Fermilab FERMILAB-Conf-02/123-E August 2002

STUDY OF THE HEAVY FLAVOR CONTENT OF JETS PRODUCED IN W EVENTS AT THE TEVATRON COLLIDER

Giorgio Apollinari
Fermi National Accelerator Laboratory, Batavia, Illinois 60510 USA
(on behalf of the CDF Collaboration)

Abstract

We present a detailed study of the heavy flavor content in the W + jet data sample collected with the CDF detector during the 1992-1995 collider run at Fermilab. Rates of heavy flavor jets, identified via the observation of secondary vertices or semileptonic decays of b and c quarks, are in good agreement with a standard model simulation including production of the top quark. An exception is the number of events in which a single jet has both a secondary vertex and a semileptonic decay tag. In the $W+$ 2,3 jet data, we expect 4.4 ± 0.6 and we find 13 such events. The kinematic properties of this small sample of events are statistically difficult to reconcile with the simulation of standard model processes.

1 Introduction

W bosons produced in association with jets in $p\bar{p}$ collisions at the Fermilab Tevatron Collider provide the opportunity to test many Standard Model (SM) [1)] predictions. CDF has measured the b and c-quark content of a data set containing 11,076 $W \to \ell\nu$ ($\ell = e$ or μ) candidates produced in association with one or more jets. Since a contribution to heavy flavors in $W+$ jet sample is coming from the production and decay of top quarks, these events have been used in the past to measure $\sigma_{t\bar{t}}$ by attributing to $t\bar{t}$ production all the excesses over expectations of $W+\geq 3$ jet events with an identified heavy flavor jet.

In this paper we adopt a different approach by studying the $W+$ jet sample assuming the theoretical estimate of $\sigma_{t\bar{t}}$ to test if the complete SM prediction is compatible with the observed yield of different tags as a function of the jet multiplicity. This is of interest for top quark studies and searches for new physics, since some mechanisms proposed to explain electroweak symmetry breaking, such as the Higgs mechanism [2)] or the dynamics of a new interaction [3)], predict the existence of new particles which can be produced in association with a W boson and decay into $b\bar{b}$.

2 Data Analysis

The CDF detector as well as the data sample and the physics objects used in the $W+$ jet analysis have been described in detail elsewhere [4, 5)]. The W selection requires an isolated electron (muon) with $E_T(P_T) \geq 20$ GeV and $\not\!\!E_{\rm T} \geq 20$ GeV to reduce the background from misidentified leptons and semileptonic b-hadron decays. Events containing additional lepton candidates are removed from the sample. W candidate events are binned according to the observed jet multiplicity (a jet is a calorimeter energy cluster with uncorrected $E_T \geq 15$ GeV and $|\eta| \leq 2.0$).

The heavy flavor content of the $W+$ jet sample is enhanced by selecting events with jets containing a displaced secondary vertex or a soft lepton.

The secondary vertex tagging algorithm (SECVTX) [6, 5)] is based on the determination of the primary event vertex and the reconstruction of additional secondary vertices using displaced tracks contained inside the jets[1].

[1]A jet is considered tagged by SECVTX if some of the jet tracks form a secondary vertex with $\frac{L_{xy}}{\sigma_{L_{xy}}} \geq 3.0$, where $\sigma_{L_{xy}}$ is the estimated uncertainty on L_{xy} (typically about 130 μm) and L_{xy} is the projection in the plane transverse to the beam line of the vector pointing from the primary vertex to the secondary vertex onto the jet axis.

An alternative way to tag b quarks is to search a jet for soft leptons produced by $b \to l\nu c$ or $b \to c \to l\nu s$ decays. The soft lepton tagging algorithm is applied to the tracks associated with a jet and contained in a cone of radius 0.4 in $\eta - \phi$ space centered around the jet axis. In order to maintain high efficiency, the p_T threshold for the lepton candidate is set low at 2 GeV/c. Soft electrons and muons candidates are identified by their matching signatures in the electromagnetic calorimeter and muon chambers respectively. A jet is considered tagged by SLT if it contains a soft lepton candidate.

3 Supertag Events

We call *multitag event* an event containing both an SLT and a SECVTX tag, not necessarily on the same jet. A *supertag event* is an event with an SLT and a SECVTX tag in the same jet. The corresponding jet is called *superjet* [2].

