated efficiency for identifying a lepton, sample E will show a 30% excess.

[2]The cross sections are predicted using the MNR [8] and PROSPINO [19] Monte Carlo generators, the MRS(G) set of structure functions [20], and the renormalization and factorization scales $\mu_0^2 = p_T^2 + m_{\tilde{b}}^2$. We use $m_b = 4.75\ \mathrm{GeV}/c^2$, $m_c = 1.5\ \mathrm{GeV}/c^2$, and $m_{\tilde{b}} = 3.6\ \mathrm{GeV}/c^2$. The cross section are integrated over final-state partons with $p_T \geq 18\ \mathrm{GeV}/c$; this threshold is used to mimic the generic-jet data. Bottom quarks have a 37% semileptonic branching ratio, BR, due to $b \to l$ and $b \to c \to l$ decays, whereas BR = 21% for c quarks [21].

[3]This technique also allows us to use the inclusive lepton sample that corresponds to a much larger integrated luminosity than that of generic-jet data.

III. THE CDF DETECTOR

The events used for this analysis have been collected with the CDF detector during the 1993–1995 run of the Tevatron collider at Fermilab. The CDF detector is described in detail in Ref. [22]. We review the detector components most relevant to this analysis. Inside the 1.4 T solenoid the silicon microvertex detector (SVX) [23], a vertex drift chamber (VTX) and the central tracking chamber (CTC) provide the tracking and momentum information for charged particles. The CTC is a cylindrical drift chamber containing 84 measurement layers. It covers the pseudorapidity interval $|\eta| \leq 1.1$, where $\eta = -\ln[\tan(\theta/2)]$. In the CDF, θ is the polar angle measured from the proton direction, ϕ is the azimuthal angle, and r is the radius from the beam axis (z axis). The SVX consists of four layers of silicon microstrip detectors, located at radii between 2.9 and 7.9 cm from the beam line, and provides spatial measurements in the r-ϕ plane with a resolution of 13 μm.

Electromagnetic (CEM) and hadronic (CHA) calorimeters with projective tower geometry are located outside the solenoid and cover the pseudorapidity region $|\eta| \leq 1.1$, with a segmentation of $\Delta\phi = 15^\circ$ and $\Delta\eta = 0.11$. A layer of proportional chambers (CES) is embedded near shower maximum in the CEM and provides a more precise measurement of the electromagnetic shower position. Coverage at larger pseudorapities is provided by the calorimeters PEM and PHA, and in the far forward region by the FEM and FHA. Two muon subsystems in the central rapidity region ($|\eta| \leq 0.6$) are used for muon identification: the central muon chambers (CMU), located behind the CHA calorimeter, and the central upgrade muon chambers (CMP), located behind an additional 60 cm of steel. The central muon extension (CMX) covers approximately 71% of the solid angle for $0.6 \leq |\eta| \leq 1.0$ and, in this analysis, is used only to increase the soft muon acceptance.

CDF uses a three-level trigger system. At the first two levels, decisions are made with dedicated hardware. The information available at this stage includes energy deposited in the CEM and CHA calorimeters, high-p_T tracks found in the CTC by a fast track processor (CFT), and track segments found in the muon subsystems. The data used in this study were collected using the electron and muon low-p_T triggers. The first two levels of these triggers require a track with $p_T \geq 7.5\ \mathrm{GeV}/c$ found by the CFT. In the case of the electron trigger, the CFT track must be matched to a CEM cluster

with transverse energy $E_T \geq 8$ GeV. In the case of the muon trigger, the CFT track must be matched to a reconstructed track segment in both sets of central muon detectors (CMU and CMP).

At the third level of the trigger, the event selection is based on a version of the off-line reconstruction programs optimized for speed. The lepton selection criteria used by the third level trigger are similar to those described in the next section.

IV. DATA SAMPLE SELECTION AND HEAVY FLAVOR TAGGING

Central electrons and muons that passed the trigger prerequisite are identified with the same criteria used to select the $W+$ jet sample described in Refs. [1,16]. Electron candidates are identified using information from both calorimeter and tracking detectors. We require the following: (1) the ratio of hadronic to electromagnetic energy of the cluster, $E_{had}/E_{em} \leq 0.05$; (2) the ratio of cluster energy to track momentum, $E/p \leq 1.5$; (3) a comparison of the lateral shower profile in the calorimeter cluster with that of test-beam electrons, $L_{shr} \leq 0.2$; (4) the distance between the extrapolated track position and the CES measurement in the r-ϕ and z views, $\Delta x \leq 1.5$ cm and $\Delta z \leq 3.0$ cm, respectively; (5) a χ^2 comparison of the CES shower profile with those of test-beam electrons, $\chi^2_{strip} \leq 20$; (6) the distance between the interaction vertex and the reconstructed track in the z direction, z-vertex match ≤ 5 cm. Fiducial cuts on the electromagnetic shower position, as measured in the CES, are applied to ensure that the electron candidate is away from the calorimeter boundaries and the energy is well measured. Electrons from photon conversions are removed using an algorithm based on track information [16].

Muons are identified by requiring a match between a CTC track and track segments in both the CMU and CMP muon chambers. The following variables are used to separate muons from hadrons interacting in the calorimeter and cosmic rays: (1) an energy deposition in the electromagnetic and hadronic calorimeters characteristic of minimum ionizing particles, $E_{em} \leq 2$ GeV and $E_{had} \leq 6$ GeV, respectively; (2) $E_{em} + E_{had} \geq 0.1$ GeV; (3) the distance of closest approach of the reconstructed track to the beam line in the transverse plane (impact parameter), $d \leq 0.3$ cm; (4) the z-vertex match ≤ 5 cm; (5) the distance between the extrapolated track and the track segment in the muon chamber, $\Delta x = r\Delta\phi \leq 2$ cm.

We select events containing at least one electron with $E_T \geq 8$ GeV or one muon with $p_T \geq 8$ GeV/c. This selection produces a data sample quite similar to that used for the measurement of the B^0-$\overline{B^0}$ flavor oscillation [24]. Since we are interested in semileptonic decays of heavy quarks, trigger leptons are also required to be nonisolated; we require $I \geq 0.1$, where the isolation I is defined as the ratio of the additional transverse energy deposited in the calorimeter in a cone of radius $R = \sqrt{\delta\phi^2 + \delta\eta^2} = 0.4$ around the lepton direction to the lepton transverse energy.

Further selection of the data sample is based upon jet reconstruction. Jets are reconstructed from the energy deposited in the calorimeter using a clustering algorithm with a fixed cone of radius $R = 0.4$. A detailed description of the algorithm can be found in Ref. [25]. Jet energies can be mismeasured for a variety of reasons (calorimeter nonlinearity, loss of low momentum particles because of the magnetic field, contributions from the underlying event, out-of-cone losses, undetected energy carried by muons and neutrinos). Corrections, which depend on the jet E_T and η, are applied to jet energies; they compensate for these mismeasurements on average but do not improve the jet energy resolution. In this analysis we select central jets (taggable) by requiring that they include at least two SVX tracks [26].

We require the trigger lepton to be contained in a cone of radius $R = 0.4$ around the axis of a taggable jet with uncorrected transverse energy $E_T \geq 15$ GeV. This jet will be referred to as a lepton-jet, l-jet, e-jet, or μ-jet. We also require the presence of at least one additional taggable jet (referred to as away-jet or a-jet) with $E_T \geq 15$ GeV. The request that both lepton- and away-jets are taggable restricts their pseudorapidity range to approximately $|\eta| \leq 1.5$. The requirement of a nonisolated lepton inside a jet rejects most of the leptonic decays of vector bosons and the Drell-Yan contribution. The request of two jets with $E_T \geq 15$ GeV reduces the statistics of the data sample.[4] This E_T threshold is chosen because efficiencies and backgrounds of the SECVTX, JPB, and SLT algorithms have been evaluated only for jets with transverse energy above this value [16]. We select 68 544 events with an e-jet and 14 966 events with a μ-jet.

In order to determine the bottom and charmed content of the data we use two algorithms (SECVTX and JPB) that have been studied in detail in Refs. [1,16]. SECVTX is based on the determination of the primary event vertex and the reconstruction of additional secondary vertices using displaced SVX tracks contained inside jets. JPB compares track impact parameters to measured resolution functions in order to calculate for each jet a probability that there are no long-lived particles in the jet cone [18].

The simulation of these tagging algorithms makes use of parametrizations of the detector response for single tracks, which were derived from the data. Because of the naivety of the method, these algorithms have required several empirical adjustments. SECVTX tags not produced by hadrons with heavy flavor (mistags) are underestimated by the detector simulation. Therefore SECVTX and JPB mistags are evaluated using a parametrized probability derived from generic-jet data [16], and are subtracted from the data in order to compare to the heavy flavor simulation. We estimate that the mistag removal has a 10% uncertainty [16].

The tagging efficiency of these algorithms is not well modeled by the parametrized simulation. In Ref. [16], we have used generic jets and a subset of the inclusive electron sample to determine the data-to-simulation scale factors for the tagging efficiency of these algorithms. The data-to-

[4] A jet with uncorrected transverse energy $E_T = 15$ GeV corresponds to a parton with average transverse energy $\langle E_T \rangle \simeq 20$ GeV.

simulation scale factor of the SECVTX tagging efficiency for b jets is measured to be 1.25 ± 0.08. The number of tags in the simulation is multiplied by this scale factor, and we add a 6% uncertainty to the prediction of tags. The data-to-simulation scale factor for c jets has been measured to be 0.92 ± 0.28 [16]; because of its large uncertainty, this scale factor is not implemented into the simulation, but we add a 28% uncertainty to the prediction of tags due to c jets. The data-to-simulation scale factor for the jet-probability algorithm has been measured to be 0.96 ± 0.05. The number of tags in the simulation is multiplied by this scale factor, and we add a 6% uncertainty to the prediction of tags.

In this study, we also probe the heavy-quark contribution by searching a jet for soft leptons (e and μ) produced by the decay of hadrons with heavy flavor. The soft lepton tagging algorithm is applied to sets of CTC tracks associated with jets with $E_T \geqslant 15$ GeV and $|\eta| \leqslant 2.0$. CTC tracks are associated with a jet if they are inside a cone of radius 0.4 centered around the jet axis. In order to maintain high efficiency, the lepton p_T threshold is set low at 2 GeV/c. To search for soft electrons the algorithm extrapolates each track to the calorimeter and attempts to match it to a CES cluster. The matched CES cluster is required to be consistent in shape and position with the expectation for electron showers. In addition, it is required that $0.7 \leqslant E/p \leqslant 1.5$ and $E_{had}/E_{em} \leqslant 0.1$. The track specific ionization (dE/dx), measured in the CTC, is required to be consistent with the electron hypothesis. The efficiency of the selection criteria has been determined using a sample of electrons produced by photon conversions [27].

To identify soft muons, track segments reconstructed in the CMU, CMP, and CMX systems are matched to CTC tracks. The CMU and CMX systems are used to identify muons with $2 \leqslant p_T \leqslant 3$ GeV/c and $p_T \geqslant 2$ GeV/c, respectively. Muon candidate tracks with $p_T \geqslant 3$ GeV/c within the CMU and CMP fiducial volume are required to match to track segments in both systems. The reconstruction efficiency has been measured using samples of muons from $J/\psi \rightarrow \mu^+\mu^-$ and $Z \rightarrow \mu^+\mu^-$ decays [27].

In the simulation, SLT tags are defined as tracks matching at generator level electrons or muons originating from b- or c-hadron decays (including those coming from τ or ψ cascade decays). The SLT tagging efficiency is implemented in the simulation by weighting these tracks with the efficiency of each SLT selection criteria measured using the data. The uncertainty of the SLT efficiency is estimated to be 10% and includes the uncertainty of the semileptonic branching ratios [27,28].

Rates of fake SLT tags are evaluated using a parametrized probability, P_f, derived in special samples of generic-jet data, and are subtracted from the data. This parametrization has been derived from the probability P that a track satisfying the fiducial requirements produces an SLT tag. This probability is computed separately for each lepton flavor and detector type and is parametrized as a function of the transverse momentum and isolation of the track [27,28]. In Ref. [16], by fitting the impact parameter distributions of the SLT tracks in the same generic-jet samples used to derive the P parametrization, we have estimated that $P_f = (0.740 \pm 0.074)P$. It follows that, in generic-jet data, the probability that a track corresponds to a lepton arising from heavy-flavor decays is $P_{HF} = (0.260 \pm 0.074)P$. Since we search a jet for SLT candidates in a cone of radius of 0.4 around its axis, the probabilities of finding a fake SLT tag in a jet is $P_f^{jet}(N) = \Sigma_{i=1}^{N}[1 - P_f^{jet}(i-1)]P_f^i$, where N is the number of tracks contained in the jet cone. In generic jets, the probability of finding an SLT tag due to heavy flavor is $P_{HF}^{jet}(N) = \Sigma_{i=1}^{N}[1 - P_{HF}^{jet}(i-1)]P_{HF}^i$. In Ref. [16], the uncertainty of the $P^{jet} = P_{HF}^{jet} + P_f^{jet}$ parametrization has been estimated to be no larger than 10% by comparing its prediction to the number of SLT tags observed in 7 additional generic-jet samples.

