CDF/ANAL/HEAVYFLAVOR/CDFR/2090
Version 1.0

Efficiency and Background Studies Using the JETVTX b-Tagger

Richard Hughes, Phil Koehn, N. Mark Shaw,
Paul Tipton, Kirsten Tollefson, Gordon Watts, Brian Winer
with the B-tag Working Group

May 27, 1993

Abstract

In this note we measure the efficiency and background-rejection performance of the JETVTX tagging algorithm on samples of 6.1 inclusive electrons and inclusive jets. The electron sample is used to determine the b-tagging efficiency from data. The jets sample is used to determine the rate of mis-tags and heavy flavor tags produced in jets (presumably from $g \rightarrow b\bar{b}$ or $g \rightarrow c\bar{c}$). We have compared the measured tagging efficiencies in data to a Monte Carlo. The agreement is good within statistics. We predict the expected b-tags in a sample of good conversion electrons as a check of the mis-tag calculation. The agreement is good.

1 Introduction

This note is a follow-up to the first three sections of CDF1959. We will examine the efficiency and background-rejection performance of the JETVTX b-tagging algorithm. The algorithm is described in detail in CDF2068. Briefly, JETVTX attaches SVX tracks to JETS clusters, and tries to form a secondary vertex of the tracks. If the resulting vertex lies along the JET direction and is in front of the primary event vertex, it is called a "positive" tag. If the resulting vertex lies along the JET direction and is behind the primary event vertex, it is called a "negative" tag. Objects which have long lifetimes (such as bottom, charm mesons, K^0_S, etc.) will tend to be tagged as positives. Tracking mistakes and the finite resolution of the SVX will cause some JET objects to be tagged as negatives.

The JETVTX algorithm works with JET clusters with $E_T \geq 10.0 GeV$ and detector $\eta \leq 2.0$. In addition, the following is required:

- The coordinates of the primary event vertex come from the module VXPRIM.
- Two track tags consistent with K^0_S and Λ candidates are removed.
- The maximum number of shared SVX clusters for each track is 2.
- The minimum number of SVX hits per track is 3.
- The maximum impact parameter is 0.1 mm.
- The maximum track Z_0 relative to the primary vertex Z is $3.0 cm$.

- The maximum SVX χ^2 is 25.0.
- The maximum contribution each track can make to the secondary vertex fit χ^2 is 50.0.
- The minimum angle between the track direction and the JET axis is 36 degrees.
- Each tagged jet is required to have at least 2 SVX tracks attached to it.

We then define two JETVTX cut sets, called TIGHT and LOOSE. Each cut set has in common the above cuts. In addition, the TIGHT cut set requires:

- $P_t \geq 2.0\ GeV/c$
- $\frac{|D0|}{\sigma_D} \geq 3.0\ GeV/c$

The LOOSE cut set requires:

- $P_t \geq 1.5\ GeV/c$
- $\frac{|D0|}{\sigma_D} \geq 2.0\ GeV/c$

Final, work on the exact optimal method for removing K_s^0 decays within JETVTX is still in progress. We have also seen that two-cluster tracks can be selected which are of high quality if one requires SVX $\chi^2 < 8$ and no shared clusters. These improvements have not yet been included.

2 Tagging Efficiencies in the Inclusive Electron Sample

Based on work done by F. Ukegawa and A.B. Wicklund on 1988-89 data, the inclusive electron sample is thought to be rich in semileptonic b decay. This was confirmed with new data in CDF1959 and CDF2004, where the fraction of electrons coming from semileptonic B decay, F_b, has been measured to be $37 \pm 8\%$. In this section we use the inclusive electron sample to measure the efficiency of the JETVTX tagging algorithm. We use the the double-tag rate to give us the tagging efficiency of semileptonic b's in data in a manner which is independent (to first order) from the value of F_b. We then compare to a Monte Carlo of semileptonic $b\bar{b}$ events.

The inclusive electron sample is selected from the stream 2 inclusive electron stream. Our cuts are described in CDF1959, with the exception that we now allow the electron $E_T > 9GeV$ instead of $E_T > 12GeV$. We have approximately 70,000 good electron events in this sample.

This Monte-Carlo has been generated with Isalep (nevolve=100 nhadron=1) by Guillaume Unal and the Penn Group. The P_t of the b or the $\bar{b}$ was required to be above 15 GeV. No decay was forced in Isajet. All the B mesons are decaying using CLEOMC. Only events with an electron $P_t > 8GeV$ and $|\eta| < 1.5$ or a muon, $P_t > 8GeV$ and $|\eta| < 1.0$ are kept for simulation with CDFSIM. In the simulation, the vertex is smeared by 30 cm in z and 35 microns in x and y .Figure 1 shows the E_T, distribution and SVX track multiplicity for data and Monte Carlo. The agreement is good but not perfect. We know that the SVX tracking efficiency is too high in the Monte Carlo.

In the double-tag method, we look at a sample of events in which we have an electron passing all cuts, and an "away" tagged jet. The number of events in this sample is then our

Algorithm	With tags	With - Fakes	Away tags	Away - Fakes	With + Away Tags
JETVTX Tight	559	531 ± 24	97	74 ± 10	12
MC Tight	329	329	60	60	10
JETVTX Loose	976	883 ± 31	186	123 ± 14	29
MC Loose	501	501	94	94	25

Table 1: Number of tags seen in the inclusive electron sample. "With" and "Away" are described in the text.

Algorithm	ϵ_{double}^{data}	ϵ_{double}^{mc}
JETVTX Tight	$16 \pm 5\%$	$17 \pm 5\%$
JETVTX Loose	$24 \pm 5\%$	$27 \pm 5\%$

Table 2: Comparison of tagging rates in data *vs* Monte Carlo for the double-tag method. No SVX track requirement is made on the electron jet. The Z vertex of the electron was required to be less than 30. cm.

denominator. The numerator will be the subset of events in which the electron is also tagged (i.e. "with" plus "away" tags):

$$\epsilon_{double} = \frac{\#\ of\ events\ with\ both\ \text{“}with\text{''}\ and\ \text{“}away\text{''}\ tags}{\#\ of\ events\ with\ \text{“}away\text{''}\ tags}$$

The number of "away" and "with plus away" tags observed in the data with the b-tagging algorithms is shown in table 1.

The number of tags observed with the JETVTX algorithm is listed in table 1. These tags include real tags from heavy flavor and fake tags. To extract an actual b-tag efficiency, ϵ_{double} should be corrected for fake tags present in the sample. The fake subtraction for JETVTX is performed as described in the next section. The mistag contamination to the electron tags are listed in table 1.

The tagging efficiencies extracted from this method is shown in table 2, where we also show results from the $b\bar{b}$ Monte Carlo sample. The data efficiency for the double tag method agrees with Monte Carlo. It is not surprising that data and Monte Carlo agree within statistics (a 25% error) even though we know the average SVX tracking in data is at least 10% lower in data when compared to Monte Carlo.

