CDF/ANAL/TOP/CDFR/2111
June 11, 1993
Version 1.1

Study of improved track selection for b tagging with SVX

W.-M. Yao, L. Galtieri, R.Ely, J.Lys, M. Austern
Lawrence Berkeley Laboratory
with the B-tag group

Abstract

We report on a study of the improved track selection for b tagging used in the SECVTX algorithm used in CDF-2072. A comparison between JETVTX and SECVTX is given. Clearly the new track criteria developed for SECVTX improve the signal to background rates for b tagging in jets. The rejection of K_S not originating from B decays is also discussed. A routine, BTKSEL, has been provided to apply the tracking selection and the K_S removal discussed here. It has become the standard for the B-tag group.

1 Introduction

Currently there are two b-tagging algorithms based on finding secondary vertices in jets using CTVMFT. One is JETVTX developed by Rich Hughes[1] , the other is SECVTX recently developed by Weiming Yao[2]. The main difference between the two is due to different criteria for selecting tracks used by the vertexing routine and the refitting of the tracks with clusters shared by more than one track. The clusters get appropriate weights depending on the cluster length and the new tracks obtained after the fit are also input to CTVMFT. Here we investigate the effects of the different track selection.

In section 2 we discuss the comparison of S^2/B obtained for the two algorithms in the jet data sample, in the top Monte Carlo and in the inclusive electron data sample. In section 3 we discuss the algorithm for eliminating secondary vertices due to K_S and Λ not coming from B decays. In section 4, we describe a routine BTKSEL, which can be used to apply the track selection discussed in this note. Section 5 is a summary of the results.

Criteria	JETVTX	SECVTX
χ^2 of SVX track	$\chi^2 < 25$	$\chi^2/DF < 6.0$
Good cluster	not shared	not shared, no dead strips, cluster length < 4 strips
Clusters/track	≥ 3	≥ 2
Good clusters/track	≥ 2	≥ 2
Track P_T, LOOSE cuts Track P_T, TIGHT cuts	> 1.5 > 2.0	> 1.5 > 2.0
$\frac{d0}{\sigma_{d0}}$, LOOSE $\frac{d0}{\sigma_{d0}}$, TIGHT	> 2 > 3	> 2.5 > 3.0
impact parameter (d0)	< 0.1 cm	< 0.15 cm
Refit of shared clusters Weight for shared clusters	yes 0.1 cm	yes $\frac{n \cdot pitch}{\sqrt{12}}$
Positive tag	$\frac{L_{xy}}{\sigma_{xy}} > 0$	$\frac{L_{xy}}{\sigma_{xy}} > 3$
mis-tag	$\frac{L_{xy}}{\sigma_{xy}} < 0$	$\frac{L_{xy}}{\sigma_{xy}} < -3$

Table 1: Track criteria for the JETVTX and SECVTX modules

2 Comparison between JETVTX and SECVTX

Both algorithms look for displaced tracks in a cone of radius R = 0.6 around the jet direction, as found in the JETS bank. Selection criteria are applied to the tracks to select those which are displaced from the primary vertex. The routine CTVMFT is then used to search for secondary vertices among the displaced tracks. In all cases the module VXPRIM is used to find the position of the primary vertex.

The cuts which are different in these two taggers are shown in Table 1. In this note we use "loose cut" for the SECVTX module. In fact the "loose cuts" and "tight cuts" for the SECVTX give the same S^2/B on the top search, where the S is the efficiency for tagging TOP-140 and B is the tagging rate for the positive side of the generic jet sample (ref: Weiming's talk at the b-tag meeting of 6/2/1993).

For the comparison of the two algorithms the K_S and Λ are not removed, but the effect of this is about 10% (see the section 3).

We now discuss details of the results obtained with these two modules.

2.1 Jet sample

We use Hughes' jet sample (JET2Q, JET3Q) as control sample and run the SECVTX and JETVTX modules on it. To obtain results with JETVTX we have used the JETVTX-tight.uic to make sure we use exactly the same criteria as other members of the b-tag group. The comparison of the results obtained with the two modules is based on the distributions shown in Figure 1. This shows the $\frac{L_{xy}}{\sigma_{xy}}$ distributions obtained with the two different modules. We call the events in the positive side "NP"

Module	NP	NN(B)	S	S^2/B
SECVTX	889	289	600.	1246
JETVTX($\frac{L_{xy}}{\sigma_{xy}} > 0$)	562	226	336.	500
JETVTX($\frac{L_{xy}}{\sigma_{xy}} > 3$)	483	170	313.	576

Table 2: Comparison of SECVTX and JETVTX on the jet data

module	Top140:S	eff.	Top120-nolife:B	S^2/B
SECVTX	90	0.39 ± 0.04	11	736
JETVTX	73	0.31 ± 0.04	14	380

Table 3: Comparison on the top Monte Carlo

and those in the negative side as "NN". The difference is assumed to be the signal, i.e., the heavy flavor contribution: S = NP - NN. We then calculate the significance of the signal, S^2/B, where B = NN. The results are shown in Table 2.