The efficiency for finding supertags (SLT tags in jets with SECVTX or JPB tags) in the simulation is additionally corrected with a data-to-simulation scale factor, 0.85 ± 0.05, derived in a previous study of generic-jet data [1]. The number of simulated supertags is multiplied by this factor, and we add a 6% uncertainty to the prediction of supertags. As mentioned earlier, the simulation of the SLT algorithm uses parametrized efficiencies measured using samples of electrons from photon conversions and muons from $J/\psi \rightarrow \mu^+\mu^-$ and $Z \rightarrow \mu^+\mu^-$ decays. Since these leptons are generally more isolated than leptons from heavy flavor decays, we have some evidence that the efficiency of the SLT algorithm in the simulation is overestimated. However, since a reduced efficiency for finding supertags could also be generated by a reduced efficiency of the SECVTX (JPB) algorithm in jets containing a soft lepton, we have chosen to correct the simulated efficiency for finding supertags, but not the efficiency of the simulated SLT algorithm [1].

V. DATA SAMPLE COMPOSITION

The lepton-jets in our sample come from three sources: $b\bar{b}$ production, $c\bar{c}$ production, and light quark or gluon production in which a hadron mimics the experimental signature of a lepton (fake lepton). The yield of fake leptons in light jets returned by our detector simulation cannot be trusted, and the $b\bar{b}$ and $c\bar{c}$ production cross sections have large experimental and theoretical uncertainties. Therefore, we use measured rates of lepton-jets with SECVTX and JPB tags due to heavy flavor (i.e. after mistag removal) in order to separate the fractions of lepton-jets due to $b\bar{b}$ production and $c\bar{c}$ production. The simultaneous use of the two tagging algorithms was pioneered in Ref. [16]; it allows us to separate the b- and c-quark contributions because both algorithms have the same tagging efficiency for b jets, while for c jets the efficiency of the JPB algorithm is approximately 2.5 times larger than that of the SECVTX algorithm. The b and c content of away-jets is also determined with this method.

The heavy flavor content of away-jets recoiling against a lepton-jet with heavy flavor depends on the production mechanisms (LO terms yield higher fractions of heavy flavor than NLO terms). Therefore, we tune the cross sections of the various production mechanisms predicted by the simula-

TABLE II. Number of tags due to heavy flavors in the inclusive lepton data (raw counts/removed mistags are indicated in parentheses). P_{GQCD} is the probability of tagging away-jets recoiling against lepton-jets without heavy flavor.

	Electron data			Muon data		
Tag type			P_{GQCD}			P_{GQCD}
$N_{l\text{-}jet}$	68544			14966		
$N_{a\text{-}jet}$	73335			16460		
$T^{SEC}_{l\text{-}jet}$	10115.3±101.7	(10221/105.7)		3657.3±60.8	(3689/31.7)	
$T^{JPB}_{l\text{-}jet}$	11165.4±115.8	(11591/425.6)		4068.6±66.2	(4204/135.4)	
$T^{SEC}_{a\text{-}jet}$	4353.3±68.5	(4494/140.7)	1.56%	1054.6±33.3	(1094/39.4)	1.67%
$T^{JPB}_{a\text{-}jet}$	5018.9±98.9	(5661/642.1)	2.45%	1265.2±41.1	(1427/161.8)	2.63%
DT^{SEC}	1375.2±37.6	(1405/29.8)		452.6±21.6	(465/12.4)	
DT^{JPB}	1627.8±43.7	(1754/126.2)		546.4±25.1	(600/53.6)	

tion to reproduce the observed number of lepton- and away-jets with SECVTX and JPB tags due to heavy flavor.

The fraction F_{HF} of lepton-jets due to heavy flavor, before tagging, is estimated using the tuned simulation. The remaining fraction, $(1-F_{HF})$, of lepton-jets is attributed to fake leptons in light jets. The number of tags in away-jets, which recoil against a lepton-jet without heavy flavor, is predicted as $N_{a\text{-}jet}(1-F_{HF})P_{GQCD}$, where $N_{a\text{-}jet}$ is the total number of away-jets, and P_{GQCD} is the average probability of tagging away-jets that recoil against lepton-jets without heavy flavor. The average probability P_{GQCD} is estimated by weighting all the away-jets with a parametrized probability of finding SECVTX (or JPB) tags due to heavy flavor in generic-jet data [16]. The number $N_{a\text{-}jet}(1-F_{HF})P_{GQCD}$ is subtracted from the number of tagged away-jets with heavy flavor that are used to tune the simulation. In Ref. [16], this method has been verified by using it also in a sample of data in which electrons are identified as coming from photon conversions. The heavy-flavor purity of *e*-jets due to photon conversions ($\simeq 8\%$) is depleted with respect to that of *e*-jets not due to conversions ($\simeq 50\%$). The study in Ref. [16] shows that the usage of the probability P_{GQCD} allows us to model the observed rate of tagged away-jets in both the electron and conversion samples within a 10% statistical uncertainty.

TABLE III. Number of jets before and after tagging in the inclusive lepton simulation (dir, f.exc and GSP indicate the direct production, flavor excitation and gluon splitting contributions). The row indicated as ''HF light'' lists separately the rates of away-jets with and without heavy flavors and highlights the properties of different production mechanisms. Data-to-simulation scale factors for the various tagging algorithms are not yet applied.

Tag type	*b* dir	*c* dir	*b* f.exc	*c* f.exc	*b* GSP	*c* GSP
			Electron simulation			
$HF_{l\text{-}jet}$	5671	947	10779	2786	5263	1690
$HF_{a\text{-}jet}$	5848	977	11280	2913	6025	1877
HF light	5407/441	899/78	1605/9675	367/2546	707/5318	145/1732
$HFT^{SEC}_{l\text{-}jet}$	1867	52	3624	194	1732	147
$HFT^{JPB}_{l\text{-}jet}$	2392	163	4531	602	2106	356
$HFT^{SEC}_{a\text{-}jet}$	2093	91	480	68	222	15
$HFT^{JPB}_{a\text{-}jet}$	2622	203	584	136	276	58
$HFDT^{SEC}$	678	5	157	4	78	1
$HFDT^{JPB}$	1083	43	303	25	168	18
			Muon simulation			
$HF_{l\text{-}jet}$	1285	298	2539	942	1455	747
$HF_{a\text{-}jet}$	1358	313	2705	994	1708	816
HF light	1206/152	278/35	422/2283	124/870	171/1537	48/768
$HFT^{SEC}_{l\text{-}jet}$	569	34	1131	83	652	92
$HFT^{JPB}_{l\text{-}jet}$	707	77	1386	229	830	202
$HFT^{SEC}_{a\text{-}jet}$	498	29	132	13	54	11
$HFT^{JPB}_{a\text{-}jet}$	627	62	173	34	60	21
$HFDT^{SEC}$	218	3	59	2	20	1
$HFDT^{JPB}$	347	12	105	7	50	6

Therefore we attribute a 10% uncertainty to the average probability P_{GQCD}.

Simulation of heavy flavor production and decay

We use the HERWIG Monte Carlo generator[5] to describe the fraction of data in which the lepton-jets contain hadrons with heavy flavor. We use the MRS(G) set of parton distribution functions [20], and set $m_c = 1.5\ \mathrm{GeV}/c^2$ and $m_b = 4.75\ \mathrm{GeV}/c^2$. In the generic hard parton scattering, $b\bar{b}$ and $c\bar{c}$ pairs are generated by HERWIG through processes of order α_s^2 such as $gg \to b\bar{b}$ (direct production). Processes of order α_s^3 are implemented in the generator through flavor excitation processes, such as $gb \to gb$, or gluon splitting, in which the process $gg \to gg$ is followed by $g \to b\bar{b}$. The HERWIG generator neglects virtual emission graphs, but, as all parton shower Monte Carlo generators, also includes higher than NLO diagrams.

The bottom and charmed hadrons produced in the final state are decayed using the CLEO Monte Carlo generator (QQ) [29]. At this generation level, we retain only final states that contain hadrons with heavy flavor and at least one lepton with $p_T \geq 8\ \mathrm{GeV}/c$. The accepted events are passed through a simulation of the CDF detector (QFL) that is based on parametrizations of the detector response derived from the data. After the simulation of the CDF detector, the Monte Carlo events are treated as real data. The simulated inclusive electron sample has 27 136 events, corresponding to a luminosity of $98.9\ \mathrm{pb}^{-1}$. The simulated inclusive muon sample has 7266 events, corresponding to a luminosity of $55.1\ \mathrm{pb}^{-1}$. The simulated samples have approximately the same luminosity as the data.

VI. DETERMINATION OF THE RATES OF SECVTX AND JPB TAGS DUE TO HEAVY FLAVOR IN THE DATA

The heavy flavor content of the data is estimated from the number of jets tagged with the SECVTX and JPB algorithms. The numbers of lepton-jets and away-jets in the data, $N_{l\text{-}jet}$ and $N_{a\text{-}jet}$, are listed in Table II. $N_{l\text{-}jet}$ is equal to the number of events and $N_{a\text{-}jet}$ is about 10% larger, which means that about 10% of the events have two away-jets. This table lists the following numbers of tags due to the presence of hadrons with heavy flavor:

(1) $T_{l\text{-}jet}^{SEC}$ and $T_{l\text{-}jet}^{JPB}$, the number of lepton-jets with a SECVTX and JPB tag, respectively.
(2) $T_{a\text{-}jet}^{SEC}$ and $T_{a\text{-}jet}^{JPB}$, the number of away-jets with a SECVTX and JPB tag, respectively.
(3) DT^{SEC} and DT^{JPB}, the number of events in which the lepton-jet and one away-jet are both tagged by SECVTX and JPB, respectively.

The uncertainty on the number of tags due to heavy flavor in Table II includes the 10% error of the mistag removal.

[5]We use option 1500 of version 5.6, generic $2 \to 2$ hard scattering with $p_T \geq 13\ \mathrm{GeV}/c$ (see Appendix A in Ref. [1] for more details).

TABLE IV. Result of the fit of the HERWIG simulation to the data. The fit is described in the text and yields $\chi^2/(\text{degree of freedom}) = 4.6/9$. The rescaling factors for the gluon splitting contributions predicted by the HERWIG parton-shower Monte Carlo program are of the same size as those measured by the SLC and LEP experiments [31], and are consistent with the estimated theoretical uncertainty [32].

SECVTX scale factor	SF_b	0.97 ± 0.03
SECVTX scale factor	SF_c	0.94 ± 0.22
JPB scale factor	SF_{JPB}	1.01 ± 0.02
e norm.	K_e	1.02 ± 0.05
μ norm.	K_μ	1.08 ± 0.06
c dir. prod.	c	1.01 ± 0.10
b flav. exc.	bf	1.02 ± 0.12
c flav. exc.	cf	1.10 ± 0.29
$g \to b\bar{b}$	bg	1.40 ± 0.18
$g \to c\bar{c}$	cg	1.40 ± 0.34

Events in which the lepton-jet does not contain heavy flavor are not described by the heavy flavor simulation. In these events, the number of away-jets with tags due to heavy flavor is predicted using the average tagging probabilities P_{GQCD} listed in Table II. These probabilities are used to correct the numbers of tagged away-jets that will be used to tune the heavy flavor simulation.

VII. TAGGING RATES IN THE SIMULATION

Numbers of tags in simulated events which contain heavy flavor, characterized by the prefix HF, are listed in Table III. Different production mechanisms are separated by inspecting at generator level the flavor of the initial and final state partons involved in the hard scattering. We attribute to flavor excitation the events in which at least one of the incoming partons has heavy flavor and to direct production the events in which the incoming partons have no heavy flavor and the outgoing partons both have heavy flavor. Pairs of heavy quarks that appear at the end of the evolution process are attributed to gluon splitting. The flavor type of each simulated jet is determined by inspecting its hadron composition at generator level.

VIII. TUNING OF THE SM SIMULATION USING SECVTX AND JPB TAGS

Following the procedure outlined in Sec. V, we fit the data with the heavy flavor simulation using rates of jets before and after tagging with the SECVTX and JPB algorithms. In the fit, we tune the cross sections of the different flavor production mechanisms. Starting from Table III the simulated rate of jets before tagging can be written as:

$$\mathrm{HF}_{l,i} = K_l \cdot (\mathrm{HF}_{b\ dir,l,i} + bf \cdot \mathrm{HF}_{bf.exc,l,i} + bg \cdot \mathrm{HF}_{bGSP,l,i}) + K_l \cdot (c \cdot \mathrm{HF}_{c\ dir,l,i} + cf \cdot \mathrm{HF}_{cf.exc,l,i} + cg \cdot \mathrm{HF}_{cGSP,l,i})$$

The rates of tagged jets are

TABLE V. Parameter correlation coefficients.