Figure 2 shows the "with" tags vs. the electron E_T and SVX track multiplicity, for tight and loose cuts. This shows that the combination of the b fraction times the tagging efficiency is relatively flat as a function of the electron E_T. Figure 3 shows the same quantities for the $b\bar{b}$ Monte Carlo.

We plan in the near future to check this efficiency measurement with a measurement using the single tag rate and a new measurement of the b fraction, F_b.

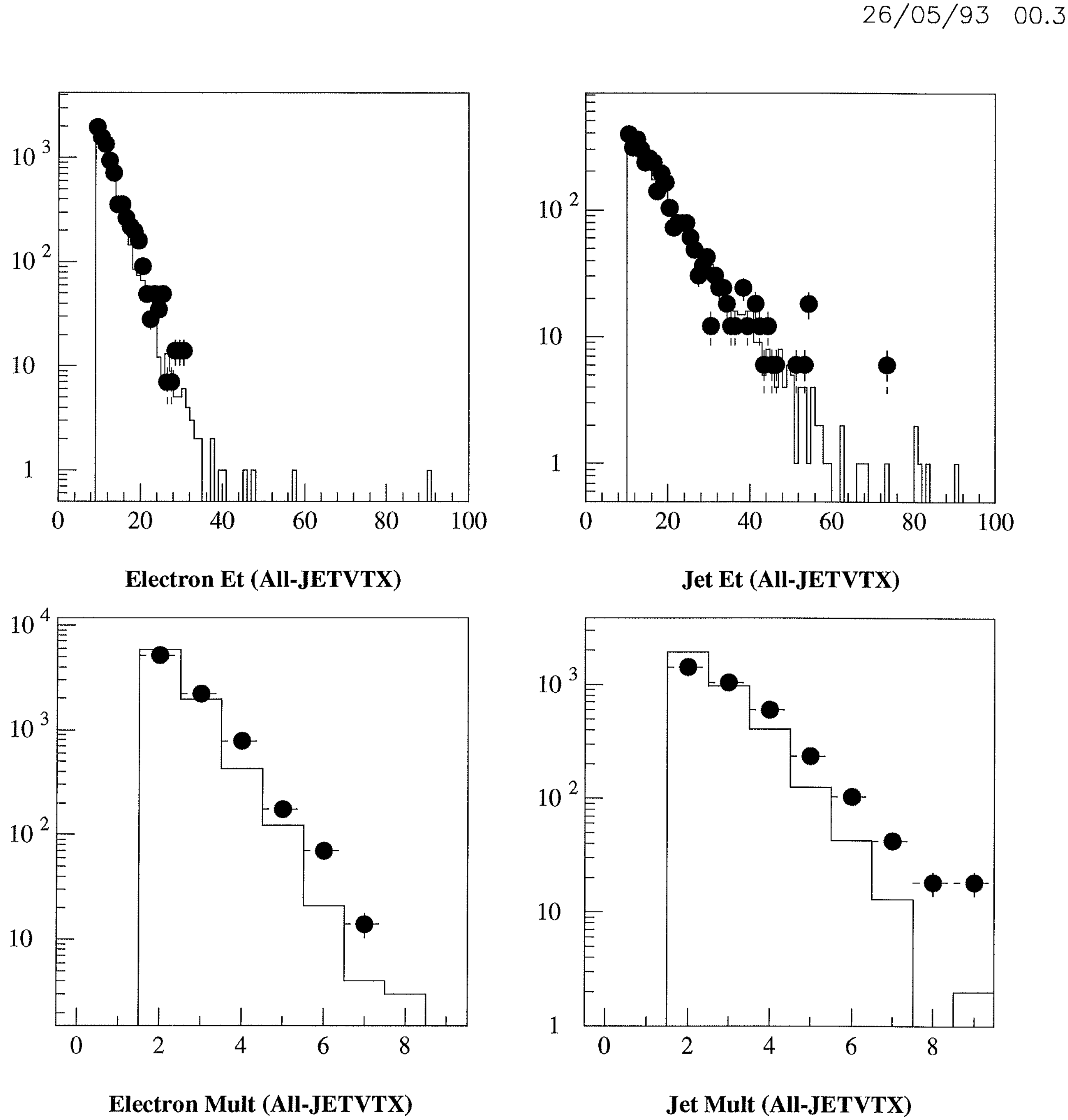

Figure 1: A) The E_T distribution for the inclusive electron sample (histogram) compared with the Monte Carlo (points) E_t distribution. B and C) The SVX track multiplicity for the electrons and all jets compared to the Monte Carlo of the same.