The study presented here starts with the comparison between observation and SM expectations for events with at least a single tag. The expectations are based on Monte Carlo calculations with the settings and the calibration described in Ref. [5]. The non-$t\bar{t}$ SM contributions, together with the $t\bar{t}$ contribution (derived using the theoretical prediction $\sigma_{t\bar{t}} = 5.1 \;\; pb$ [7]) and the data are shown in figure 1 (right) for $W+$ jet events with at least one SECVTX tag. We observe good agreement between data and expectations. The probability [8] that the observed numbers of events with at least one SECVTX tag are consistent with the SM predictions in all four jet bins is 80%. The similar analysis for events with at least one SLT tag gives a probability of 56%.

We continue the study by selecting $W+$ jet events with both SECVTX and SLT tags, where we expect the events to be mostly contributed by real heavy flavor. As shown in figure 1 (left) the numbers of *multitag events* are not well predicted by the simulation. In particular figure 2 shows the comparison between the rate of observed *supertag events* and the expected production and decay of hadrons with heavy flavor. It is interesting to notice that 16 of the 18 multitag events shown in figure 1 are due to *superjets.*

The probability [8] that the observed numbers of events with at least one superjet are consistent with the prediction in all four jet bins is 0.4%. This low probability value is mostly driven by the excess in the $W+$ 2,3 jet bins where 13

[2] The prefix "super" is used as a generalized term of high quality. No reference to any particular physics model is implied.

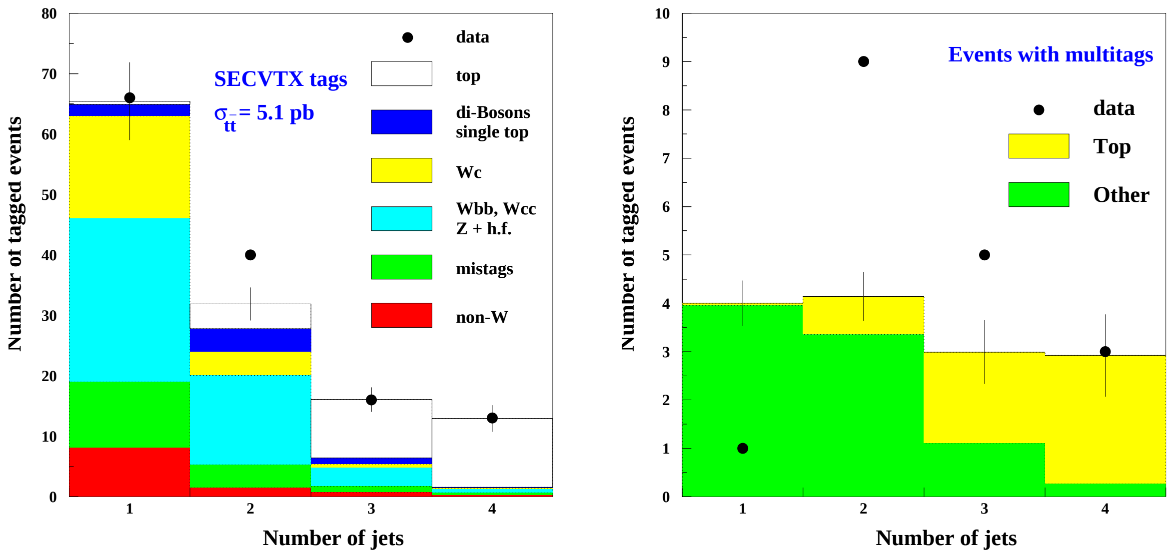

Figure 1: *Summary of observed and predicted number of W events with at least one SECVTX tag as a function of the jet multiplicity (left) and with a multitag (right).*

Table 1: *Fractions of SECVTX with a supertag (* $\frac{\text{SLT+SECVTX}}{\text{SECVTX}}$ *) in generic-jet data and in the corresponding simulation. In the simulation the fraction of supertags is slightly higher than in the data, independent of the jet transverse energy and the heavy flavor type.*

	JET 20	JET 50	JET 100
Data	0.054±0.005	0.067±0.006	0.069±0.007
Sim.	0.065±0.007	0.076±0.004	0.086±0.006
Data/Sim.	0.83±0.12	0.88±0.09	0.80±0.10

events are observed[3] and 4.4 ± 0.6 are expected from SM sources. The *a posteriori* probability of observing no less than 13 events is 0.1%.