	SF_c	SF_{JPB}	K_e	c	bf	cf	bg	cg	K_μ
SF_b	−0.073	0.718	−0.747	0.054	0.346	0.297	−0.062	0.066	−0.715
SF_c		0.358	−0.238	−0.002	0.038	0.147	−0.071	0.086	−0.306
SF_{JPB}			−0.810	0.010	0.363	0.127	−0.009	−0.049	−0.802
K_e				−0.092	−0.641	−0.302	0.071	0.077	0.933
c					0.053	0.020	0.008	0.002	−0.098
bf						0.245	−0.680	−0.199	−0.526
cf							−0.321	−0.164	−0.274
bg								−0.029	−0.019
cg									−0.018

$$\mathrm{HFT}^j_{l,i} = K_l \cdot \mathrm{SF}^j_b (\mathrm{HFT}^j_{b\ dir,l,i} + bf \cdot \mathrm{HFT}^j_{bf.exc,l,i} + bg \cdot \mathrm{HFT}^j_{bGSP,l,i}) + K_l \cdot \mathrm{SF}^j_c (c \cdot \mathrm{HFT}^j_{c\ dir,l,i} + cf \cdot \mathrm{HFT}^j_{cf.ex,l,i} + cg \cdot \mathrm{HFT}^j_{cGSP,l,i})$$

and the rates of events with a double tag are

$$\mathrm{HFDT}^j_l = K_l \cdot \mathrm{SF}^{j\,2}_b (\mathrm{HFDT}^j_{b\ dir,l} + bf \cdot \mathrm{HFDT}^j_{bf.ex,l} + bg \cdot \mathrm{HFDT}^j_{bGSP,l}) + K_l \cdot \mathrm{SF}^{j\,2}_c (c \cdot \mathrm{HFDT}^j_{c\ dir,l} + cf \cdot \mathrm{HFDT}^j_{cf.ex,l} + cg \cdot \mathrm{HFDT}^j_{cGSP,l}),$$

where the index l indicates electron or muon data, i indicates the lepton- or the away-jet, and j indicates the type of tag (SECVTX or JPB). The fit parameters K_l account for the slightly different luminosity between data and simulation; they also include the normalization of the direct b-production cross section. The factors c, cf, cg, bf, and bg are fit parameters used to adjust the remaining cross sections calculated by HERWIG with respect to the direct $b\bar{b}$ production. The number of tags predicted by the simulation is obtained by multiplying the numbers in Table III by the appropriate scale factor. The fit parameters SF^j_b and SF^j_c are used to account for the uncertainties of the corresponding scale factors. The simulated rates $\mathrm{HFT}^j_{l,i}$ and HFDT^j_l have statistical errors $\delta^j_{T,l,i}$ and $\delta^j_{DT,l}$.

As mentioned at the end of Sec. V, the fraction of the data, which contains heavy flavor and is described by the simulation, is $F^l_{HF} = \mathrm{HF}_{l,l\text{-}jet}/N_{l\text{-}jet}$. Therefore we fit the simulated rates to the quantities

$$\mathrm{HFT}^j_{l,l\text{-}jet}(\mathrm{Data}) = T^j_{l,l\text{-}jet},$$

$$\mathrm{HFT}^j_{l,a\text{-}jet}(\mathrm{Data}) = T^j_{l,a\text{-}jet} - N_{l,a\text{-}jet}(1 - F^l_{HF}) P^j_{GQCD,l},$$

$$\mathrm{HFDT}^j_l(\mathrm{Data}) = \mathrm{DT}^j_l.$$

where $P^j_{GQCD,l}$ is the probability of finding a type-j tag due to heavy flavor in a-jets recoiling against a l-jet without heavy flavor (see Table II). The errors $\epsilon^j_{T,l,i}$ of the rates $\mathrm{HFT}^j_{l,i}(\mathrm{Data})$ include also the 10% uncertainty of $P^j_{GQCD,l,i}$.

Following the same procedure pioneered in Ref. [16], in which the HERWIG simulation was tuned to generic-jet data, we constrain the following fit parameters X_i to their measured or expected value $\overline{X_i}$ using the term

$$G_i = \frac{(X_i - \overline{X_i})^2}{\sigma^2_{\overline{X_i}}}.$$

(1) The ratio of the b and c direct production cross sections; it is constrained to the HERWIG default value with a 14% Gaussian error to account for the uncertainty of the parton fragmentation and for the fact that all quarks are treated as massless by the generator.

TABLE VI. Rates of tags due to heavy flavor in the data and in the fitted HERWIG simulation. The heavy flavor purity of the lepton-jets in the data returned by the best fit is $F_{HF} = (45.3 \pm 1.9)\%$ in the electron sample and $F_{HF} = (59.7 \pm 3.6)\%$ in the muon sample. The contribution of a-jets recoiling against l-jets without heavy flavor has been subtracted; the 10% uncertainty of this contribution is included in the errors.

	Electrons		Muons	
Tag type	Data	Simulation	Data	Simulation
$\mathrm{HFT}^{SEC}_{l\text{-}jet}$	10115.3 ± 101.7	10156.8 ± 159.3	3657.3 ± 60.8	3636.7 ± 95.8
$\mathrm{HFT}^{JPB}_{l\text{-}jet}$	11165.4 ± 115.8	11139.8 ± 159.7	4068.6 ± 66.2	4059.7 ± 95.8
$\mathrm{HFT}^{SEC}_{a\text{-}jet}$	3729.0 ± 92.8	3691.5 ± 109.7	943.8 ± 35.2	967.4 ± 43.2
$\mathrm{HFT}^{JPB}_{a\text{-}jet}$	4035.8 ± 139.7	3984.0 ± 111.0	1090.8 ± 44.9	1059.3 ± 42.8
HFDT^{SEC}	1375.2 ± 37.6	1380.8 ± 59.4	452.6 ± 21.6	474.3 ± 31.1
HFDT^{JPB}	1627.8 ± 43.7	1644.0 ± 57.1	546.4 ± 25.1	556.6 ± 28.7

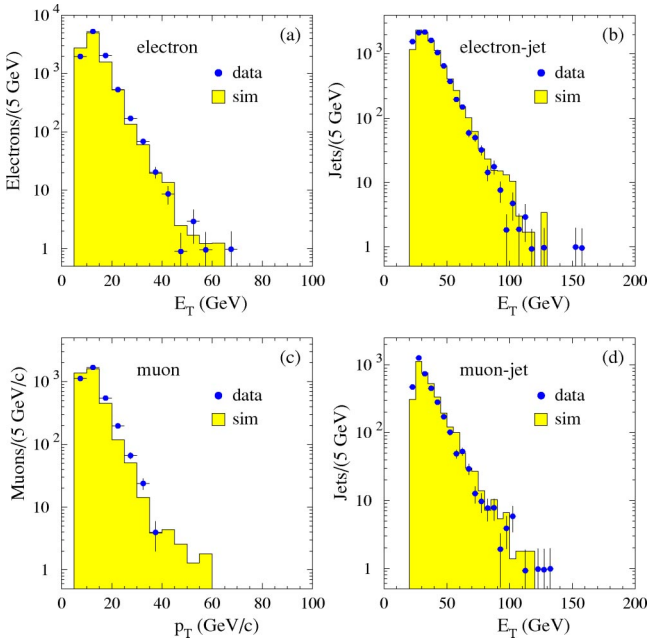

FIG. 1. Distributions of transverse energy, E_T, or momentum, p_T, for lepton-jets tagged by SECVTX. (a) electrons; (b) electron-jets; (c) muons; (d) muon-jets. Jet energies are corrected for detector effects and out-of-cone losses.

(2) The ratio of the b to c flavor excitation cross sections; it is constrained to the HERWIG default value with a 28% error to account for the uncertainty of the parton structure functions.

(3) The correction bg to the rate of gluon splitting; $g \to b\bar{b}$ is constrained to the value 1.4 ± 0.19 returned by the fit to generic-jet data [16].

(4) The correction cg to $g \to c\bar{c}$; it is constrained to the value 1.35 ± 0.36 returned by the fit to generic-jet data [16].

(5) We constrain SF_b for SECVTX to unity with a 6% error.

(6) We constrain SF_c to unity with a 28% error.

(7) We constrain SF_b^{JPB} and SF_c^{JPB} to unity with a 6% error.

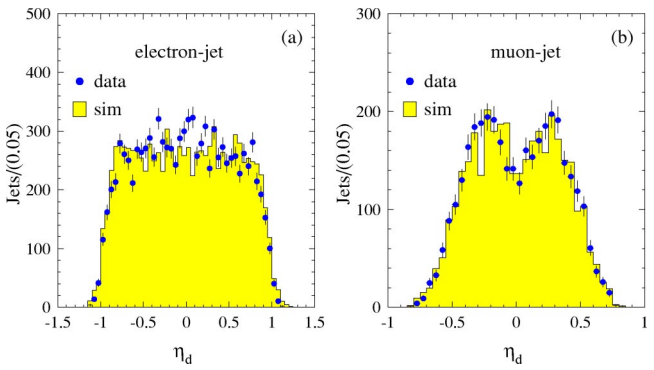

FIG. 2. Pseudorapidity distributions of (a) electron- and (b) muon-jets tagged by SECVTX.

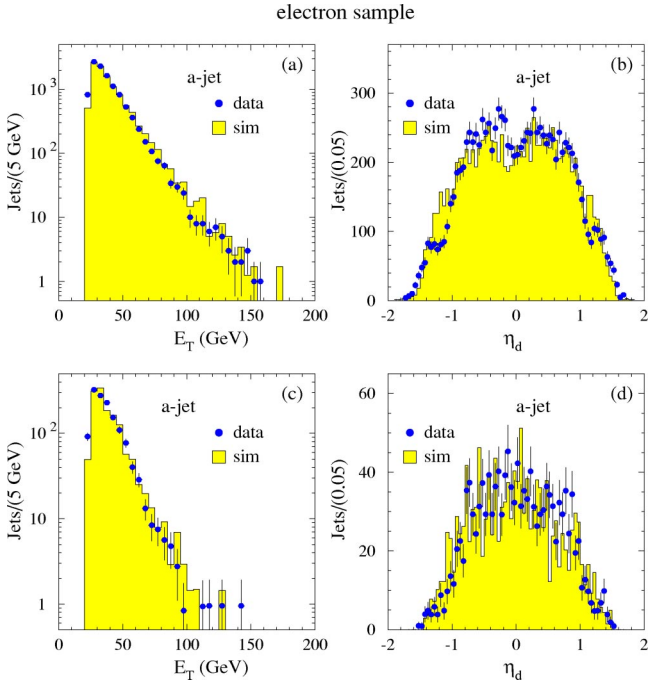

FIG. 3. Away-jet distributions in events where the electron-jet is tagged by SECVTX. (a) a-jet transverse energy; (b) a-jet pseudo-rapidity; (c) transverse energy of a-jets tagged by SECVTX; (d) pseudo-rapidity of a-jets tagged by SECVTX. Jet energies are corrected for detector effects and out-of-cone losses.

In summary the fit minimizes the function

$$\chi^2 = \sum_{l=e,\mu} \sum_{j=\text{tag type}} \left(\sum_{i=\text{jet type}} \frac{[\text{HFT}_{l,i}^j(\text{Data}) - \text{HFT}_{l,i}^j]^2}{\delta_{T,l,i}^{j\,2} + \epsilon_{T,l,i}^{j\,2}} + \frac{[\text{HFDT}_l^j(\text{Data}) - \text{HFDT}_l^j]^2}{\delta_{DT,l}^{j\,2} + \epsilon_{DT,l}^{j\,2}} \right) + \sum_{i=1}^{7} G_i \,.$$

In total we fit 12 rates with 10 free parameters and 7 constraints. The best fit returns a χ^2 value of 4.6 for 9 degrees of freedom. The values of the parameters returned by the fit and their correlation coefficients are shown in Tables IV and V. Tagging rates in the data and in the fitted simulation are listed in Table VI.