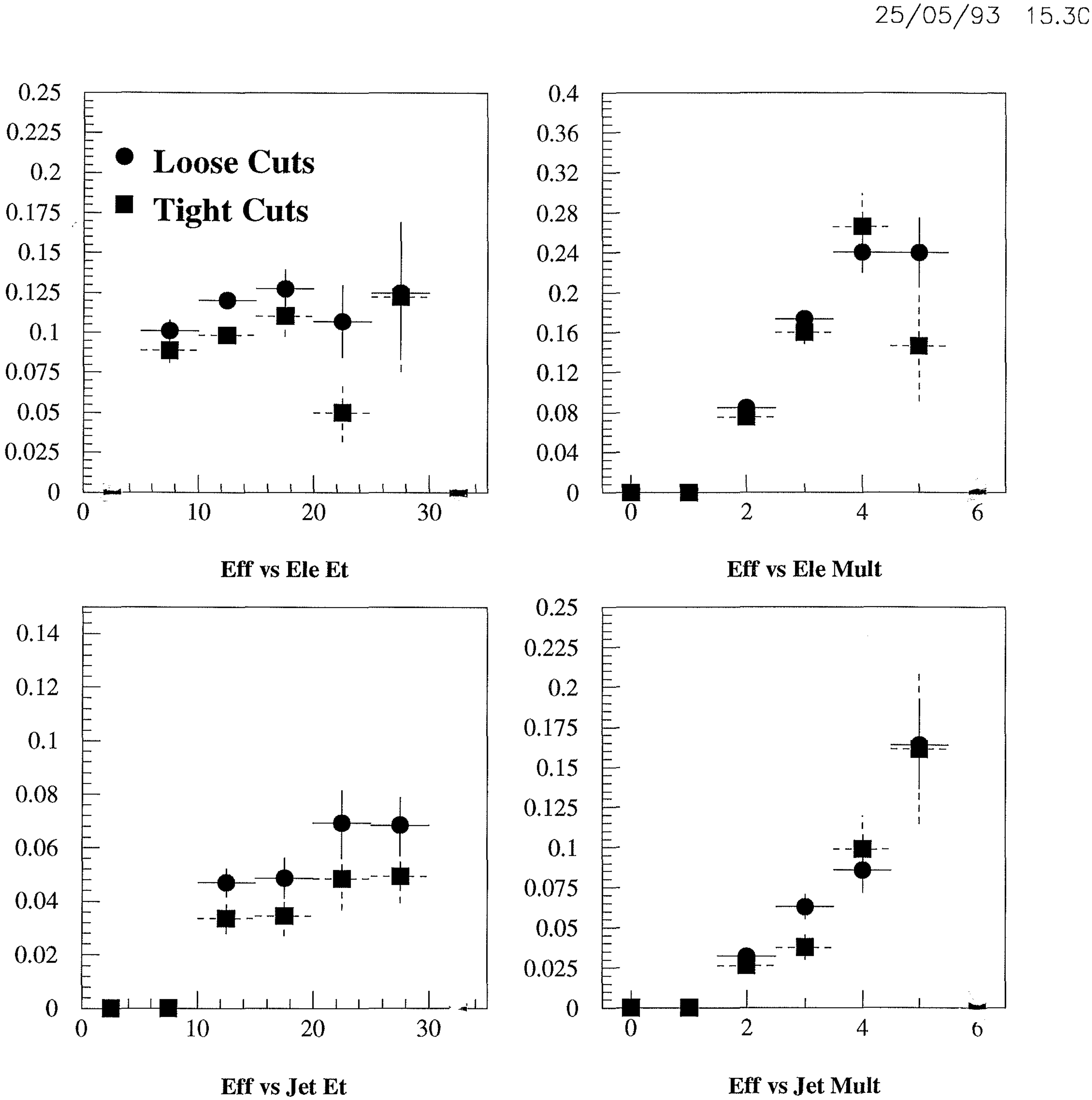

Figure 2: The electron tagging rate as a function of electron E_T and multiplicity for the loose and tight JETVTX tags of the electron jet and the "away" jet. No correction is made for F_b as a function of E_T.

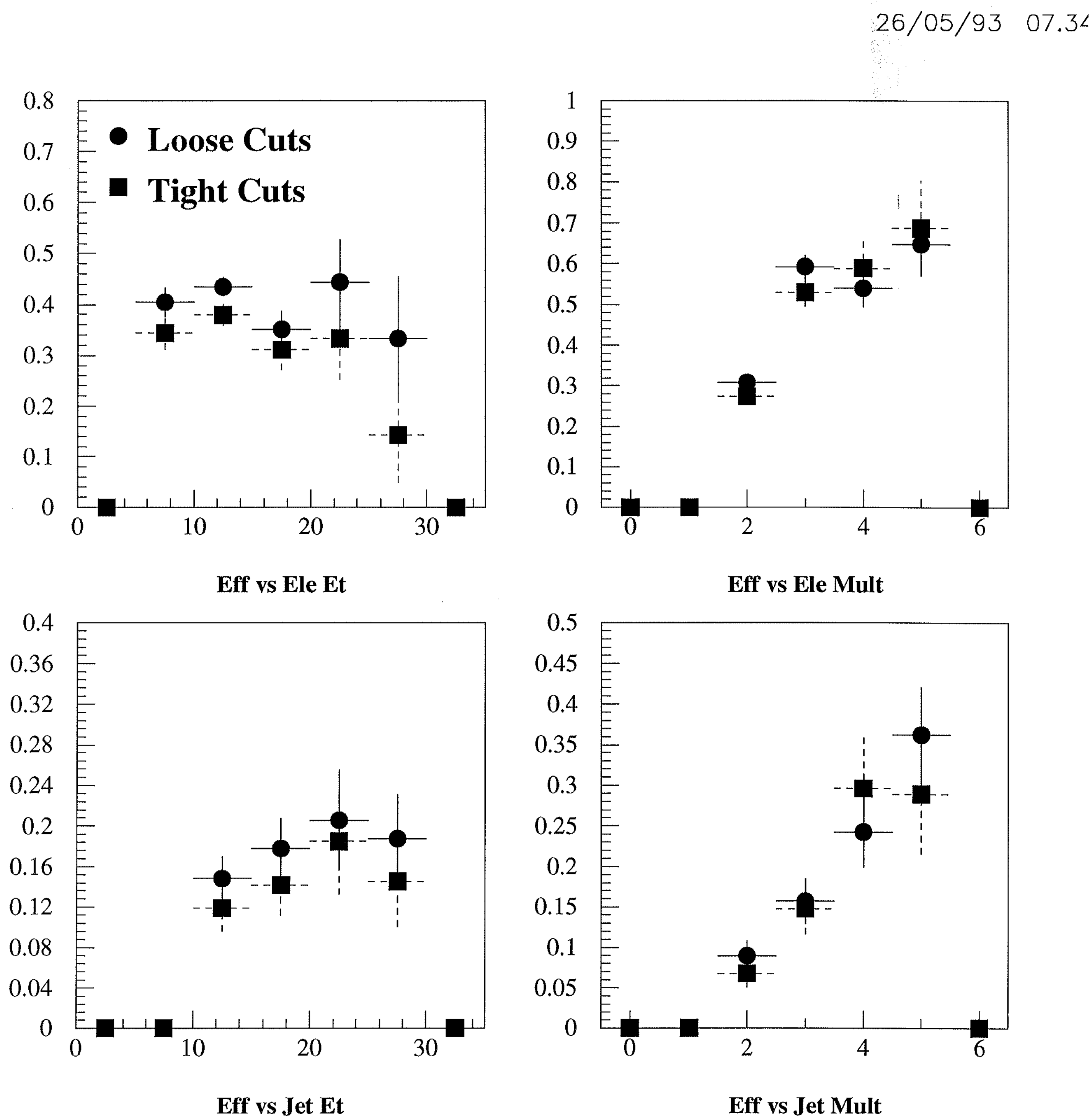

Figure 3: Monte Carlo of the electron tagging rate as a function of electron E_T and multiplicity for the loose and tight JETVTX tags of the electron jet and the "away" jet.

3 Tagging Efficiencies (Mistag rates) in the Jet Sample

To test the ability of the tagging algorithms to reject gluon and light quark jets we have used the stream 1 jet triggers J1Q1, J2Q1 and J3Q1. These triggers require one jet with an E_T greater than $100GeV$, $70GeV$, and $50GeV$, respectively. This jet sample should be dominated by non bottom/charm jets. The rate of jets containing charm or bottom may not be small, however, and at some point our tagging algorithms might start to see contamination from these sources.

First, we define the sample of jets that are "taggable" as a jet that passes the jet E_T requirement and the SVX track requirements of JETVTX. We then apply the tagging algorithms to the events which contain "taggable" jets. In CDF 1959 we found the mistag rate (the rate of tagged generic jets) as a function of Jet Et and SVX track multiplicity. Fig. 4 shows a strong correlation between Et and Jet Multiplicity, using loose cuts. Fig. 5 shows that efficiency(mistagged jets/all jets) for Et > 10.0 and track mult≥ 2, is also a function of detector η. Because of strong functional dependencies, the tagging efficiency for the inclusive jets is parametrized in terms of a two dimensional mistag matrix; Et vs Multiplicity. In addition, the jets are also characterized in the regions of detector $|\eta| < 1.0$ and ≥ 1.0. To determine the predicted number of mistags in an independent sample, we multiply the Et vs Multiplicity matrix for all taggable jets in the sample by the mistag matrix. These predictions are then compared to the observed number of tags in terms of Et and Multiplicity. This is done for both loose and tight JETVTX cuts. Furthermore, we split the tags into positive and negative decay lengths, so each tag type may be examined individually, and their relative rates measured.