Clearly the SECVTX module is much more efficient (about a factor of 2) for tagging heavy flavor with little increase in the background.

2.2 Top-140 Monte Carlo

Next, to check if the SECVTX criteria are indeed more efficient than those of JETVTX for top search, we used the JETVTX-tight.uic and standard SECVTX cuts to make a similar comparison on the Top-140 Monte Carlo events as a signal and on the Top-120-nolife (i.e. for the case that the secondaries from B decays come from the primary vertex) for the background. The results are shown in Table 3.

The statistics for this sample is not very large, but again SECVTX is the more efficient module.

2.3 Inclusive electron sample

We have run JETVTX and SECVTX on the electron sample provided by Brian Winer, which includes V6.1 inclusive electron data from STREAM1.

For this sample the total energy in the event is rather small, since the jet containing the electron is relative soft (mean at 19 GeV, see Figure 2) and the electron took about half of the total energy. We expect the environment of such events to be much cleaner than that of the generic high E_T jets considered earlier, therefore the effect of the track quality cuts used in SECVTX may be smaller. Here the loose cuts are used for JETVTX.

Figure 3 shows the distribution of $\frac{L_{xy}}{\sigma_{xy}}$ for the tagged electron jets obtained by the two modules after removal of events which contained a conversion electron. The

module	NP	NN(B)	S	S^2/B
SECVTX	989	61	928	14117
JETVTX($\frac{L_{xy}}{\sigma_{xy}} > 0$	876	92	784.	6681
JETVTX($\frac{L_{xy}}{\sigma_{xy}} > 3$	798	49	749.	11449

Table 4: Comparison of SECVTX and JETVTX for the inclusive electron sample

number of events in the positive and negative side are shown in Table 4.

3 K_S and Λ removal

Secondary vertices in jets are due to decays of charm and bottom particles, however they can also be due to K_S and Λ. These particles have a much longer mean life, but a small fraction of them will decay close to the primary vertex. Since they are copiuously produced, they could give a large contribution to the b-tag signal. Ideally, one would like to remove these long leaved particles if they are not associated with heavy flavor production and retain those that come from heavy flavor decays. In order to maximize the efficiency for heavy flavor tagging while reducing the backgrond, we remove them according to the criteria which we discuss in this section. We illustrate the method with the K_S case.

First, we look for K_S in generic jets, the JET2Q and JET3Q samples (70 and 50 GeV trigger respectively). Tracks of opposite sign and with at least one track with $P_T > 1.5$ GeV are paired and fitted by CTVMFT. The resulting $\pi\pi$ invariant mass is shown in Figure 4a. A clear peak is seen with a width of 6 MeV. In Figure 4b we plot the impact parameter significance $\frac{d0_{K_S}}{\sigma}$, for the events in the peak region ($|M_{K_S} - M(\pi\pi)| < 20$ MeV and the side bands (M($\pi\pi$)=0.44-0.46 and M($\pi\pi$)=0.54-0.56). The K_S signal region exhibits a clear peak with a resolution of $\sigma = 2.0$. This sample of K_S appears to be a prompt signal, i.e., the K_S come from the primary vertex.

Next, we study the K_S associated with the secondary vertices. First, we plot the invariant mass $M(\pi\pi)$ for all the two track pairs which are part of the daughters of the tagged secondary vertex. This is shown in Figure 5a and it shows a clear peak at the K_S mass. To find the additional K_S, which have only one track in the tagged vertex, we pair up each track which was a daughter of the tagged vertex with any track in the event wich has opposite sign and is displaced from the primary vertex. The impact parameter significance for the second track, $\frac{d0}{\sigma_{d0}}$, is required to be > 2.5, and no P_T cut is used. Each pair is then fitted by CTVMFT with the $\pi\pi$ mass or $p\pi$ mass assignments. Figure 5b shows the $\pi\pi$ mass distributions obtained with these criteria. Note that the track pairs shown in Figure 5a, are included in this plot. Again, a peak is observed, with a worse resolution than the one in Figure 5a, as expected as the second π can have a momentum as low as 200 MeV/c.

If the pair has a K_S mass ($\pm$20 MeV) or Λ mass ($\pm$ 6 MeV), the 2D decay length,

L_{xy}, and the impact parameter, $d0_{K_S}$, are calculated for this pair. The pair is removed if $\frac{L_{xy}}{\sigma_{lxy}} > 5.$ and $\frac{d0_{K_S}}{\sigma_{d0}} < 6.0$, i.e., the K_S or Λ points to the primary vertex. Figure 6a shows the $\pi\pi$ mass distribution for the prompt pairs and Figure 6b shows the distribution for the pairs with $\frac{d0_{K_S}}{\sigma_{d0}} > 6$. There is no significant K_S signal in the latter plot.