As shown by Table IV, the correction factors to the parton-level cross sections predicted by HERWIG are close to unity. As also noted in Ref. [30], the HERWIG generator predicts a single b cross section at the Tevatron that is approximately a factor of two larger than the NLO prediction of the MNR Monte Carlo generator [8] and is in fair agreement with the CDF and DØ measurements. As shown in Table III, LO (labeled as direct production) and higher-order (labeled as flavor excitation and gluon splitting) terms produce events with quite different kinematics. The LO contribution mostly consists of events which contain two jets with b (or c) flavor in the detector acceptance. In contrast, only a small fraction of the events due to higher-order terms contains two jets with heavy flavor in the detector acceptance. Therefore, the observed ratio of tagged a-jets to tagged l-jets constrains the

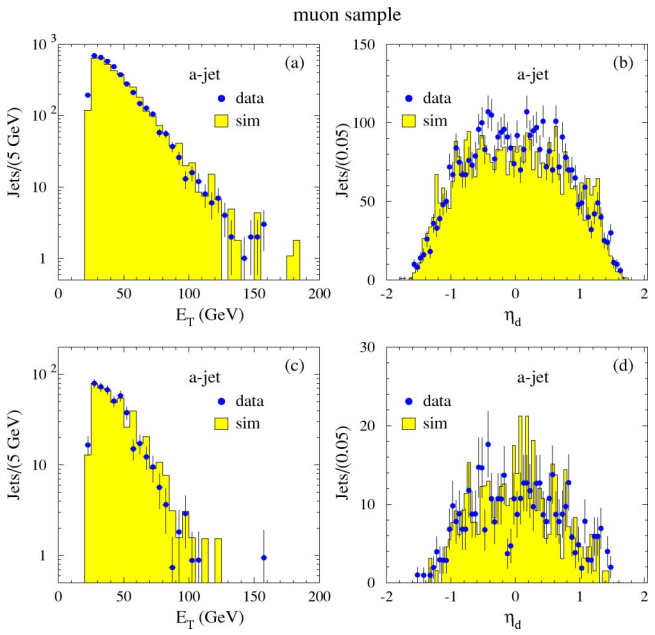

FIG. 4. Away-jet distributions in events where the muon-jet is tagged by SECVTX. (a) *a*-jet transverse energy; (b) *a*-jet pseudorapidity; (c) transverse energy of *a*-jets tagged by SECVTX; (d) pseudorapidity of *a*-jets tagged by SECVTX. Jet energies are corrected for detector effects and out-of-cone losses.

relative weight of LO and higher-order contributions. In the HERWIG simulation tuned to reproduce the data, the contribution of higher-order terms to the single b cross section is approximately four times larger than the LO contribution. For the same kinematics, the MNR generator, which uses normalization and factorization scales $\mu_0=(p_T{}_b{}^2+m_b^2)^{1/2}$ and the same parton distribution functions, predicts a ratio of NLO to LO contributions that is approximately two.

As shown by the comparison between data and tuned simulation in Table VI (rows 3 to 6), the number of events

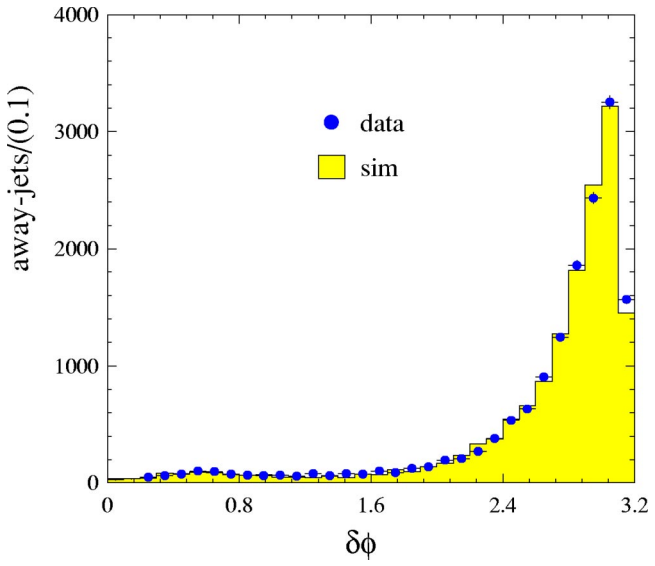

FIG. 5. Distribution of the azimuthal angle $\delta\phi$ between lepton-jets tagged by SECVTX and away-jets in the same event.

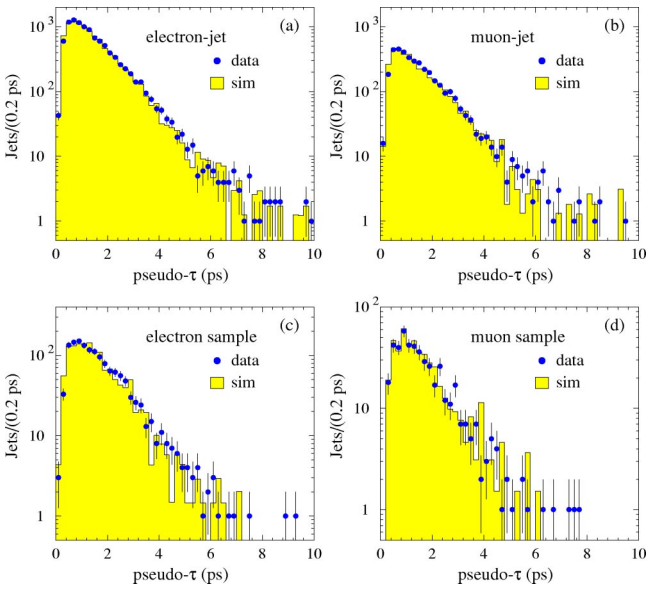

FIG. 6. Pseudo-τ distributions of (a) electron-jets and (b) muon-jets tagged by SECVTX and for tagged away-jets in events where the (c) electron-jet or the (d) muon-jet is also tagged.

containing two jets with heavy flavor, corresponding to $\sigma_{b\bar{b}}$, is well modeled by the HERWIG generator in which, as shown in Table III, approximately 30% of the production is due to higher-than-LO terms. In this case, the NLO prediction of $\sigma_{b\bar{b}}$ underestimates the data by 20%, whereas, as mentioned in the introduction, the NLO predictions of $\sigma_{b\bar{b}}\times$BR and $\sigma_{b\bar{b}}\times$BR^2 underestimate the data by a much larger factor.

Kinematics

Because of the large flavor excitation contribution, the cross section evaluated with HERWIG depends strongly on the pseudorapidity and transverse momentum of the heavy quarks in the final state. The $2\rightarrow2$ hard scattering with $p_T^{\min}\geqslant13$ GeV/c used to generate simulated events does not cover some of the available phase space, such as the production of massive gluons with small transverse momentum, which then branch into pairs of heavy quarks. In addition, the detector simulation (QFL), which is based upon parametrizations of single particle kinematics, may not accurately model the jet-E_T and trigger thresholds used in the analysis. It is therefore important to show that the simulation, which reproduces correctly the tagging rates and the away-jet multiplicity distribution, also models the event kinematics. Figures 1–4 compare transverse energy and pseudorapidity distributions in the data and in the simulation, normalized according to the fit listed in Table IV.[6]

[6]The systematic discrepancy in the first bin of each E_T distribution is the reflection of the slightly inaccurate modeling of the efficiency of the lepton trigger near the threshold. A few local discrepancies in some pseudorapidity distributions at $|\eta|\simeq0$ and $|\eta|\simeq1$ are due to an inaccurate modeling of the calorimetry cracks. These small discrepancies are not relevant in this analysis.

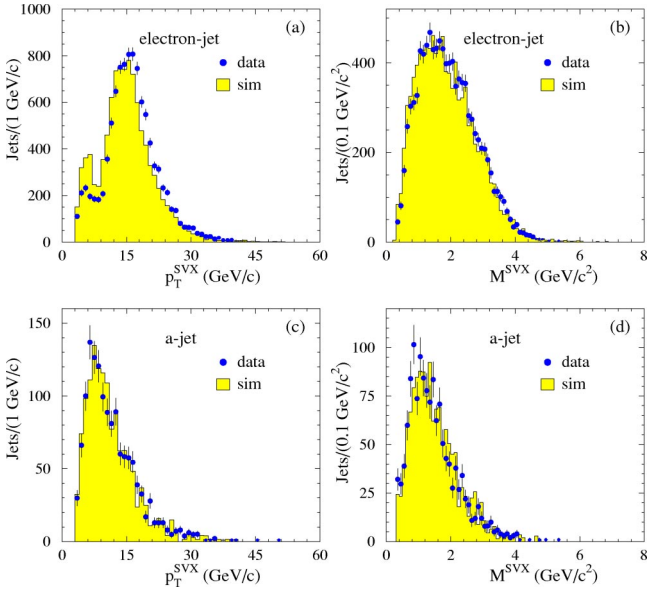

FIG. 7. Distributions of the (a) transverse momentum and (b) invariant mass of SECVTX tags in electron-jets; (c) and (d) are analogous distributions for away-jets in events in which the *e*-jet is also tagged.

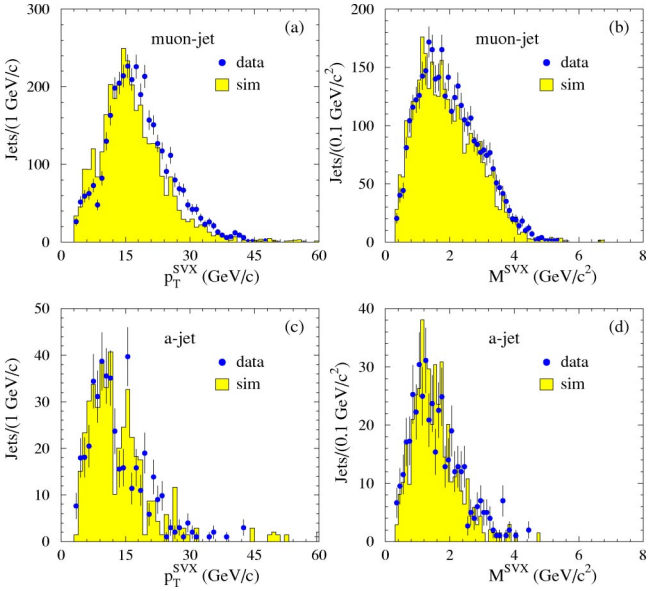

FIG. 8. Distributions of the (a) transverse momentum and (b) invariant mass of SECVTX tags in muon-jets; (c) and (d) are analogous distributions for away-jets in events in which the muon-jet is also tagged.

Figure 5 compares distributions of the azimuthal angle $\delta\phi$ between the lepton-jet and the away-jets. The region at $\delta\phi$ smaller than 1.2, which is well modeled by the tuned simulation, is mostly populated by the gluon splitting contribution. The good agreement between data and prediction supports the 40% increase of the gluon splitting cross sections (see Table IV).

Figure 6 compares pseudolifetime distributions of SECVTX tags. The pseudolifetime is defined as

$$\text{pseudo-}\tau = \frac{L_{xy} M^{SVX}}{c p_T^{SVX}},$$

where L_{xy} is the projection of the two-dimensional vector pointing from the primary vertex to the secondary vertex on the jet direction, and M^{SVX} and p_T^{SVX} are the invariant mass and the transverse momentum of all tracks forming the SECVTX tag.

Distributions of M^{SVX} and p_T^{SVX}, which is sensitive to the heavy-quark fragmentation, are shown in Figs. 7 and 8. In Figs. 7(a) and 8(a), the simulated p_T^{SVX} distributions of SECVTX tags in lepton-jets are above the data near the p_T threshold. This discrepancy follows from the fact that the tagging efficiency in the simulation is smaller than in the data and we take care of it with an overall multiplicative factor. This procedure does not account for the fact that the probability that a 8 GeV/c lepton is part of a tag is also higher in the data than in the simulation. In away-jets, where high-p_T tracks are not a selection prerequisite, there is better agreement between data and simulation. In conclusion, our simulation calibrated within the theoretical and experimental uncertainties models correctly the heavy flavor production at the Tevatron.

IX. RATES OF SLT TAGS

Following the strategy outlined in Sec. II, we search away-jets for soft leptons (e or μ) with $p_T \geqslant 2$ GeV/c and contained in a cone of radius 0.4 around the jet axis. We then compare rates of away-jets containing soft lepton tags due to heavy flavor in the data and in the simulation tuned as in Table IV. Table VII lists the following rates of away-jets with SLT tags:

TABLE VII. Number of away-jets with SLT tags due to heavy flavors in the inclusive lepton sample. Raw counts and removed mistags are listed in parentheses. When appropriate, mistags include fake SECVTX (JPB) contributions. P_{GQCD} is the probability of finding a tag due to heavy flavor in away-jets recoiling against a lepton-jet without heavy flavor.

	Electron data			Muon data		
Tag type			P_{GQCD}			P_{GQCD}
$T_{a\text{-}jet}^{SLT}$	1063.8±113.0	(2097/1033.2)	0.49%	308.6±34.7	(562/253.4)	0.54%
$T_{a\text{-}jet}^{SLT\cdot SEC}$	356.3±22.8	(444/87.7)	0.08%	69.3±9.9	(92/22.7)	0.09%
$T_{a\text{-}jet}^{SLT\cdot JPB}$	401.3±25.3	(513/111.7)	0.13%	112.3±12.3	(143/30.7)	0.14%

TABLE VIII. Rates of away-jets with SLT tag due to heavy flavors in the inclusive lepton simulation. The data-to-simulation scale factor for the supertag efficiency is not yet applied.

Tag type	b dir	c dir	b f.exc	c f.exc	b GSP	c GSP
			Electron simulation			
$\mathrm{HFT}^{SLT}_{a\text{-}jet}$	362	26	93	30	41	9
$\mathrm{HFT}^{SLT\cdot SEC}_{a\text{-}jet}$	159	1	47	2	18	0
$\mathrm{HFT}^{SLT\cdot JPB}_{a\text{-}jet}$	200	7	53	6	21	2
			Muon simulation			
$\mathrm{HFT}^{SLT}_{a\text{-}jet}$	82	10	21	5	9	5
$\mathrm{HFT}^{SLT\cdot SEC}_{a\text{-}jet}$	33	2	9	0	4	0
$\mathrm{HFT}^{SLT\cdot JPB}_{a\text{-}jet}$	44	3	13	3	5	2

(1) $T^{SLT}_{a\text{-}jet}$, the number of away-jets with a soft lepton tag.