In figure 6, we show the dependence of the negative tag rate as a function of E_T and SVX multiplicity In figure 7, we show the same information for jet eta greater than 1.0.

In figure 8, we plot the 2d decay length for the tight JETVTX tags. In figure 9, we plot the 2d decay length for the loose JETVTX tags. The resultant 2D decay length distribution shows a clear enhancement in positive decay lengths.

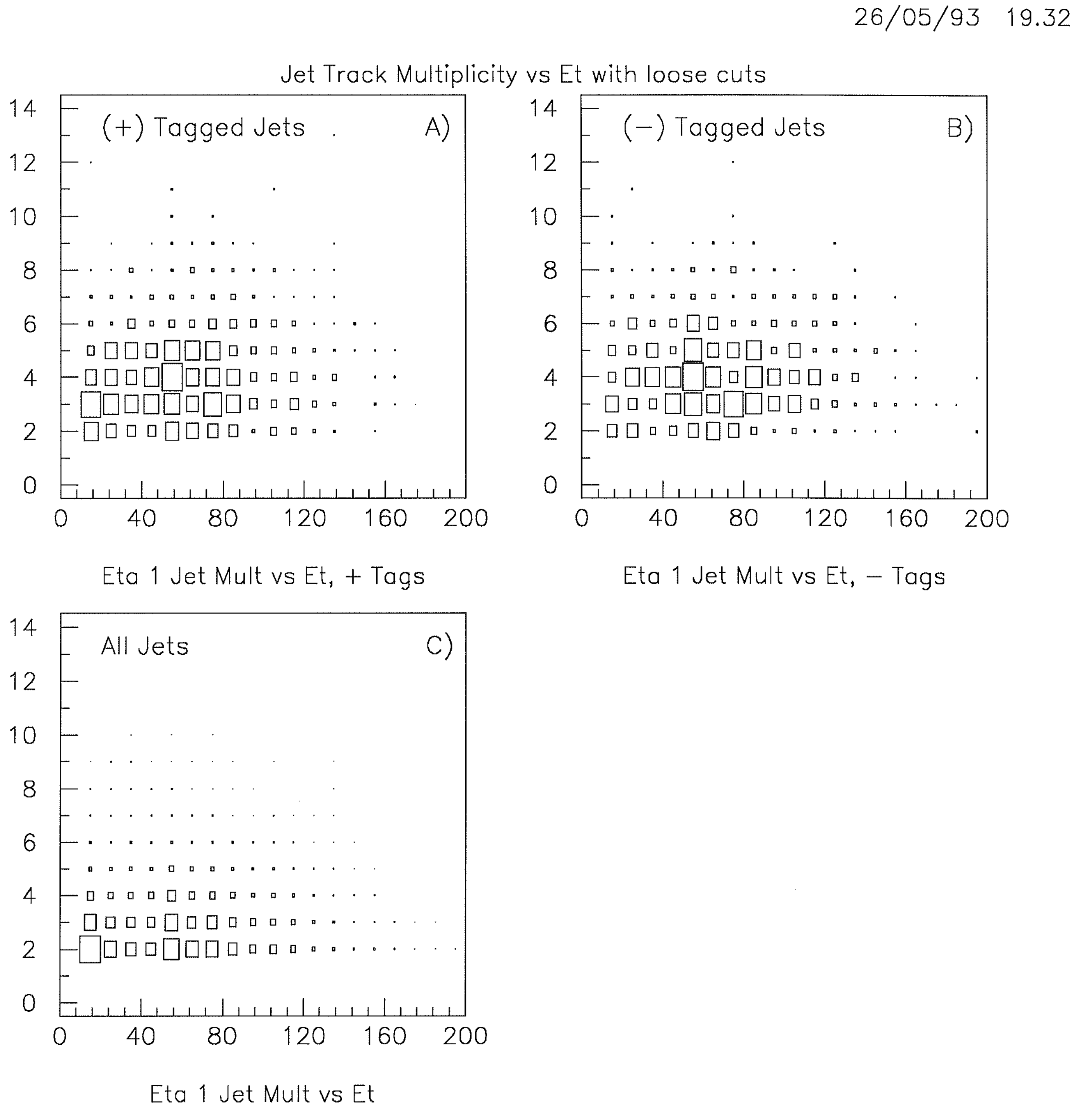

Figure 4: E_T **vs Mult for the Inclusive Jet Sample.** Plots A) and B) are scatter plots of E_t vs Jet multiplicity for positive and negative decay length tags. While C) is the same plot for all *taggable* jets. All plots are for $|\eta| < 1.0$ and loose JETVTX cuts.