The cause of the excess of $W+$ 2,3 jet events with supertags could be a discrepancy in the correlation between the SLT and SECVTX efficiencies in the data and simulation. These simulated efficiencies have been tuned separately using the data and, in principle, the SLT tagging efficiency in jets already tagged by SECVTX could be higher in the data than in the simulation. We have checked this possible bias using generic-jet data (table 1) and we conclude that the excess of $W+$ 2,3 jet events with a supertag cannot be explained by this type of simulation deficiency.

[3]The 13 events include $t\bar{t}$ candidates and four of these events are included in the sample used to measure the top quark mass [5].

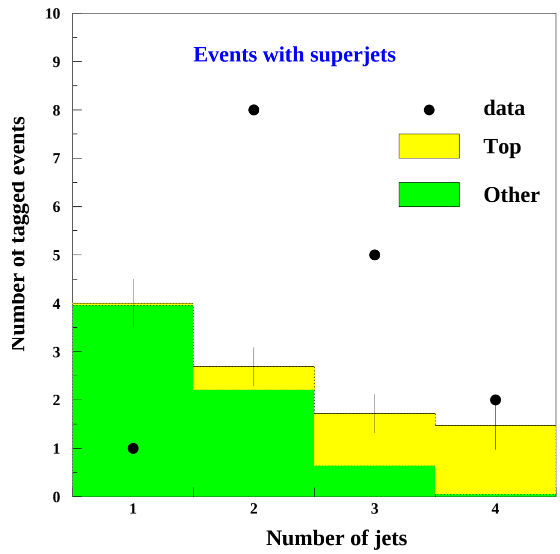

Figure 2: *Summary of observed and predicted number of W events with at supertag as a function of the jet multiplicity.*

4 Supertag Events and SM Simulation

The events with a supertag are a subset of all the events with a SECVTX tag and a track in the jet that is "SLT taggable". In order to define an ortogonal and complementary sample for comparison studies, we select all the events where the "SLT taggable" track does not generate an actual SLT tag. By construction these two samples have similar heavy flavor compositions. The complementary sample of $W+$ 2,3 jet events consists of 42 events, while 41.2 $\pm$ 3.1 events are predicted by the SM simulartion.

We compare the distributions of several simple kinematic variables for the 13 supertag events in the $W+$ 2,3 jet sample and for the complementary sample of 42 events with the expectations from the SM simulation. The comparison is performed by measuring the Kolmogorov-Smirnov (K-S) distance [9] between the data and the simulation distributions. The probability of the measured K-S distance is determined by pseudoexperiments. We examined the transverse energy and pseudorapidity of each physics object in the final state as well as the invariant masses or other kinematical quantities for various combinations of the physics objects. The probabilities of consistency with the SM for all the kinematical variables we studied are listed in table 2.

Figure 3 shows the measured probabilities for the superjet sample (left) and the complementary sample (right). The probabilities of the complementary sample appear to be flatly distributed, as expected for a set of distributions consistent

Table 2: *Probabilities of the K-S comparisons between data and simulation for the superjet sample and the complementary sample.*

	Superjets Sample	Compl. Sample		Superjets Sample	Compl. Sample
Variable	P (%)	P (%)	Variable	P (%)	P (%)
E_T^l	2.6	70.9	$\not{E}_{\rm T}$	27.1	57.1
η^l	0.10	72.7	M_T^W	13.1	38.2
E_T^{suj}	11.1	43.0	M^{suj+b}	4.0	58.9
η^{suj}	15.2	73.4	y^{suj+b}	7.1	34.9
E_T^b	6.7	8.6	E_T^{suj+b}	24.0	60.1
η^b	6.8	80.0	$M^{l+suj+b}$	21.0	33.6
$E_T^{l+b+suj}$	2.5	18.8	$\theta^{suj,b}$	30.1	41.1
$y^{l+b+suj}$	13.8	83.8	$\phi^{suj,b}$	15.3	83.8
$\delta\phi^{l,b+suj}$	1.0	77.9	$\theta^{l,suj+b}$	37.3	35.7

with the simulation. In contrast, the probabilities of the superjet events cluster at low values. This indicates the difficulty of our SM simulation to describe the kinematics of $W+$ 2,3 jet events with a jet in which we require the two prominent features of heavy flavors: long lifetime (SECVTX) and semileptonic decay (SLT). The same simulation describes well the characteristics of other data samples, like the complementary sample or the QCD sample.