(2) $T^{SLT\cdot SEC}_{a\text{-}jet}$ ($T^{SLT\cdot JPB}_{a\text{-}jet}$), the number of away-jets with an SLT tag and a SECVTX (JPB) tag (called supertag in Ref. [1]).

The uncertainty on the number of tags due to heavy flavor in Table VII includes the 10% error of the mistag removal. In events in which the lepton-jet does not contain heavy flavor, the number of away-jets with an SLT tag due to heavy flavor is predicted using the average probability P_{GQCD}. This average probability is estimated by weighting all the away-jets with the parametrized probability P^{jet}_{hf}, derived in generic-jet data and described in Sec. IV. In these events, the uncertainty of the average probability of finding a real or a fake SLT tags is estimated to be no larger than 10%. We cross-check the estimate of these uncertainties in Sec. X.

Rates of SLT tags in the simulation before tuning are shown in Table VIII. The uncertainty of the SLT efficiency is estimated to be 10% and includes the uncertainty of the semileptonic branching ratios [27,28]. The numbers of supertags predicted by the simulation are obtained by multiplying the numbers in Table VIII by the scale factor 0.85 ± 0.05.

Following the notations of Sec. VIII, rates of tagged away-jets with heavy flavor in the fitted simulation are defined as

$$\begin{aligned}\mathrm{HFT}^{SLT}_{l,a\text{-}jet} = K_l \cdot (&\mathrm{HFT}^{SLT}_{b\ dir,l,a\text{-}jet} + bf \cdot \mathrm{HFT}^{SLT}_{bf.exc,l,a\text{-}jet} \\ &+ bg \cdot \mathrm{HFT}^{SLT}_{bGSP,l,a\text{-}jet}) \\ &+ K_l \cdot (c \cdot \mathrm{HFT}^{SLT}_{c\ dir,l,a\text{-}jet} \\ &+ cf \cdot \mathrm{HFT}^{SLT}_{cf.exc,l,a\text{-}jet} + cg \cdot \mathrm{HFT}^{SLT}_{cGSP,l,a\text{-}jet}),\end{aligned}$$

$$\begin{aligned}\mathrm{HFT}^{SLT\cdot j}_{l,a\text{-}jet} = K_l \cdot \mathrm{SF}^j_b (&\mathrm{HFT}^{SLT\cdot j}_{b\ dir,l,a\text{-}jet} \\ &+ bf \cdot \mathrm{HFT}^{SLT\cdot j}_{bf.exc,l,a\text{-}jet} \\ &+ bg \cdot \mathrm{HFT}^{SLT\cdot j}_{bGSP,l,a\text{-}jet}) \\ &+ K_l \cdot \mathrm{SF}^j_c (c \cdot \mathrm{HFT}^{SLT\cdot j}_{c\ dir,l,a\text{-}jet} \\ &+ cf \cdot \mathrm{HFT}^{SLT\cdot j}_{cf.exc,l,a\text{-}jet} \\ &+ cg \cdot \mathrm{HFT}^{SLT\cdot j}_{cGSP,l,a\text{-}jet}),\end{aligned}$$

where $\mathrm{HFT}^{SLT}_{l,a\text{-}jet}$ is the rate of a-jets containing heavy flavor tagged by the SLT algorithm, and $\mathrm{HFT}^{SLT\cdot j}_{l,a\text{-}jet}$ is the rate of a-jets containing heavy flavor with a supertag j (SECVTX or JPB). The errors on the simulated rates include the statistical error, the systematic uncertainty for finding SLT tags and supertags, and the uncertainties of the parameters (K_l, bf, bg, c, cf, cg, and SF) listed in Tables IV and VI. In the data the analogous rates are

$$\mathrm{HFT}^{SLT}_{l,a\text{-}jet}(\mathrm{Data}) = T^{SLT}_{l,a\text{-}jet} - N_{l,a\text{-}jet}(1 - F^l_{HF}) P^{SLT}_{GQCD,l},$$

$$\mathrm{HFT}^{SLT\cdot j}_{l,a\text{-}jet}(\mathrm{Data}) = T^{SLT\cdot j}_{l,a\text{-}jet} - N_{l,a\text{-}jet}(1 - F^l_{HF}) P^{SLT\cdot j}_{GQCD,l}.$$

Rates of soft leptons due to heavy flavor in the data and in the tuned simulation

The comparison of the yields of away-jets with SLT tags due to heavy flavor in the data and in the tuned simulation is shown in Table IX. Table X lists the numbers of tags in the tuned simulation split by flavor type and production mechanism, and Table XI summarizes the different contributions to the observed number of tags. In the data there are $\mathrm{HF}^{SLT}_{a\text{-}jet} = 1138 \pm 140$ a-jets with a soft lepton tag due to heavy flavor. The ± 140 error is dominated by the 10% systematic uncertainty of the fake and generic QCD contributions to SLT tags; the statistical error is ± 51 jets. The simulation predicts 747 ± 75 a-jets with soft lepton tags due to $b\bar{b}$ and $c\bar{c}$ production (most of the error is systematic and due to the 10% uncertainty on the SLT tagging efficiency). The discrepancy is a 2.5σ systematic effect.

The comparison of the yields of supertags in the data and in the tuned simulation is also listed in Table XI. The subset of data, in which a-jets have both SLT and JPB tags due to heavy flavor, contains 453 ± 29 supertags (in this case the ± 25 statistical error is larger than the ± 15 systematic error due to the fake-tag subtraction). The simulation predicts 317 ± 25 a-jets with a supertag due to $b\bar{b}$ and $c\bar{c}$ production. The ± 25 systematic error is obtained combining in quadrature the uncertainty of the SLT efficiency (± 16) with the uncertainty (± 20) due to the fit in Table IV and to the simulation statistical error. This discrepancy is a 3.5σ effect

TABLE IX. Number of a-jets with an SLT tag due to heavy flavor decay. The contribution of a-jets recoiling against l-jets without heavy flavor has been subtracted (see text).

	Electrons		Muons	
Tag type	Data	Simulation	Data	Simulation
$\mathrm{HFT}^{SLT}_{a\text{-}jet}$	865.1 ± 114.8	597.6 ± 69.3	272.7 ± 34.9	149.3 ± 21.0
$\mathrm{HFT}^{SLT\cdot SEC}_{a\text{-}jet}$	322.6 ± 23.3	242.4 ± 22.5	63.3 ± 9.9	53.8 ± 8.7
$\mathrm{HFT}^{SLT\cdot JPB}_{a\text{-}jet}$	350.2 ± 26.3	251.5 ± 21.7	103.2 ± 12.4	65.0 ± 8.9

TABLE X. Tagging rates in the normalized simulation listed by production mechanisms.

Tag type	*b* dir	*c* dir	*b* f.exc	*c* f.exc	*b* GSP	*c* GSP
			Electron simulation			
$HF_{l\text{-}jet}$	5781.0±320.8	973.2±109.8	11247.8±1073.9	3115.7±790.1	7504.6±1081.6	2411.0±593.8
$HF_{a\text{-}jet}$	5961.4±330.6	1004.0±113.2	11770.6±1123.6	3257.7±826.0	8591.1±1237.4	2677.8±659.2
$HFT^{SEC}_{l\text{-}jet}$	2267.5±101.6	49.1±19.4	4505.5±451.7	199.5±81.7	2942.4±408.7	192.8±87.8
$HFT^{JPB}_{l\text{-}jet}$	2358.3±99.0	162.0±20.7	4572.8±454.2	651.1±167.3	2904.3±404.4	491.2±122.2
$HFT^{SEC}_{a\text{-}jet}$	2542.0±112.3	86.0±33.1	596.8±65.0	69.9±29.4	377.1±57.5	19.7±10.2
$HFT^{JPB}_{a\text{-}jet}$	2585.1±107.3	201.8±24.8	589.4±62.8	147.1±39.4	380.6±57.1	80.0±22.1
$HFDT^{SEC}$	981.1±52.5	4.3±3.6	232.5±31.4	3.8±3.3	157.9±27.8	1.2±1.5
$HFDT^{JPB}$	1032.7±45.8	41.3±7.5	295.7±36.0	26.2±8.5	224.1±35.0	24.0±8.1
$HFT^{SLT}_{a\text{-}jet}$	369.0±46.2	26.7±6.6	97.0±16.7	33.6±11.0	58.5±13.7	12.8±5.5
$HFT^{SLT\cdot SEC}_{a\text{-}jet}$	164.1±17.4	0.8±0.9	49.7±9.2	1.7±1.4	26.0±7.2	0
$HFT^{SLT\cdot JPB}_{a\text{-}jet}$	167.6±16.6	5.9±2.3	45.5±8.1	5.5±2.7	24.6±6.5	2.3±1.8
			Muon simulation			
$HF_{l\text{-}jet}$	1383.7±84.4	323.5±39.6	2798.6±292.4	1112.8±285.4	2191.5±310.5	1125.7±284.9
$HF_{a\text{-}jet}$	1462.3±88.7	339.8±41.4	2981.5±311.2	1174.2±301.0	2572.5±363.5	1229.7±310.9
$HFT^{SEC}_{l\text{-}jet}$	730.0±42.3	33.9±14.0	1485.2±164.3	90.1±38.0	1170.0±161.8	127.5±59.3
$HFT^{JPB}_{l\text{-}jet}$	736.3±39.0	80.8±12.3	1477.5±160.9	261.6±69.0	1209.1±166.3	294.4±76.0
$HFT^{SEC}_{a\text{-}jet}$	638.9±38.4	28.9±12.1	173.3±23.8	14.1±7.0	96.9±18.4	15.2±8.3
$HFT^{JPB}_{a\text{-}jet}$	653.0±35.6	65.1±10.5	184.4±24.0	38.8±11.9	87.4±16.2	30.6±10.1
$HFDT^{SEC}$	333.2±26.2	2.8±2.5	92.3±16.1	2.0±2.0	42.8±11.1	1.3±1.6
$HFDT^{JPB}$	349.5±22.0	12.2±3.7	108.3±16.3	7.7±3.5	70.4±13.6	8.5±4.0
$HFT^{SLT}_{a\text{-}jet}$	88.3±14.0	10.9±3.8	23.1±6.0	5.9±3.1	13.6±5.1	7.5±3.9
$HFT^{SLT\cdot SEC}_{a\text{-}jet}$	36.0±6.8	1.7±1.4	10.0±3.6	0	6.1±3.2	0
$HFT^{SLT\cdot JPB}_{a\text{-}jet}$	38.9±6.5	2.7±1.6	11.8±3.6	2.9±1.8	6.2±2.9	2.5±1.9

dominated by systematic uncertainties. In the even smaller subset of events, in which *a*-jets contain both SECVTX and SLT tags due to heavy flavor, the discrepancy between data and simulation is a 2.4σ effect, also dominated by the same systematic errors.

There is no gain in combining the three results because the uncertainties on the number of *a*-jets with SLT tags due to heavy flavor, before and after tagging with the SECVTX and JPB algorithms, are highly correlated. Away-jets with supertags are a subset of the *a*-jets with SLT tags, and there is overlap between the subsets with JPB and SECVTX supertags. However, it is important to note that the discrepancy between observed and expected number of SLT tags is of the same size before and after tagging with the SECVTX and JPB algorithms. This disfavors the possibility that the disagreement between data and simulation arises from jets containing hadrons with a lifetime much shorter than that of conventional heavy flavor.

We have considered the impact on the number of expected supertags due to the 0.85 ± 0.05 scale factor derived in generic-jet data. If we had evaluated the number of simulated supertags using the product of simulated efficiencies of the SECVTX (JPB) algorithm and of the SLT algorithm, which has a 10% uncertainty, the discrepancy between data and simu-

TABLE XI. Summary of the observed and predicted numbers of *a*-jets with SLT tags or supertags in the inclusive lepton sample. Mistags are the expected fake-tag contributions in *a*-jets recoiling against *l*-jets with heavy flavor (HF). QCD are the predicted numbers of tags, which include mistags, in *a*-jets recoiling against *l*-jets without heavy flavor. $HFT_{a\text{-}jet}$ (data and HF simulation) are the numbers of tagged *a*-jets with heavy flavor recoiling against *l*-jets with heavy flavor; in the data, this contribution is obtained by subtracting the second plus third rows of this table from the first one.