26/05/93 19.21

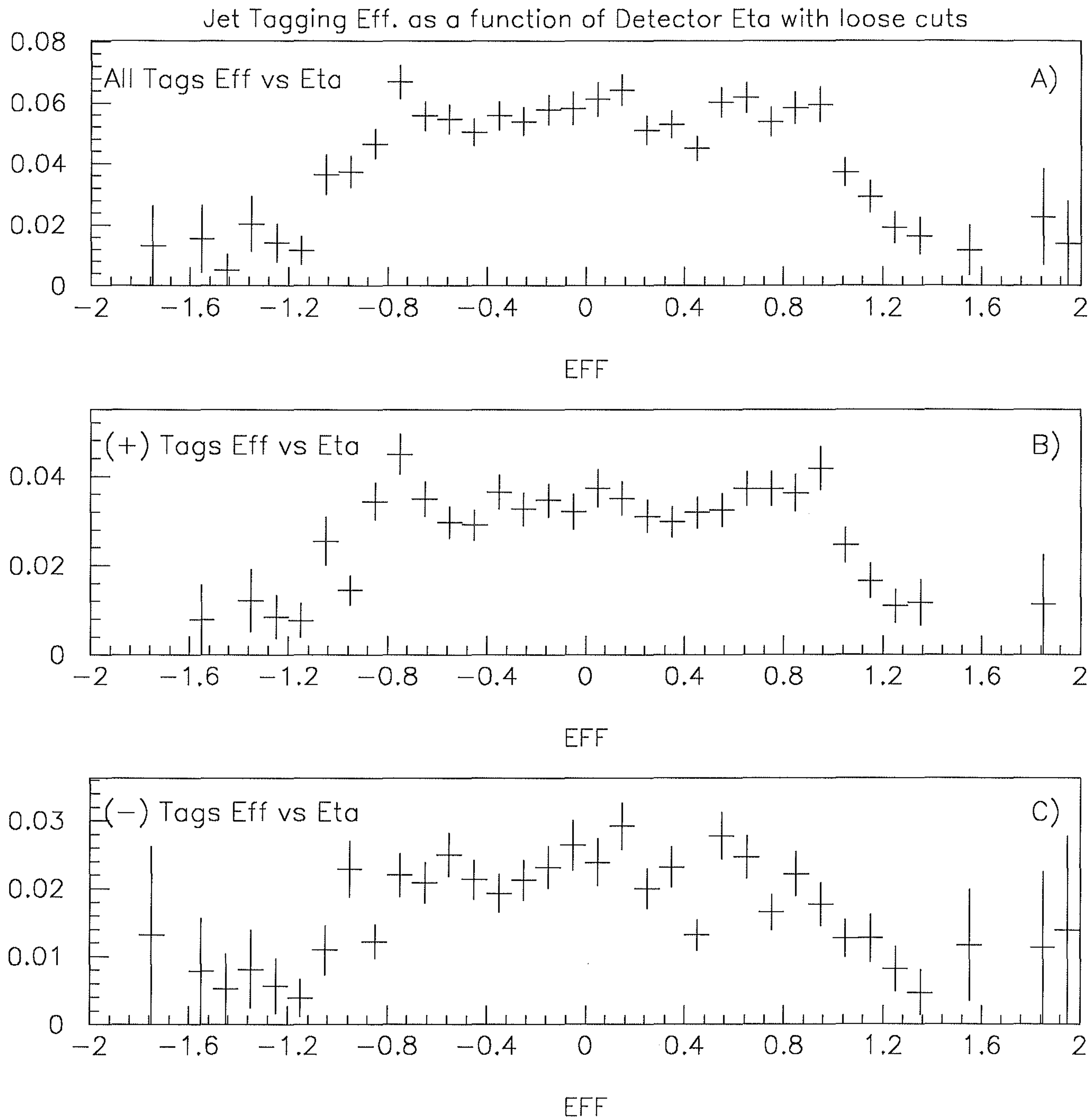

Figure 5: **Jet Tagging Efficiency as a function of detector η.** Plot A) contains *all* jet tags vs detector η, plots B) and C) contain positive and negative decay length tags respectively. All plots are for loose JETVTX cuts.

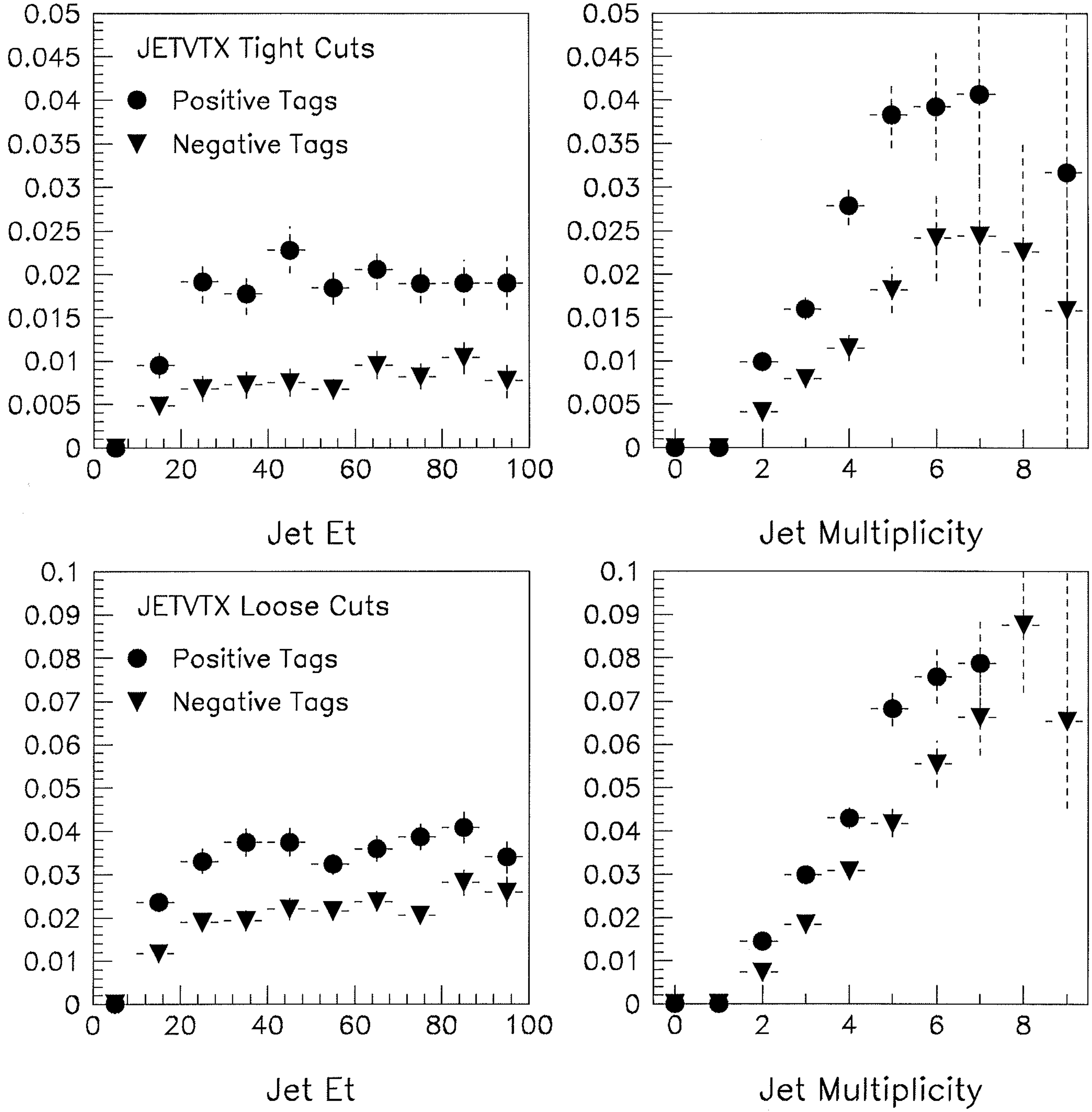

Figure 6: The dependence of the JETVTX tag rate on Jet E_T, SVX track multiplicity for positive and negative decay lengths and tight and loose JETVTX cuts. These plots are for jet eta less than 1.0