5 Properties of the Superjets and Superjet Events

5.1 *Primary Lepton Properties*

The kinematics of the primary leptons in events with a superjet has a low probability of being compatible with a W decay as predicted by the SM simulation (table 2). However both the signed impact parameter significance and the isolation of the primary lepton (shown in figure 4 (left) together with the same variables for the leptons in the complementary sample) indicate that the primary leptons in the superjet events are indeed *primary* and cannot be attributed to non-W background

5.2 *Soft Lepton Properties*

Based on the SM expectation, the average transverse momenta of primary and soft leptons are expected to differ by an order of magnitude. However, in the superjet events the average transverse momenta are 35 and 13 GeV/c, respectively. The superjets could be due to dilepton events (either from Drell-Yan or poorly

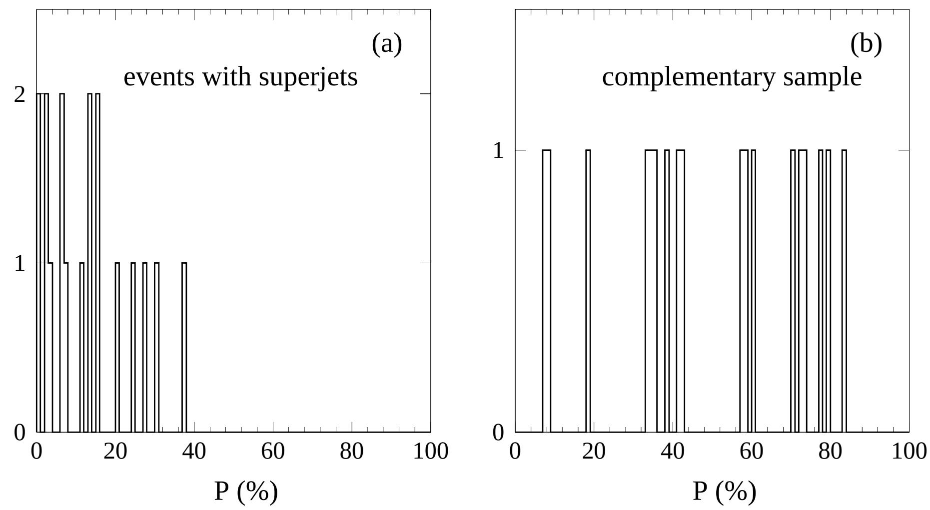

Figure 3: *Distribution of the probabilities P that the 13 events with a superjet (left) and the complementary sample (right) are consistent with the SM prediction. The superjet sample distribution has a mean of 0.13 and a RMS of 0.11, the complementary sample distribution has a mean of 0.50 and a RMS of 0.24.*

measured Z decays) which are not removed because the second lepton happens to be merged with a jet and is not isolated. Figure 4 (right) shows that soft leptons are mostly found close to the superjet axis and are not uniformly distributed over the jet clustering cone as would be expected in the random overlap scenario described above. In addition, the figure shows that soft leptons inside a superjet are not prompt. As expected from the simulation of heavy flavor decays, the soft lepton track is part of the SECVTX tag in 8 out of 13 superjets.

5.3 *Superjet Lifetime Properties*

A measure of the lifetime of an hadron producing a secondary vertex is

$$pseudo-\tau = \frac{L_{xy}}{c}\frac{M^{SVX}}{p_T^{SVX}} \qquad (1)$$

where L_{xy} is the projection of the transverse displacement of the secondary vertex on the jet-axis, M^{SVX} is the invariant mass and p_T^{SVX} is the total transverse momentum of all tracks associated with the secondary vertex. This approach approximates the Lorentz boost of the heavy flavor hadron with the Lorentz boost of the tracks contributing to the SECVTX tag. A kinematical correction factor is typically introduced to account for the undetected neutral particles emitted at the decay point