In this sample, about 10% of the secondary vertices include a K_S, of which 50% have a K_S pair formed among daughters of the tagged vertex (Figure 5a), and 50% have only one track among the daughters of the tagged vertex. Clearly, these values depend upon the P_T distribution of K_S.

With these criteria, as used by SECVTX, in the TOP-140 sample there is a loss in efficiency of $< 2\%$. This is much smaller than the 17% loss on the top-160 quoted by another analysis (Brian Winer talk, top meeting, 5/27/93). This high value was obtained by making only a cut on the $\pi\pi$ mass of ± 20 MeV. The loss clearly depends on how clean $\pi\pi$ mass plot looks. By making the requirements discussed above, SECVTX reduces the loss in efficiency considerably.

4 BTKSEL

Weiming has made the track selection routine, BTKSEL.CDF, available to everybody (see Appendix).

Guillame Unal and Dave Gerdes have already used this routine and report improvements in their analyses which use the JETVTX[3] and the JPBTAG[4] algorithms. This routine is now the official track selection and K_S removal routine adopted by the b-tag group.

5 Conclusion

We have improved S^2/B by a factor 2 compared to the JETVTX cuts. We have developped an algorithm for K_S and Λ removal from the tagged vertex tracks, which appears to be efficient without much loss in tagging efficiency.

References

[1] R. Hughes et al., CDF note: CDF/ANAL/TOP/CDFR/2068
R. Hughes et a., CDF note: CDF/ANAL/TOP/CDFR/2090

[2] . M. Yao et al., CDF note: CDF/ANAL/TOP/CDFR/2072

[3] . Benton et al, CDF note CDF/ANAL/TOP/CDFR/2003

[4] .B. Kim et al, CDF note CDF/ANAL/TOP/CDFR/2091

Figure Captions

Fig. 1. Decay length significance, $\frac{L_{xy}}{\sigma_{xy}}$, as obtained by the SECVTX and JETVTX modules in the generic jets sample.

Fig. 2. Electron E_T distribution for the inclusive eletron sample.

Fig. 3. The $\frac{L_{xy}}{\sigma_{xy}}$ distribution in the inclusive electron sample.

Fig. 4. K_S signal in the inclusive jet sample. a) the $M(\pi\pi)$ distribution for track pairs in the jets, b) the $\frac{d0_{K_S}}{\sigma_{d0}}$ for the events in the K_S peak and sidebands.

Fig. 5. High E_T jet sample. a) Invariant mass, $M(\pi\pi)$ for tracks which are daughters of the secondary vertex. b) same as a), where one track is a daughter of the tagged vertex, the other one is a displaced track anywhere in the event.

Fig. 6. Same as Fig. 5b. a) prompt K_S, b) displaced K_S

- APPENDIX A.

```
      SUBROUTINE btksel(num, vexnew, vexerr, lgood, nhit, lvee, ipair)
c  Purpose of this routine:
c                         to select the standard svx tracks for b tagger
c                         also, veto primary ks and lambda daughters.
   Defination of good track:
           CTC:
c               1) at least 2 axial SL with at least 4 hits
c               2) at least 2 stero SL with at least 2 hits
c          SVX:
c               1) at least 2 good clusters on the track
c               2) svxchi/ndof < 6. ( with ctc scale factor = 2.0 )
c          Good cluster:
c               1) No shared with other track
c               2) No dead strip in the cluster
c               3) cluster length < 4
c
c  weiming yao (LBL) --- 5/30/93
c
c  Bank needed: SVXS, QTRK
c  Procedure: call SVHFOU to get all clusters info:    --- thanks to Oliver
c             loop over rest of displaced tracks to form a possible ks and
c             lambda mass with CTVMFT.
c             1) significance of impact parameter b/sig > 2.5
c             2) lxy/sig > 5.0
c             3) chisqr < 50
c             4) delta m < 20 mev for ks and 6 mev for lambda
c There are two option for ks:
c killing:    5) promt ks, lambda: |Imb_ks/sig_ks| < 6.0
c using it:   6) displaced ks:     |imb_ks/sig_ks| >= 6.0
c

c
c  Input:  SVXS bank no --- you are interested in: displaced or high pt
c          vexnew(3)    --- xyz beam position (cm)
c          vexerr(3,3)  --- error of beam position (cm)
c
c  output: lgood      ---- logical if(T) -> good track
c          nhit       ---- number of svx hits on the track
c          lvee(1)    ---- if(T)-> prompt ks
c          lvee(2)    ---- if(T)-> possible displaced ks
c          lvee(3)    ---- if(T)-> prompt lambda
c          Ipair(1)   ---- Partner SVX bank number of prompt ks
c          Ipair(2)   ---- Partner SVX bank number of displaced ks
c          Ipair(3)   ---- Partner SVX bank number of prompt lambda
c
c  Warning: It will over-write the partner bank no if there are more than ones
c             in the same catalog.
c
c So far, I will provide these arguements, If you need more, please let me
c know.
C===========================
C
C
$$IMPLICIT
C
C==Global Declarations:

$$INCLUDE 'UIPACK$LIBRARY:UIERROR.INC'
$$INCLUDE 'YBOS$LIBRARY:ERRCOD.INC'
$$INCLUDE 'C$INC:BCS.INC'
```