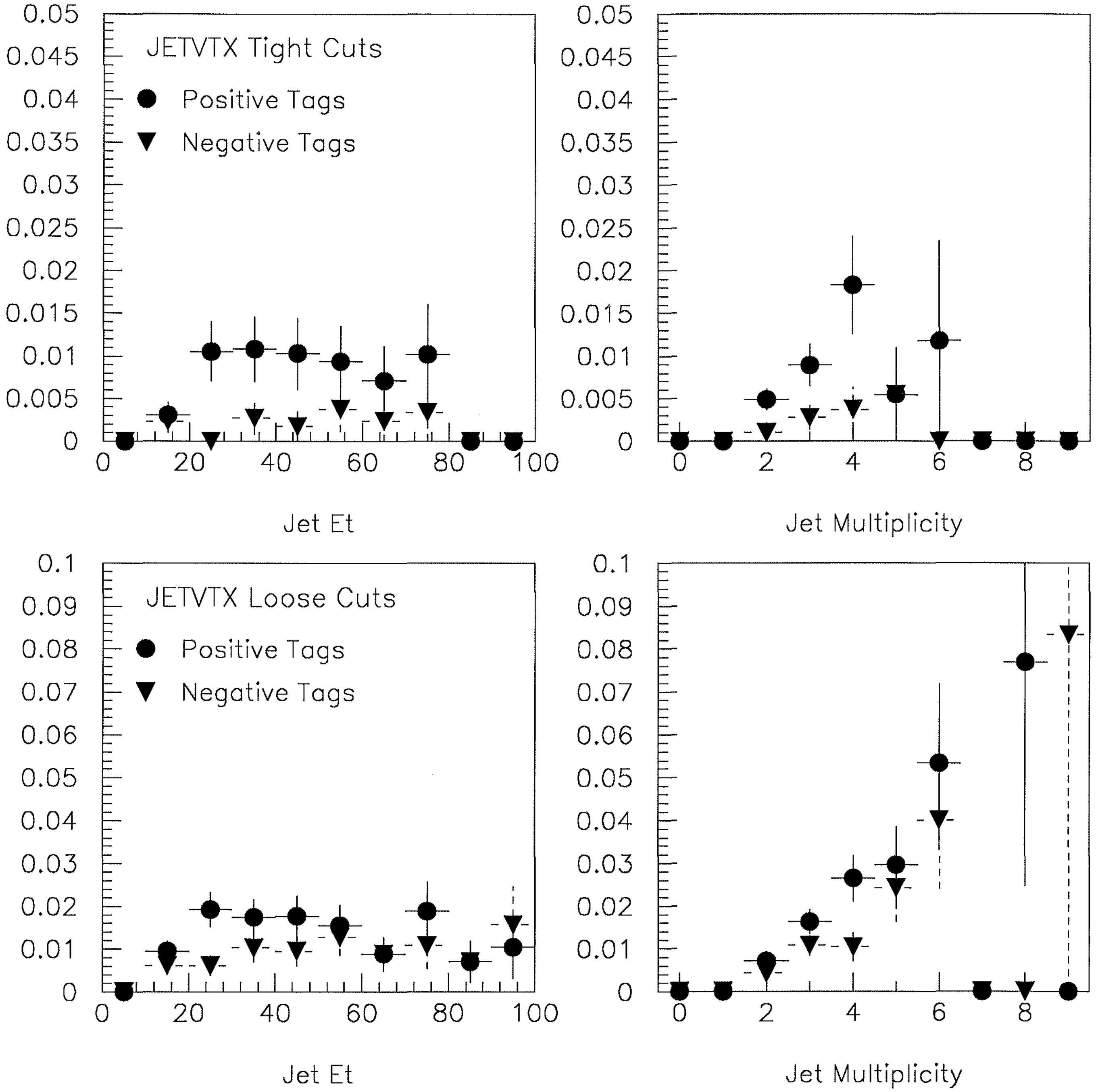

Figure 7: The dependence of the JETVTX tag rate on Jet E_T, SVX track multiplicity for positive and negative decay lengths and tight and loose JETVTX cuts. These plots are for jet eta greater than 1.0

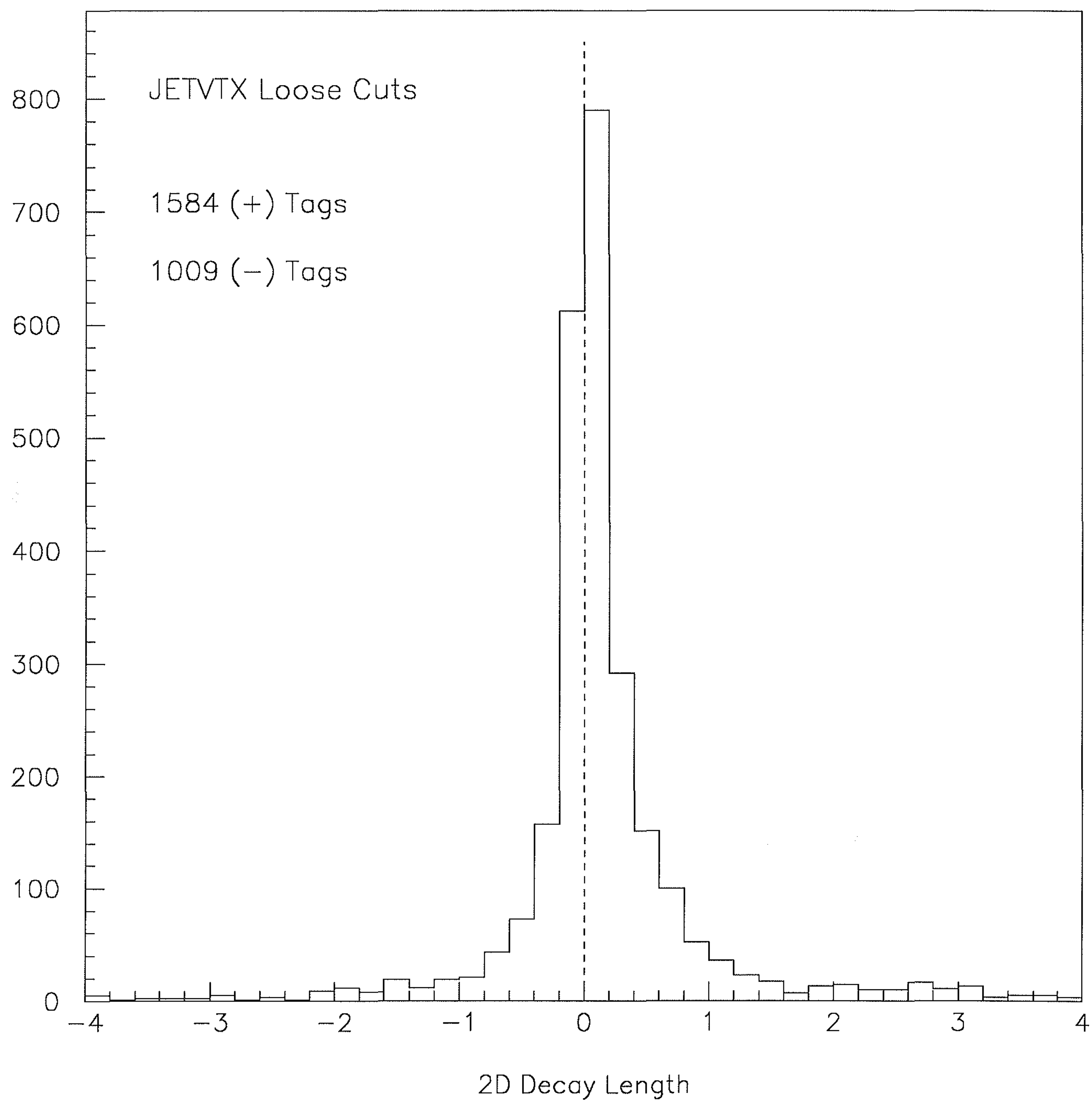

Figure 8: The 2D decay length for the loose JETVTX tagged jets.