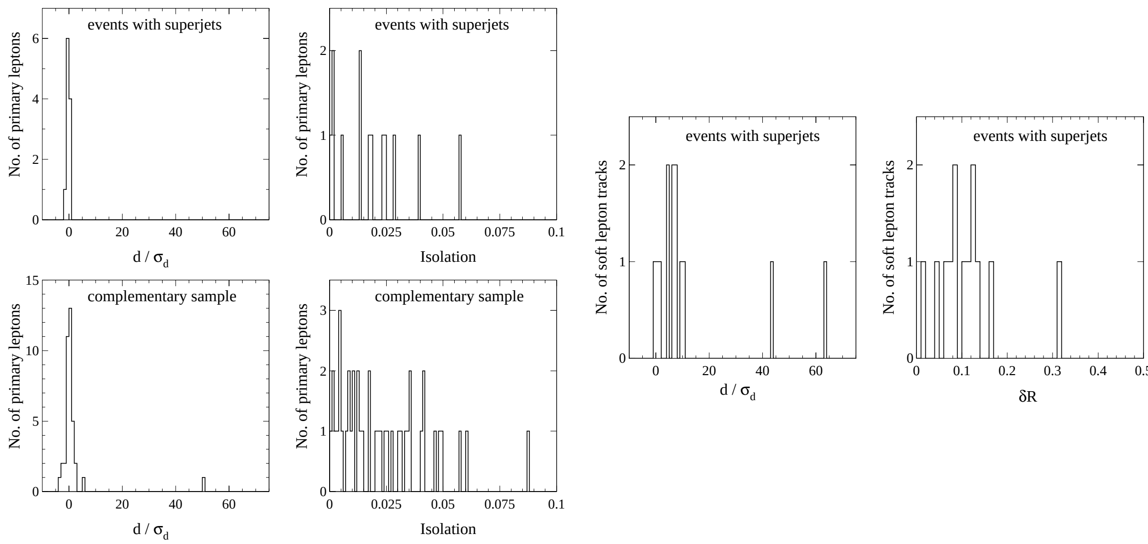

Figure 4: *Distribution of the signed impact parameter significance d/σ_d and the isolation I of primary lepton for the superjet sample (top left) and the complementary sample (bottom left). The plots on the right show the distribution of the soft lepton impact parameter significance and distance $\delta R = \sqrt{\delta\phi^2 + \delta\eta^2}$ from the superjet axis for the soft leptons in the superjet sample.*

In the data we observe a pseudo-τ distribution for the 13 superjets in qualitatively agreement with the one observed in the complementary sample. Assuming a correction factor similar to the one used for the decay of beauty hadrons, this observation suggests a superjet lifetime of the order of 1 ps.

Additional discussion of the superjet lifetime is provided in [10]. The same work describes additional studies performed on the superjet samples, including a detailed examination of the fragmentation properties of the superjets and a comparison to generic QCD samples and SM simulation. No anomalous behaviour has been observed in any of these studies.

6 Acceptance studies

We study the rate of superjet events as a function of the acceptance of our analysis cuts.

The selection criteria used in this analysis were optimized for finding the top quark [6]. The high-p_T inclusive lepton data set, from which we have selected the sample used in this study, consists of about 82,000 events. A large fraction of the events are due to multi-jet production with one jet containing a fake lepton. Small amounts of $b\bar{b}$ and $c\bar{c}$ production are also included. The $p_T \geq 20$ GeV/c, $I \leq 0.1$ and $\not{E}_{\mathrm{T}} \geq 20$ GeV cuts reduce this data set to an almost pure $W+$ jet sample of about 11,000 events. By releasing the p_T cuts ($p_T \geq 18$ GeV/c), the Isolation cut ($0.1 \leq I \leq 0.2$) or the $\not\!\!E_{\rm T}$ cut ($\not\!\!E_{\rm T} \leq 20$ GeV), no unexpected superjet event is found.