Tag type	SLT	SLT+SECVTX	SLT+JPB
Observed	2659	536	656
Mistag	619±62	53±5	69±7
QCD	902±91	97±10	134±13
$HFT_{a\text{-}jet}$ (data)	1138±140	386±26	453±29
$HFT_{a\text{-}jet}$ (HF simulation)	747±75	296±26	317±25
Excess	391±159	90±37	136±38

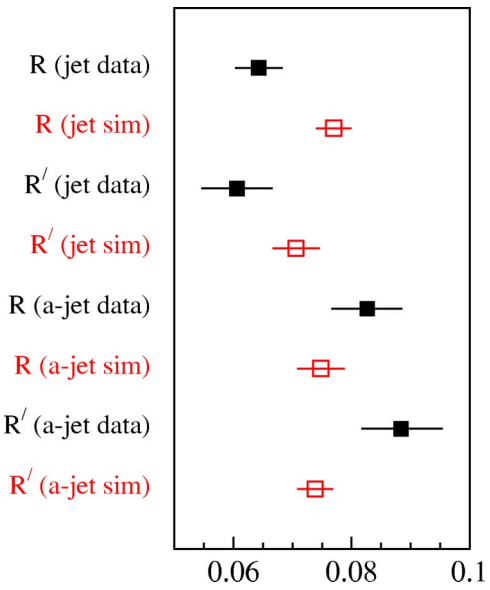

FIG. 9. Yield of R, the ratio of the number of jets with a SECVTX and SLT tag to that with a SECVTX tag in the data (square) and the corresponding simulations (open square). R' is the analogous ratio for JPB tags. The error in the simulation comes from the uncertainty of the relative ratio of bottom and charmed hadron in the data; this uncertainty results from the tuning of the heavy flavor cross sections predicted by HERWIG to model the rates of SECVTX and JPB tags observed in the data. The simulation is not corrected for the scale factor 0.85 ± 0.05, which is used to equalize data and prediction in generic jets.

lation would be smaller: 1.6σ and 1.0σ for a-jets with JPB and SECVTX tags, respectively. However, analogous rates of tags in generic-jet data would be approximately 1.5σ lower than in the simulation. Figure 9 shows the yield of R, the ratio of the number of supertags (SECVTX+SLT) to that of SECVTX tags produced by heavy flavor, in generic jets and in the away-jets recoiling against a lepton-jet. The ratio R' is derived in analogy replacing SECVTX with JPB tags. The comparison of these ratios in the generic-jet data and their simulation has been used in Ref. [1] to calibrate the efficiency for finding supertags in the simulation. In Fig. 9, the efficiency for finding supertags in the simulation has not been corrected with the 0.85 ± 0.05 scale factor. For the simulation, the plotted errors of $R(R')$ account for the uncertainty of the relative contribution of b and c quarks, but not for the uncertainty of the supertag efficiency, which is no smaller than 10%. One notes that the simulation predicts the same value of $R(R')$ for generic jets and away-jets in lepton-triggered events, whereas, in the data, the value of $R(R')$ for away-jets is approximately 20% higher than for generic jets.

Finally, we have investigated the dependence of the predicted yield of away-jets with SLT tags on the ratio of the $c\bar{c}$ to $b\bar{b}$ productions predicted by the simulation. To a good approximation, the predicted yield does not depend on the tuning of the simulation. Since the ratio of the tagging efficiency for c jets to that for b jets is approximately equal for the JPB and SLT algorithms,[7] the expected number of away-jets with SLT tags is

$$\begin{aligned} \mathrm{HFT}^{SLT}_{a\text{-}jet} &= \epsilon_b^{SLT} \times (N_b + \epsilon_c^{SLT}/\epsilon_b^{SLT} \times N_c) \\ &= \epsilon_b^{SLT}/\epsilon_b^{JPB} \times \epsilon_b^{JPB} \times (N_b + \epsilon_c^{JPB}/\epsilon_b^{JPB} \times N_c) \\ &= \epsilon_b^{SLT}/\epsilon_b^{JPB} \times \mathrm{HFT}^{JPB}_{a\text{-}jet}(\mathrm{data}) \\ &= \epsilon_b^{SLT}/\epsilon_b^{JPB} \times (5126.6 \pm 146.7) \\ &= 763 \pm 80 \end{aligned}$$

and does not depend on the size of N_b and N_c, the numbers of away-jets attributed by the fit to bottom and charmed flavor, respectively. As an example of this, without constraining the ratio of the c to b direct productions to the nominal value within a 14% error, we have misled the fit to return a very different, and not correct, local minimum ($c = 2.8 \pm 1.6$ instead of $c = 1.01 \pm 0.10$ in Table IV). The number of a-jets with SLT tags remains approximately constant (in the electron sample, 598 ± 69 becomes 603 ± 66; in the muon sample, 149 ± 21 becomes 156 ± 21).

X. SYSTEMATICS

This section reviews and verifies systematic effects that could reduce the discrepancy between observed and predicted numbers of away-jets with a soft lepton tag due to heavy flavor. The discrepancy depends on the estimate of the mistag rate in the data and on the simulated efficiency of the SLT algorithm, and also on the size of the $b\bar{b}$ contribution in the simulation. We verify these estimates in Secs. X A and X B, respectively. In Sec. X C, we verify the discrepancy between data and simulation found in this study with a sample of jets that recoil against J/ψ mesons arising from B decays.

A. Fake SLT tags and the simulated SLT efficiency

Table XI shows an excess of 391 away-jets with SLT tags due to heavy flavor with respect to the number, 747 ± 75, predicted by the heavy flavor simulation. In the data, we have removed a fake contribution of 619 ± 62 SLT tags.[8] If the estimate of the fake rate could be increased by 60% (6 times the estimated uncertainty), this excess would disappear. The simulated efficiency of the SLT algorithm has been tuned using the data and we estimate its uncertainty to be 10%; however, if the simulated efficiency could be increased by 50%, the disagreement between data and simulation would also disappear.

[7]The average tagging efficiencies in this data set are $\epsilon_b^{JPB} = 0.43$, $\epsilon_c^{JPB} = 0.30$, $\epsilon_b^{SLT} = 0.064$, and $\epsilon_c^{SLT} = 0.046$.

[8]In the data, we have also subtracted the generic-jet contribution of SLT tags due to a-jets recoiling against l-jets without heavy flavor (see Table XI). This contribution is slightly overestimated because the tagging probability P^{jet} has been constructed using also events in which both jets contain heavy flavor.

TABLE XII. Number of tags due to heavy flavors in three samples of generic jets [33] and in their tuned simulation. The amount of mistags removed from the data is indicated in parentheses; errors include a 10% uncertainty in the mistag evaluation. The yields of tags in the simulation have been corrected with the appropriate scale factors (see Sec. IV). The error of the number of simulated SLT tags includes the 10% uncertainty of the SLT tagging efficiency in the simulation; the simulation efficiency for finding supertags (SLT+SECVTX and SLT+JPB) has been empirically reduced by 15% to reproduce generic-jet data with a 6% accuracy.

Tag type	Data (removed fakes)	Simulation
	JET 20 (194,009 events)	
SECVTX	4058 ± 92 (616.0)	4052 ± 143
JPB	5542 ± 295 (2801.0)	5573 ± 173
SLT	1032 ± 402 (3962.0)	826 ± 122
SLT+SECVTX	219.8 ± 20 (94.2)	223 ± 16
SLT+JPB	287.3 ± 28 (166.7)	280 ± 19
	JET 50 (151,270 events)	
SECVTX	5176 ± 158 (1360.0)	5314 ± 142
JPB	6833 ± 482 (4700.0)	6740 ± 171
SLT	1167 ± 530 (5241.0)	1116 ± 111
SLT+SECVTX	347 ± 29 (169.0)	343 ± 23
SLT+JPB	427.5 ± 42 (288.5)	416 ± 27
	JET 100 (129,434 events)	
SECVTX	5455 ± 239 (2227.0)	5889 ± 176
JPB	6871 ± 659 (6494.0)	7263 ± 202
SLT	1116 ± 642 (6367.0)	1160 ± 168
SLT+SECVTX	377.6 ± 36 (243.4)	432 ± 29
SLT+JPB	451.8 ± 55 (401.2)	478 ± 32

Table XI also shows an excess of 137 a-jets with SLT+JPB supertags due to heavy flavor with respect to the number 316 ± 25 predicted by the simulation. In the data, we have removed 142 ± 14 fake tags; in this case, one would need to increase the fake-rate estimate by 10σ in order to cancel the excess in the data. The simulated supertag efficiency has been calibrated with generic-jet data to a 6% accuracy; in order to cancel the discrepancy, the supertag efficiency in the simulation should be increased by 8.7σ.

We verify the uncertainty of the fake rate and heavy flavor contributions by comparing rates of SLT tags in three generic-jet samples to their corresponding simulations fitted to the data using rates of SECVTX and JPB tags. These rates of tags, together with the fake contributions evaluated with the same fake parametrizations used in the present study, are listed in Table XII, which is derived from the study presented in Ref. [1]. A summary of Table XII is presented in Table XIII. The observed number of SLT tags in generic jets (sample A in Table XIII) is dominated by the fake contribution, and we use the difference between the observed number of SLT tags and the number of SLT tags due to heavy flavor predicted by the simulation to reduce the uncertainty of the fake rate. Generic-jet data contain 18885 SLT tags. The parametrized probability predicts 15570 ± 1557 fake tags. The simulation predicts 3102 SLT tags due to heavy flavor with a 13% uncertainty (dominated by the 10% uncertainty of the SLT tag-

TABLE XIII. Number of SLT tags in all generic jets listed in Table XII (sample A) and in away-jets recoiling a lepton-jet with heavy flavor (sample D). Samples B and C are generic jets tagged with the SECVTX and JPB algorithms, respectively. Before tagging with the SLT algorithm, the heavy flavor purity is 13% for sample A, 78% for sample B, 58% for sample C, and 26% for the sample D used in this study. The prediction of the fake SLT rate is calculated with the same parametrized probability for all samples; the heavy flavor (HF) contributions are predicted with the same simulation.

Sample	Number of SLT tags	Predicted fakes	Predicted HF
A: JET 20+JET 50+JET 100	18885	15570 ± 1557	3102 ± 403
B: generic jets with SECVTX tags	1451	507 ± 51	998 ± 60
C: generic jets with JPB tags	2023	856 ± 86	1174 ± 71
D: away-jets	1757	619 ± 62	747 ± 75

ging efficiency). By removing from the data the heavy flavor contribution predicted by the simulation, one derives an independent and consistent estimate for the fake contribution of 15783 ± 403 SLT tags. The latter determination of the fake contribution has a 2.6% uncertainty.

Before tagging with the SLT algorithm, away-jets in the inclusive lepton sample have a larger heavy flavor content ($\simeq 26\%$) than that of sample A in Table XIII ($\simeq 13\%$). However, generic jets tagged by SECVTX and JPB algorithms (samples B and C, respectively) have a heavy-flavor purity of 78% and 58%, respectively. Because these latter samples have a larger heavy flavor content, the discrepancy between the observed and predicted yields of away-jets with SLT tags observed in the present study cannot arise from deficiencies of the heavy flavor simulation or from an increase of the fake probability in jets with heavy favor.

In addition, the total number of SLT tags observed in generic jets can be used to achieve a better determination of the sum of the predicted numbers of fake SLT tags plus SLT tags due to heavy flavor (HF) with respect to that presented in Sec. IX A. To obtain this, we fit the observed rate of SLT tags in both samples A and C with the predicted number of fake and HF tags weighted with unknown parameters P_f and P_{HF}, respectively. The data constrain the parameter values to be $P_f = 1.017 \pm 0.013$ and $P_{HF} = 0.981 \pm 0.045$ with a correlation coefficient $\rho = -0.77$.

After having removed the contribution of events in which the lepton-jet does not contain heavy flavor, away-jets contain 1757 ± 104 SLT tags; in Sec. IX A, this number was compared to a prediction of 619 ± 62 fake and 747 ± 75 HF tags. When using the weights, errors, and parameter correlation derived using generic jets, the prediction of the total number of SLT tags becomes 1362 ± 28. The systematic uncertainty of the prediction is reduced by a factor of 2.8 with respect to that presented in Sec. IX A, while the disagreement remains the same. In conclusion, the discrepancy observed in this study cannot arise from obvious deficiencies of the prediction.

We have investigated the possibility that the rate of fake SLT tags might be higher in jets with heavy flavor than in jets due to light partons. The correlation between the fake and HF predictions, established by the previous comparison between the total number of observed and predicted tags in generic jets, would require that an increase of the fake rate is compensated by a smaller efficiency of the SLT algorithm in the simulation, and it would not reduce the disagreement between data and prediction observed in the inclusive lepton sample. However, it is of interest to show this study in anticipation of the next subsection.

The parametrization of the SLT fake rate has been derived in generic-jet data without distinguishing between muons faked by hadrons not contained by the calorimeter and muons produced by in-flight decays of π and K mesons. The second contribution is believed to be small because the reconstruction algorithms reject tracks that exhibit large kinks, but this has never been carefully checked. Away-jets in the inclusive lepton sample have a larger heavy flavor content ($\simeq 26\%$) than the generic jets used to determine the SLT fake rate ($\simeq 13\%$), and possibly a larger kaon content. Since kaons have a shorter lifetime than pions, in-flight decays of kaons could increase the SLT fake rate in the inclusive lepton sample with respect to generic-jet data. We verify the contribution of kaon in-flight decays by using a combination of data and simulation. First we extend the simulation of the SLT algorithm to match tracks not only to leptons originating from heavy quark decays at generation level but also to muons originating from kaon decays at detector simulation level. With this implementation, the rate of SLT tags in the simulation increases by only 1% (from 746.9 to 754.4 tags).