26/05/93 10.47

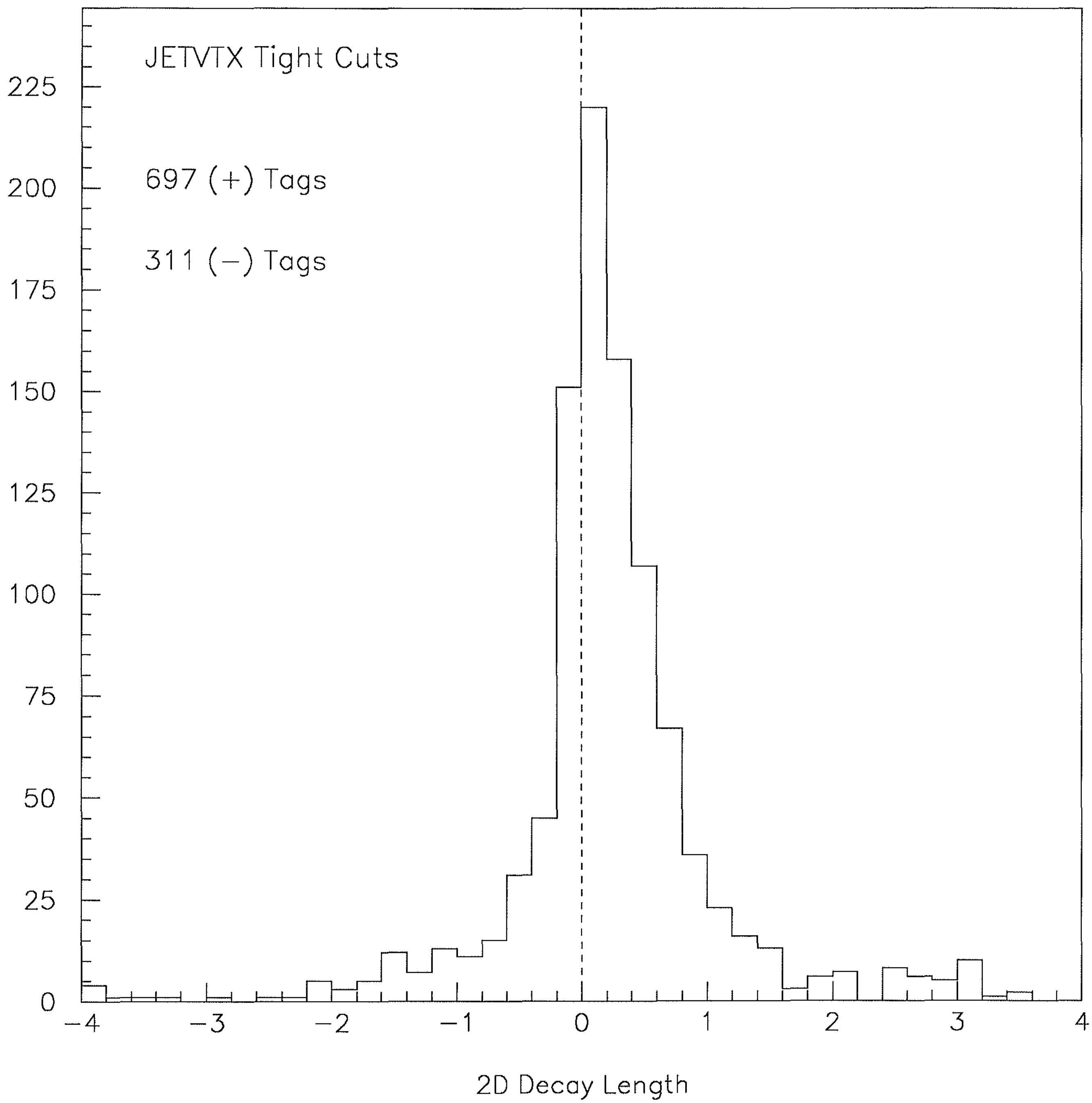

Figure 9: The 2D decay length for the tight JETVTX tagged jets.

4 Using Mistags in Inclusive Jets to Predict Mistags in Other Data Samples

We have tested the mistag calculation in two samples, the first being the inclusive jet sample, the second being the "good conversion" sample. The first test of the mistag rate calculation was done by separating the inclusive jet sample into two independent parts. The first part consisted of $\sim$ 50k jets and was used to form the mistag matrix, the second part, $\sim$ 25k jets and was used to find the observed mistags. Fig. 10 shows excellent agreement between the observed and the predicted number of positive and negative decay length mistags as functions of Et and Multiplicity, using loose cuts.

As a second test of the mistag rate calculation, we look at sample of tagged *good* conversion electrons, using loose cuts. A *good* conversion passes the following cuts:

1. $|\Delta \cot(\theta)| < 0.06$
2. $|\Delta(sep)| < 0.3$
3. $M_{ee} < 500 MeV/c^2$
4. $R_c < 50$ cm

When we exclude the conversion jet, we essentially have a generic jet sample. Fig. 11 is a similar plot for tagged "good" electrons. Again, the number of mistags as determined from the full inclusive jet sample matrix agrees very well with the observed number of mistags seen in the conversions.

5 Conclusions

We have shown the JETVTX algorithm is $16 \pm 5\%(24 \pm 5\%)$ efficient on taggable semileptonic B decays with a tight (loose) cutset. These agree well with Monte Carlo predictions of the same. We have measured the background rate with a generic jet sample and parameterized it in a matrix as a function of jet E_T, SVX track multiplicity, and jet $|\eta|$. Checks of the background calculation technique in an independent jet sample and a "golden conversion" sample show good agreement with the observed number of JETVTX tags.