Additionally, we studied the rate of superjet events by removing the request that the primary lepton has fired the appropriate second level trigger in CDF, which is approximately 70% efficient [5] for muons. Based on the observed 13 events with a superjet, we should have lost about two such events because the primary muon is failing the muon trigger. On the other hand, since 85% of the superjet events contain a soft lepton with transverse momentum comparable or larger than the L2 trigger threshold (and assuming the hard spectrum of the soft leptons to be a real feature of the superjets) we should find in the original data sample one or two additional events with a supertag in which the primary muon failed the trigger but the event was rescued by the soft muon. In the data, after removing the trigger requirement on the primary muon, we recover three $W+$ 1 jet events, none of which contains supertags. We also recover one $W+$ 2 jet and one $W+$ 3 jet event, both with a supertag. No extra $W+$ 4 jet event is found. Scaling from the SM, we would have expected to find only 0.08 $W+$ 2,3 jet events with a supertag

A similar study has been performed extending the acceptance of CDF to electrons detected in the plug region ($1 \leq |\eta| \leq 1.5$). We select $W+$ jet events in the plug requiring an isolated electron with $E_T \geq 20$ GeV and $\not\!\!E_{\rm T} \geq 20$ GeV. We observe two additional $W+$ 2,3 jet events with a supertag, while 0.34 ± 0.04 events are expected from known processes scaling the SM expectations in the central region of the CDF detector.

It is interesting to notice that in 13 of the 17 events with a superjet the charges of the primary lepton and the soft lepton are opposite. No correlation between the charges is expected from the SM simulation. The probability of an equal or larger fluctuation is 2.4 %.

7 Conclusions

We have carried out a detailed study of the heavy flavor content of jets produced in association with W bosons. Comparisons of the observed heavy flavor rates with standard model predictions, including NLO calculations of single and pair produced top quarks, are generally in good agreement. However, we find an excess of events when we exploit the two basic characteristics of an heavy flavor decay by requiring a jet to be tagged by both a long lifetime (SECVTX) and by a semileptonic decay (SLT). The standard model expectation for these $W+$ 2,3 jet events is 4.4 ± 0.6 events, while 13 are observed. A detailed examination of the kinematic properties of these events finds that they are statistically difficult to reconcile with a simulation of standard model processes, which well reproduces closely related samples of data.

We are not aware of any model for new physics which incorporates the production and decay properties necessary to explain all features of these events. With much larger data samples from the Run II of the Tevatron, we will be able to explore in greater detail this class of events.

8 Acknowledgments

I wish to thank the organizers of this informative and enjoyable conference for their hospitality and interest in this new result from CDF. I acknowledge the support of the US Department of Energy, the CDF Collaboration and their funding Agencies.

9 Questions and Answers

Q: What is the effect of the supertag events on the measurement of $\sigma_{t\bar{t}}$?

A: CDF has prepared no revision to the measured value of $\sigma_{t\bar{t}}$. However Ref. [11] has performed a phenomenological analysis to try to quantify the production of the anomalous superjet events and their effect on the measurement of $\sigma_{t\bar{t}}$.

An interpretation in terms of new physics of the superjet properties would require the production of a low-mass, strongly interacting object, decaying semi-leptonically with a branching ratio close to 1 and with a lifetime of the order of a picosecond. The low mass is required by the fact that the SECVTX and SLT tags are contained in a cone of radius 0.4 in the $\eta - \phi$ space around the jet axis. The request that the object is strongly interacting is motivated by the fact that several tracks in the superjet point directly to the primary vertex, suggesting an emission of gluons during the evolution prior to the decay. The large semileptonic branching ratio, finally, is required to explain the large fraction of soft lepton tags in jets tagged by SECVTX.

Since there are no experimental limits on the existence of a charge-1/3 scalar quark with mass smaller than 7.4 GeV/c^2 [12], the supersymmetric partner of the bottom quark is a potential candidate. As an example, Ref. [11] uses the ansatz that superjets are due to the decay into $lc\tilde{\nu}$ of a scalar quark $\tilde{b}$ with a mass of 3.6 GeV/c^2 and a lifetime of 1.0 ps; the scalar neutrino is assumed to be massless and the presence of a charm quark in the final state is not necessary (i.e. the c can be substituted by any other appropriately charged and kinematically allowed quark). Furthermore, the event kinematics is modeled with the fictitious production of a heavy state N of mass 220 GeV/c^2 which decays into $\tilde{b}\bar{\tilde{b}}$ according to phase space.

In this framework, and assuming the $\tilde{b}$ decay to be mediated by the higgsino coupling to the right handed matter, Ref. [11] fit the rates of $W + n$ jets events ($n = 1, 4$) before and after tagging, with the inclusion of the state N in the SM simulation. The fit is performed using a likelihood technique with the $t\bar{t}$ cross section and the number of events due to N production as unconstrained parameters of the fit.