We check the simulation result within a factor of two by selecting $D^0 \to K\pi$ decays in the data and in the tuned simulation. As done in previous analyses [34], we search the inclusive lepton sample for $D^0 \to K^-\pi^+$ decays near the trigger leptons. To increase the sample statistics we do not require that leptons are contained in a jet with transverse energy larger than 15 GeV. The $D^0 \to K^-\pi^+$ decays are reconstructed as follows. We select events in which a cone of radius 0.6 around the lepton direction contains only two SVX tracks with opposite charge, $p_T \geqslant 1.0$ GeV/c, and an impact parameter significance larger than two.[9] We reconstruct the two-track invariant mass attributing the kaon mass to the track with the same charge as the lepton as is the case in semileptonic B decays. The resulting $K^-\pi^+$ invariant mass spectrum is shown in Fig. 10 together with a polynomial fit to the background which ignores the mass region between 1.7 and 2.0 GeV/c^2. According to the fit, in the mass range 1.82–1.92 GeV/c^2 the simulation contains 563 D^0 mesons on top of a background of 95 events (the corresponding 563 kaons are also identified at generator level). We find that one kaon in 563 D^0 decays produces a soft muon tag, which corresponds to 0.0018 SLT tags per kaon.

The data contain 1117 $K^-\pi^+$ pairs in the mass range 1.82–1.92 GeV/c^2 (891 are attributed by the fit to D^0 mesons and 226 to the background). The 1117 kaon tracks produce 6 SLT tags. The contribution of the background is estimated from the sidebands (1.64–1.74 and 2.0–2.1 GeV/c^2) to be 3.8 ± 1.0 events. It follows that 891 kaons from D^0 decays produce 2.2 ± 2.6 SLT tags. The fraction of SLT tags per kaon, 0.0024 ± 0.0029, includes the fake-tag contribution, and is consistent with the small fraction predicted by the simulation. We conclude that in-flight decays of K mesons are a negligible background contribution.

B. *b* purity of the data sample

The discrepancy between observed and predicted number of a-jets with SLT tags due to heavy flavor would be reduced if the $b\bar{b}$ contribution was underestimated by the simulation. In this section, we verify that the $b\bar{b}$ contribution is predicted correctly. As shown in Table X, the inclusive electron simulation predicts that 79% of the away-jets with heavy flavor are due to $b\bar{b}$ production. This table also shows that the fraction of away-jets with an SLT tag is higher in events due

[9]The impact parameter is the distance of closest approach to the primary vertex in the transverse plane.

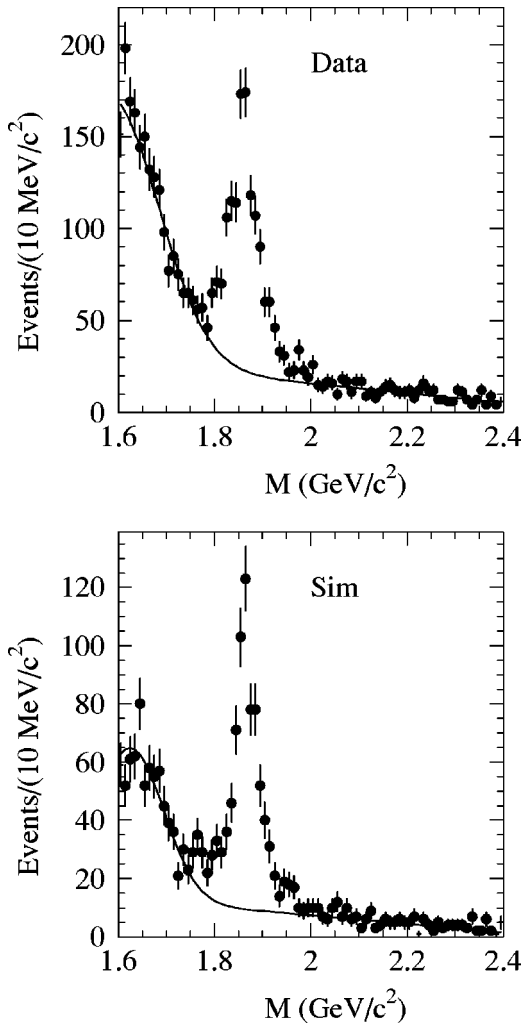

FIG. 10. Distributions of the $K\pi$ invariant mass, M. The solid line is a polynomial fit to the distributions excluding the window between 1.7 and 2.0 GeV/c^2.

to $b\bar{b}$ production (2%) than in events due to $c\bar{c}$ production (1%). If one had a reason to increase the b purity in the simulation from 79% to 100%, one could increase the predicted number of a-jets with a SLT tag in Table IX from 598 to 756, which is closer to the 865 ± 115 a-jets with a SLT tag due to heavy flavor in the data. We provide an independent check of the b purity of the inclusive lepton sample by comparing the number of D^0, $D^{\pm}$, and J/ψ mesons from B decays which are contained in lepton-jets in the data and in the normalized simulation.

1. l^-D^0 and l^+D^- candidates

We identify l^-D^0 candidates searching for $D^0\to K^-\pi^+$ decays inside the lepton-jet, as explained in the previous section. In a similar way, we identify l^+D^- pairs searching for $D^-\to K^+\pi^-\pi^-$ decays inside the lepton-jet. In this case, we select jets containing one positive and two negative tracks with $p_T\geqslant 0.6$ GeV/c and impact parameter significance larger than 2.5 in a cone of radius 0.6 around its axis. When reconstructing the three-track invariant mass, we attribute the

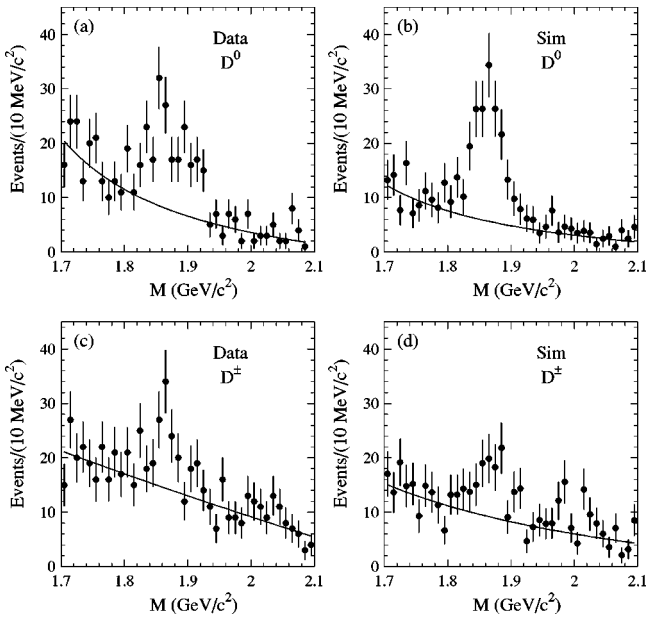

FIG. 11. Invariant mass distributions of D^0 candidates in the data (a) and in the simulation (b) and of $D^{\pm}$ candidates in the data (c) and in the simulation (d). The solid line is a polynomial fit to the mass distributions excluding the region 1.75–2.0 GeV/c^2.

kaon mass to the track with the same charge as the lepton as is the case in semileptonic B decays.

Figure 11 shows the invariant mass distributions of D^0 and $D^{\pm}$ candidates found in the data and in the fitted simulation. By comparing with Fig. 10, one notes that the mass resolution is degraded when using tracks inside a jet and is degraded slightly differently in the data and in the simulation.

There are 83510 lepton-jets in the data with an estimated heavy flavor purity $F_{HF}=(47.9\pm2.0)\%$. The simulation normalized according to Table IV contains 39989 lepton-jets with heavy flavor. In the mass range 1.82–1.92 GeV/c^2, we find 205 D^0 candidates in the data and 195.5 D^0 candidates in the simulation. By fitting the sidebands with a polynomial function (solid line in Fig. 11), we evaluate a background of 79.6 ± 6.0 events in the data and of 55.6 ± 5.5 events in the simulation. After background subtraction, there are 126.0 ± 15.5 D^0 mesons in the data and 139.9 ± 15.0 D^0 mesons in the simulation.

In the mass range 1.82–1.92 GeV/c^2, there are 216 $D^{\pm}$ candidates in the data and 159.2 in the simulation. By fitting the sidebands with a polynomial function we estimate a background of 142.3 ± 10.0 events in the data and of 90.7 ± 6.4 events in the simulation. After background subtraction we find 73.7 ± 17.8 $D^{\pm}$ mesons in the data and 68.5 ± 14.1 in the simulation. From the ratio of the numbers of lD candidates, we derive that the ratio of the $b\bar{b}$ production in the simulation to that in the data is 1.09 ± 0.15.

2. J/ψ candidates

We look for J/ψ candidates by searching the electron- or muon-jet for additional soft lepton tags with the same flavor and opposite sign charge (OS). Dileptons with invariant mass

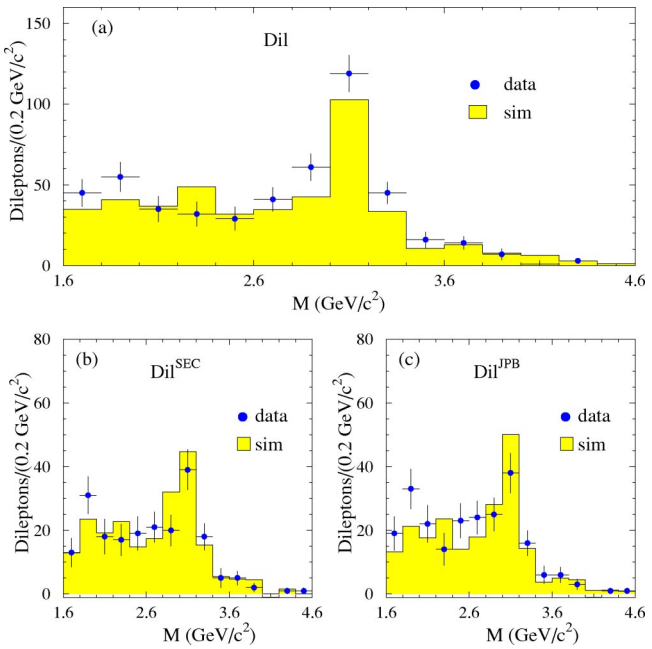

FIG. 12. Distributions of the invariant mass of same flavor dileptons inside the same jet before (a) and after tagging with SECVTX (b) and JPB (c).

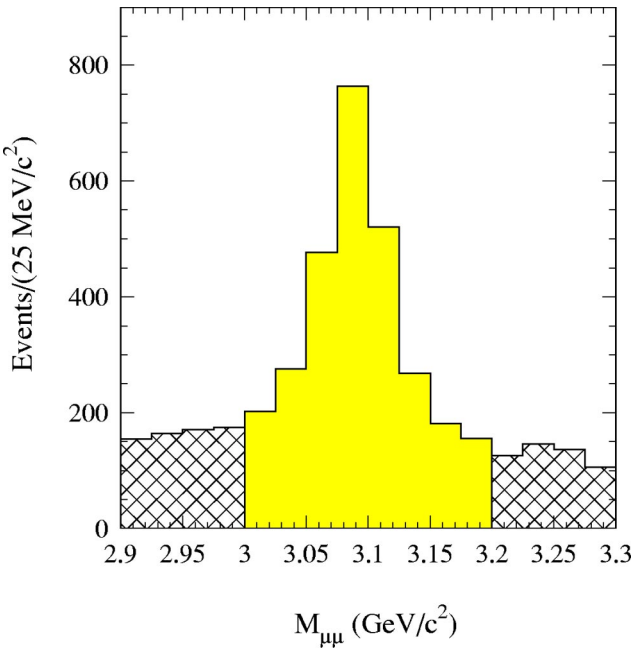

FIG. 13. Invariant mass distribution of muon pairs. The shaded area indicates the J/ψ signal region and the cross-hatched area indicates the sideband region, SB, used to estimate the background.

$2.6 \leqslant m_{ee} \leqslant 3.6$ GeV/c^2 and $2.9 \leqslant m_{\mu\mu} \leqslant 3.3$ GeV/c^2 are considered J/ψ candidates (Dil$_\psi$). Dil^{SEC} and Dil^{JPB} are the numbers of J/ψ candidates in lepton-jets tagged by SECVTX and JPB, respectively. We use the number of dileptons with same sign charge (SS) with a 10% error to estimate and remove the background to OS dileptons due to misidentified leptons [35].

Figure 12 compares invariant mass distributions of same flavor dileptons including J/ψ mesons in the data and in the simulation (in the simulation J/ψ mesons are only produced by B decays). Rates of J/ψ mesons in the data and in the normalized simulation are listed in Table XIV. One notes that the simulation contains a number of J/ψ mesons in jets tagged by SECVTX or JPB, which is slightly higher than, but consistent with the data. Before tagging, the rate of J/ψ mesons in the data is 20% larger than in the simulation, whereas it was expected to be larger by a factor of two according to the CDF measurement of the fraction of J/ψ's coming from B decays [36]. This would happen if the $b\bar{b}$ cross section had been overestimated in normalizing the simulation.