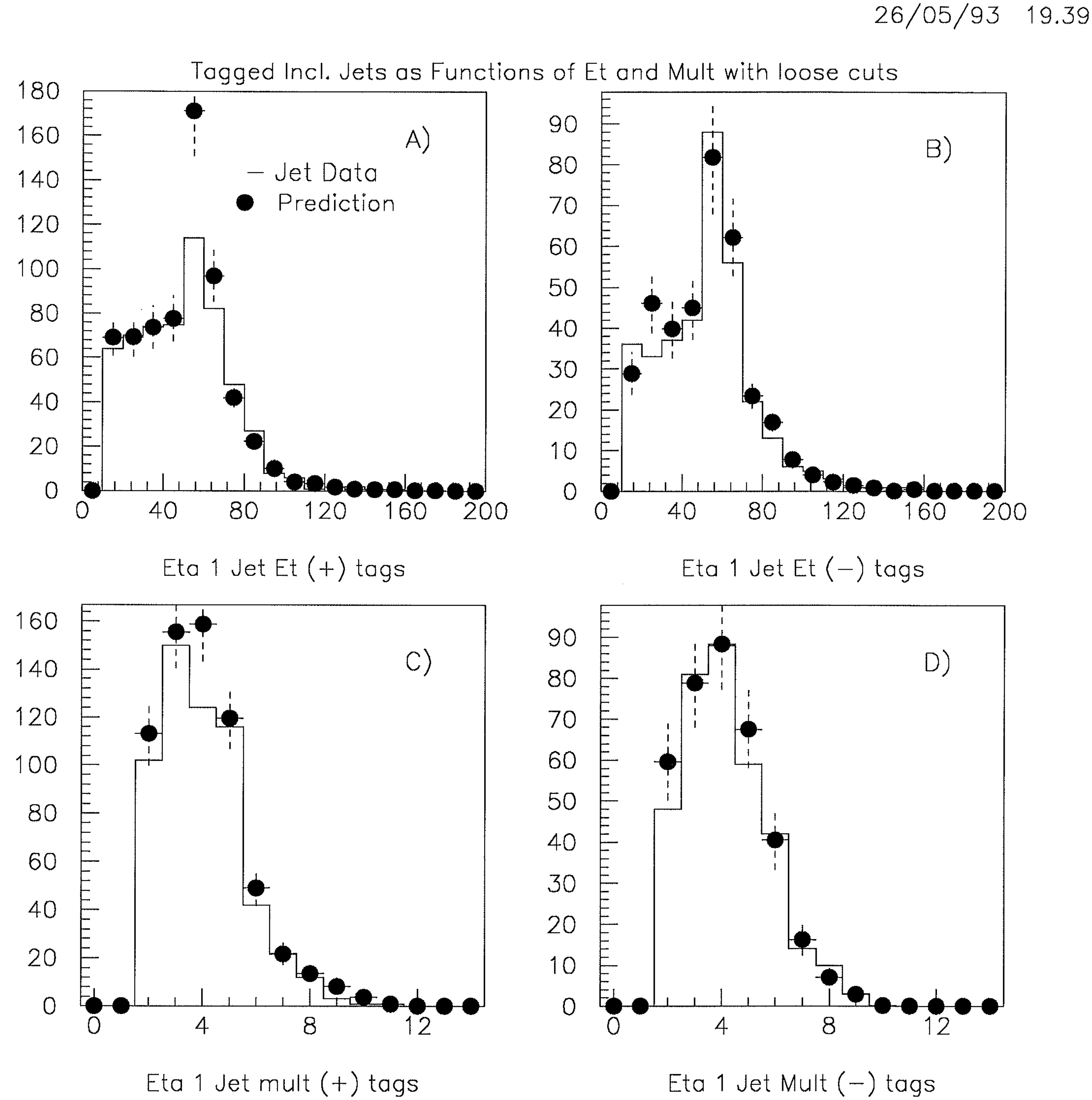

Figure 10: **Tagged Jets as functions of E_t and Multiplicity** Plot A) compares the positive decay length tagged jet *data* with the *predicted* number of mistags as a function of E_t. Plot B) makes the same comparison for negative tags. Similarly, plots C) and D) make the comparison as a function of jet track multiplicity. All plots are for $|\eta| < 1.0$ and loose JETVTX cuts.

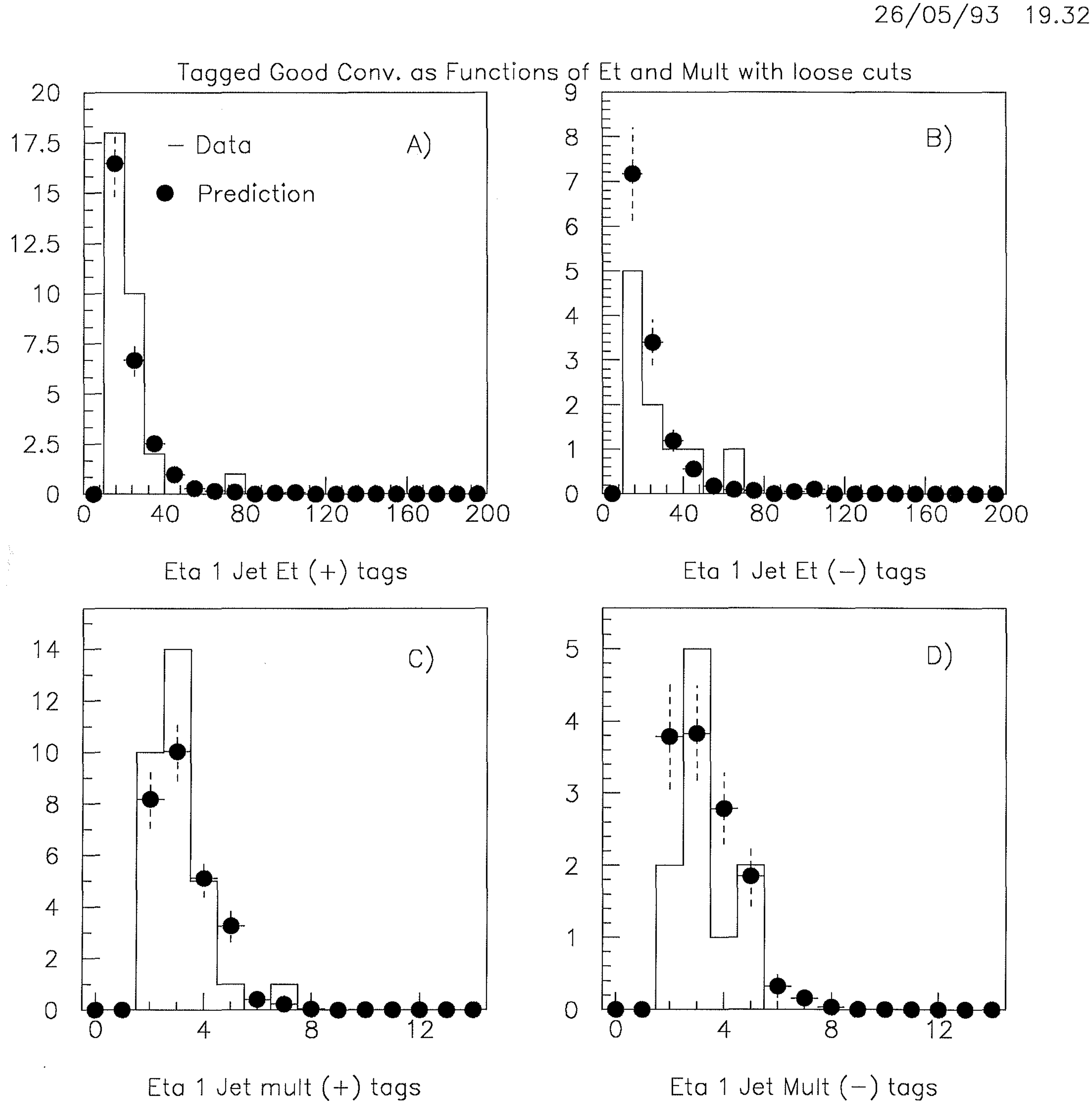

Figure 11: **Tagged Good Conversions as functions of E_t and Multiplicity** Plot A) compares the number of positive decay length tagged jets in an event that has at least one good conversion, with the *predicted* number of mistags as a function of E_t. Note: the conversion jet is excluded from the event. Plot B) makes the same comparison for negative tags. Similarly, plots C) and D) make the comparison as a function of jet track multiplicity. All plots are for $|\eta| < 1.0$ and loose JETVTX cuts.