The value of the $t\bar{t}$ cross section returned by the fit is 4.0 $\pm$ 1.5 pb, in agreement with theoretical expectations [7] and the DO result [13]. The fit also returns 52.8 $\pm$ 22.1 events due to the production of a massive $\tilde{b}\bar{\tilde{b}}$ pair. Since the lepton and missing energy kinematics have been ignored, the acceptance cannot be readily calculated and Ref. [11] quotes the number of events rather than a cross-section.

Additionally, Ref. [11] describes two approaches used to compute the supertag events acceptance. The first approach uses a W+Higgs simulation where the Higgs is identified with the state N and the primary lepton angular distribution is arbitrarily weighted to obtain agreement with the data. The second approach models the supertag events production with a general $2 \to 4$ hard scattering process described by an effective lagrangian fitting the data. Both approaches provide an acceptance of approximately 10%. Therefore the 52.8 $\pm$ 22.1 events with a superjet returned by the previous fit correspond to a production cross section of approximately 5 $\pm$ 2 pb.

Q: A low-mass scalar quark should be visible at e^+e^- colliders.

A: The CLEO Collaboration [14] has reported a negative result for such a search. The CLEO Collaboration has assumed $\tilde{B} \to D^{(*)} l \tilde{\nu} + X$ as a possible decay chain for a dressed $\tilde{B}$ meson. Furthermore they have assumed a three body phase space for the decay matrix element. Decays resulting in leptons with a softer spectrum, such as the one used in Ref. [11], or without charmed particles in the final state have not been performed yet at e^+e^- colliders.

References

1. S. L. Glashow, Nucl. Phys. **22**, 579 (1961); S. Weinberg, Phys. Rev. Lett. **19**, 1264 (1967); A. Salam, in *Elementary Particle Theory: Relativistic Groups and Analyticity (Nobel Symposium No. 8)*, edited by N. Svartholm (Almqvist and

Wiksell, Sweden, 1968), p. 367; M. Gell-Mann, in *QCD, 20 years later*, edited by P. M. Zervas and H. A. Kastrup (World Scientific, 1993).

2. J. Gunion *et al., The Higgs Hunter's Guide* (Addison-Wesley, 1990).

3. E. Eichten and K. Lane, Phys. Lett. **B388**, 803 (1996).

4. D. Amidei *et al.,* Nucl. Inst. and Methods, **A271**, 387 (1988); Fermilab Report 94/024-E (1994).

5. T. Affolder *et al.,* Phys. Rev. **D63**, 032003 (2001) and references therein.

6. F. Abe *et al.,* Phys. Rev. **D50**, 2966 (1994); Phys. Rev. Lett. **73**, 225 (1994)

7. F. Bonciani *et al.,* Nucl. Phys. **B529**, 450 (1998); E. Berger and H. Contopanagos, Phys. Rev. **D54**, 3035 (1996); S. Catani *et al.,* Phys. Lett. **B378**, 329 (1996); E. Laenen *et al.,* Phys. Lett. **B321**, 254 (1994).

8. The probability is estimated with Monte Carlo pseudo-experiments which include both Poisson fluctuations and Gaussian uncertainties in the prediction. Using these experiments we derive the probability of observing a likelihood $\mathcal{L} = \prod_{i=1}^{4}[\mu_i^{n_i} \exp^{-\mu_i}/n_i!]$, where n_i and μ_i are the observed and predicted numbers of tags in the i-th jet bin, no larger than that of the data.

9. N. H. Kuiper, *Proceeedings of the Koninklijke Nederlandse Akademie van Wetenschappen*, ser. A, 28 (1962);
W. T. Eadie, D. Dryard, F. E. James, M. Roos, and B Sadoulet, *Statistical Methods in Experimental Physics* (American Elsevier, 1971).

10. D.Acosta *et al.*, Phys. Rev. **D65**,052007 (2002).

11. G.Apollinari*et al.*, hep-ex/0109020

12. H.E.Haber and G.L.Kane, Phys. Rep. **117C**,76 (1985);
C.Nappi, Phys. Rev. **D25**,84 (1982);
M. Carena *et al.*, Phys. Rev. Lett. **86**,4463 (2001)

13. A.Abachi *et al.*, Phys. Rev. Lett. **79**, 12003 (1997)

14. V.Savinov *et al.*, Phys. Rev. **D63**, 051101 (2001)