After combining the ratio of lD candidates in the data to that in the simulation with the ratio of lJ/ψ candidates with a JPB tag listed in Table XIV, we estimate that the ratio of the $b\bar{b}$ production in the simulation to that in the data is 1.09 ± 0.11. This ratio is consistent with unity, and does not support the possibility that the b purity in the fitted simulation is underestimated by 21%.

TABLE XIV. Number of J/ψ mesons identified in the data and in the fitted simulation.

	Electrons		Muons	
Tag type	Data	Simulation	Data	Simulation
Dil$_\psi$	176.0 ± 14.4	155.2 ± 21.5	83.0 ± 9.4	54.0 ± 10.1
Dil$_\psi^{SEC}$	57.8 ± 8.8	71.8 ± 10.7	31.9 ± 5.8	28.7 ± 6.2
Dil$_\psi^{JPB}$	61.2 ± 8.4	68.9 ± 9.4	29.6 ± 5.7	33.0 ± 6.4

C. $J/\psi \to \mu\mu$ data

As shown in Table X, away-jets with a supertag are mostly due to $b\bar{b}$ production as it is the case for generic jets with a supertag. However, we see a discrepancy between observed and predicted number of supertags after having calibrated the supertag efficiency in the simulation by using generic jets. Since this is suggestive that the excess of SLT tags in the away-jets is related to the request that a jet contains a presumed semileptonic b decay (lepton-jet), we study a complementary data sample enriched in $b\bar{b}$ production but not in semileptonic b decays, i.e. events containing $J/\psi \to \mu^+\mu^-$ decays. The data sample consists of $\simeq 110$ pb^{-1} of $p\bar{p}$ collisions collected by CDF during the 1992–1995 collider run. This sample has been used for many analyses and is described in detail in Ref. [37]. Approximately 18% of these J/ψ mesons come from B decays [36]. Muon candidates are selected as in Ref. [37]. Since we want to make use of the B lifetime to remove the contribution of prompt J/ψ mesons, we select muons with SVX tracks. The dimuon invariant mass is calculated without constraining the two muon tracks to a common vertex since the mass resolution is not important in this check. In addition we require a jet with transverse energy larger than 15 GeV lying in the hemisphere opposite to the J/ψ and contained in the SVX acceptance.

The dimuon invariant mass distribution in these events is shown in Fig. 13. In the mass range between 3 and

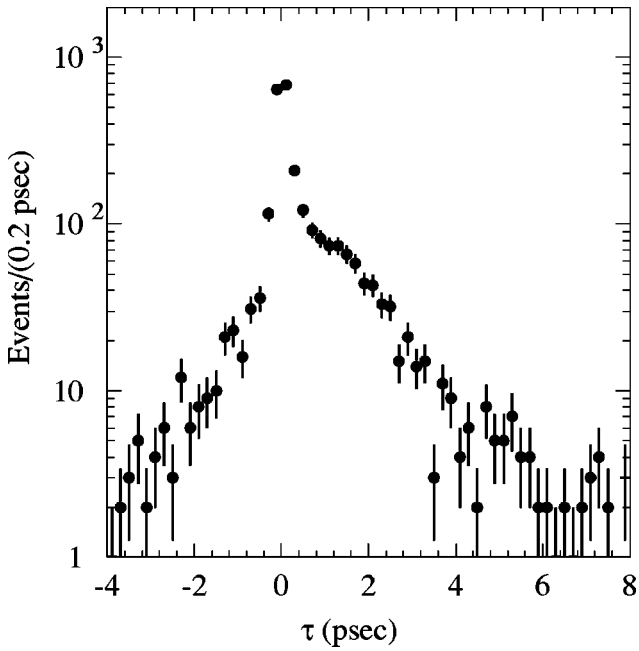

FIG. 14. Lifetime distribution of J/ψ candidates.

3.2 GeV/c^2 there are 1163 J/ψ events over a background of 1179 events estimated from the sideband region (see Fig. 13).[10]

The J/ψ lifetime is defined as

$$\tau = \frac{(\vec{L}\cdot\vec{p}_T)\cdot M}{c\cdot p_T^2},$$

where M and p_T are the dimuon invariant mass and transverse momentum and L is the distance between the event vertex and the origin of the muon tracks. The lifetime distribution of J/ψ candidates is shown in Fig. 14. As studied in Ref. [37], prompt J/ψ candidates produce a symmetric τ distribution peaking at $\tau=0$. We call ψ^+ and ψ^- the numbers of J/ψ candidates with positive and negative lifetime; SB^+ and SB^- are the analogous numbers for the sideband region, which is used to estimate the background in the invariant mass distribution. The number of J/ψ mesons from B decays is then $N_\psi = \psi^+ - \psi^- - (SB^+ - SB^-) = 561$, which is 48% of the initial sample. In the opposite hemisphere we find 572 away-jets. In these *a*-jets we measure the following numbers of tags after mistag removal:

(1) 48.0 ± 15.1 SECVTX tags
(2) 61.7 ± 17.3 JPB tags
(3) -9.4 ± 14.4 SLT tags

For 54.8 ± 11.5 lifetime tags (average of the observed number of SECVTX and JPB tags) the simulation predicts 8.1 ± 1.7 SLT tags. The observed number of SLT tags is 1.2σ lower than the prediction rather than 50% larger as in the inclusive lepton sample.

[10]The request of a recoiling away-jet reduces the number of J/ψ mesons in the original data set by a factor of $\simeq 200$.

XI. CONCLUSIONS

We have studied the heavy flavor properties of jets produced at the Tevatron collider. This study is motivated by the evidence, reported in Ref. [1], for a class of jets that contain long-lived objects consistent with *b*- or *c*-quark decays, identified by the presence of secondary vertices (SECVTX tags) or of tracks with large impact parameters (JPB tags), but which also have an anomalously large content of soft leptons (SLT tags); we refer to these as superjets and supertags. The study in Ref. [1] focused on high-p_T jets produced in association with W bosons. The analysis reported here uses a much larger data set collected with low-p_T lepton triggers ($p_T \geqslant 8$ GeV/c). This data set has been previously used to study bottom and charmed semileptonic decays, and to provide calibrations for the measurement of the pair production of top quarks [16].

In the present analysis, we study events having two or more central jets with $E_T \geqslant 15$ GeV, one of which (lepton-jet) is consistent with a semileptonic bottom or charmed decay to a lepton with $p_T \geqslant 8$ GeV/c. The measurement is a comparison between the data and a HERWIG-based simulation of the semileptonic decay rate for the additional jets (away-jets), which have no lepton trigger requirement. We first use measured rates of lepton- and away-jets with SECVTX and JPB tags in order to determine the bottom and charmed content of the data; we then tune the simulation to match the observed heavy-flavor content. Rates of SECVTX and JPB tags and the kinematics of these events are well modeled after tuning the parton-level cross sections predicted by HERWIG within the experimental and theoretical uncertainties. The tuned parton-level prediction of HERWIG indicates that, in order to model the single *b* production cross section measured at the Tevatron, any theoretical calculation should predict higher-order-term contributions approximately four times larger than the LO contribution.

We then measure the yields of soft ($p_T \geqslant 2$ GeV/c) leptons due to heavy-flavor decays in the away-jets, and compare them to the prediction of the tuned simulation. The latter depends on the bottom and charmed semileptonic decay rates and on the soft lepton reconstruction efficiency. To calibrate the predictions of the simulation, we perform the same analysis on samples of generic jets with 20, 50, and 100 GeV E_T thresholds; these samples have also been previously used to calibrate the simulation of heavy flavor background to pair production of top quarks [16].

Finally, with these calibrations we find that away-jets have a 30–50% excess of soft lepton tags as compared with the simulation, corresponding to $(2.5\text{–}3.5)\sigma$, depending on the selection of the away-jets; the selections include (a) all away-jets, (b) a subset with SECVTX tags, and (c) another subset with JPB tags (the three results are highly correlated and should not be combined). The size of this excess is consistent with the differences between the NLO prediction and the $b\bar{b}$ cross section measurements at the Tevatron that are based upon the detection of one and two leptons from *b*-quark decays. A possible interpretation of this excess, the one that motivated this study, is the pair production of light scalar quarks with a 100% semileptonic branching ratio. Due

to the $p_T \geqslant 8$ GeV/c lepton-trigger requirement, we expected such a signature to be enhanced in this sample as compared with generic-jet data. However, alternative explanations for the excess are not excluded by this study, the interpretation of which requires independent confirmations.

ACKNOWLEDGMENTS

We thank the Fermilab staff and the technical staff of the participating Institutions for their contributions. This work was supported by the U.S. Department of Energy and National Science Foundation; the Istituto Nazionale di Fisica Nucleare; the Ministry of Education, Culture, Sports, Science and Technology of Japan; 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; the Korea Science and Engineering Foundation (KoSEF); the Korea Research Foundation; and the Comision Interministerial de Ciencia y Tecnologia, Spain. We benefitted from many useful discussions with M. Seymour and S. Moretti.

[1] D. Acosta *et al.*, Phys. Rev. D **65**, 052007 (2002).
[2] P. Dawson *et al.*, Nucl. Phys. **B327**, 49 (1988).
[3] E.L. Berger *et al.*, Phys. Rev. Lett. **86**, 4231 (2001).
[4] E. Braaten *et al.*, Phys. Rev. D **66**, 034003 (2002).
[5] F. Abe *et al.*, Phys. Rev. D **53**, 1051 (1996).
[6] F. Abe *et al.*, Phys. Rev. D **55**, 2546 (1997).
[7] B. Abbott *et al.*, Phys. Lett. B **487**, 264 (2000).
[8] M. Mangano *et al.*, Nucl. Phys. **B373**, 295 (1992). The FORTRAN code is available from the authors, and was used to evaluate the cross sections for these particular topologies.
[9] S. Frixione *et al.*, Adv. Ser. Dir. High Energy Phys. **15**, 609 (1998).
[10] H. Jung, J. Phys. G **28**, 971 (2002).
[11] H.E. Haber and G.L. Kane, Phys. Rep. **117C**, 76 (1985).
[12] C. Nappi, Phys. Rev. D **25**, 84 (1982).
[13] M. Carena *et al.*, Phys. Rev. Lett. **86**, 4463 (2001).
[14] J. Cao *et al.*, Phys. Rev. Lett. **88**, 111802 (2002).
[15] S. Baek, Phys. Lett. B **541**, 161 (2002).
[16] T. Affolder *et al.*, Phys. Rev. D **64**, 032002 (2001).
[17] G. Marchesini and B.R. Webber, Nucl. Phys. **B310**, 461 (1988); G. Marchesini *et al.*, Comput. Phys. Commun. **67**, 465 (1992).
[18] D. Buskulic *et al.*, Phys. Lett. B **313**, 535 (1993).
[19] W. Beenakker *et al.*, Nucl. Phys. **B492**, 51 (1997); hep-ph/9611232.
[20] A.D. Martin, R.G. Roberts, and W.J. Stirling, Phys. Lett. B **354**, 155 (1995).
[21] Particle Data Group, D.E. Groom *et al.*, Eur. Phys. J. C **15**, 1 (2000).
[22] F. Abe *et al.*, Nucl. Instrum. Methods Phys. Res. A **271**, 387 (1988).
[23] D. Amidei *et al.*, Nucl. Instrum. Methods Phys. Res. A **350**, 73 (1994).
[24] F. Abe *et al.*, Phys. Rev. D **60**, 051101 (1999).
[25] F. Abe *et al.*, Phys. Rev. D **45**, 1448 (1992).
[26] An SVX track is reconstructed in the central tracking chamber (CTC) with at least two associated SVX hits. It is refitted using the SVX hits and the CTC track parameters and covariance matrix.
[27] F. Abe *et al.*, Phys. Rev. D **50**, 2966 (1994).
[28] D. Kestenbaum, Ph.D. thesis, Harvard University, 1996.
[29] Version 9_1 of the CLEO simulation; P. Avery, K. Read, G. Trahern, Cornell Internal Note CSN-212, March 25, 1985 (unpublished).
[30] R.D. Field, Phys. Rev. D **65**, 094006 (2002).
[31] A. Ballestrero *et al.*, hep-ph/0006259; M. Mangano, in *International Europhysics Conference on High-Energy Physics* (EPS-HEP99), Tampere, 1999, p. 33, hep-ph/9911256.
[32] M. Seymour, Nucl. Phys. **B426**, 163 (1995); M. Mangano, *ibid.* **B405**, 536 (1993).
[33] JET NN data are collected with a Level 2 trigger which requires a calorimetry cluster with transverse energy larger than NN GeV.
[34] F. Abe *et al.*, Phys. Rev. D **58**, 092002 (1998); Phys. Rev. Lett. **76**, 4462 (1996).
[35] F. Abe *et al.*, Phys. Rev. D **59**, 052002 (1999); **49**, 1 (1994).
[36] F. Abe *et al.*, Phys. Rev. Lett. **79**, 572 (1997); **79**, 578 (1997).
[37] F. Abe *et al.*, Phys. Rev. D **57**, 5382 (1998); **55**, 1142 (1997).