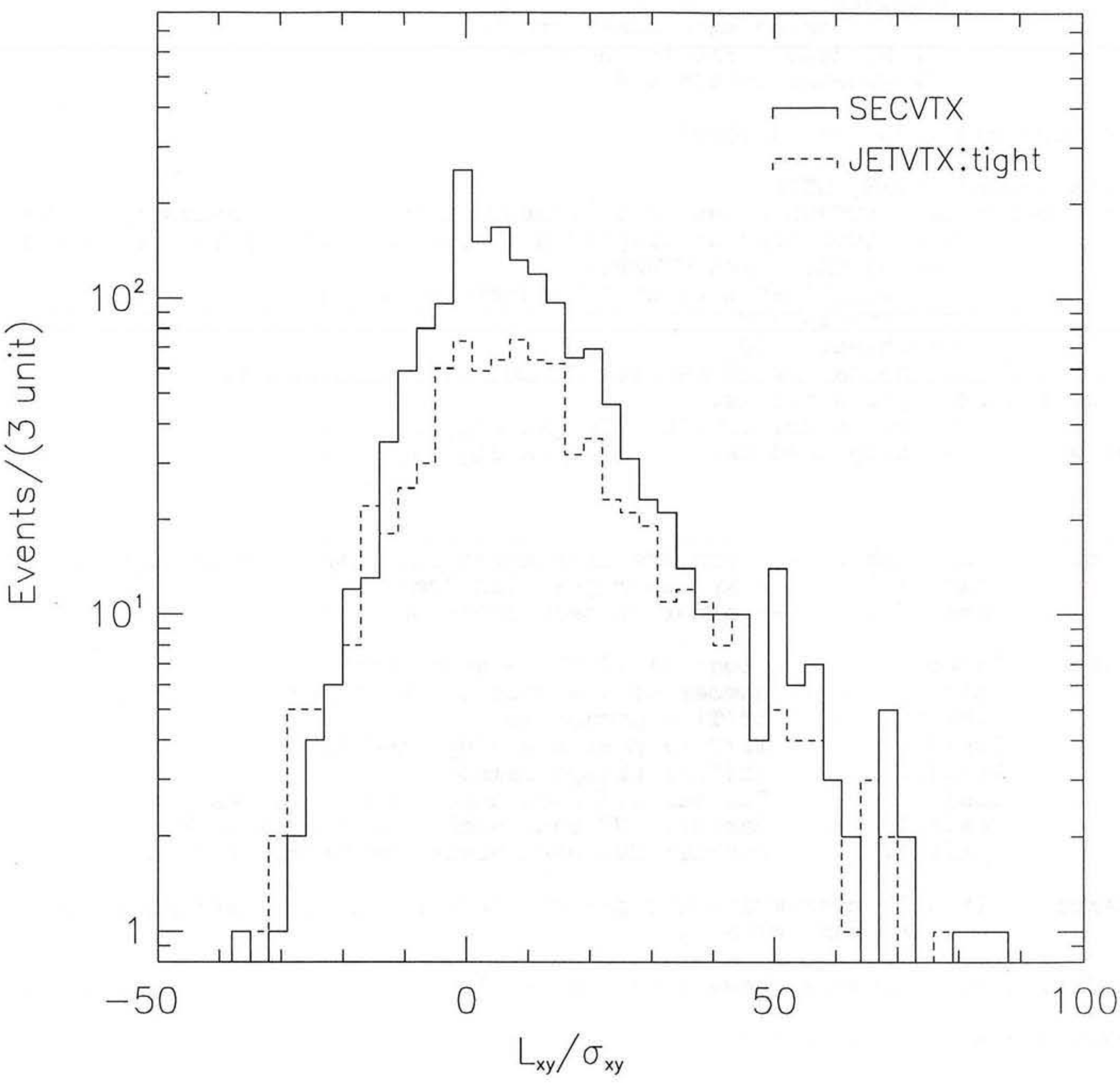
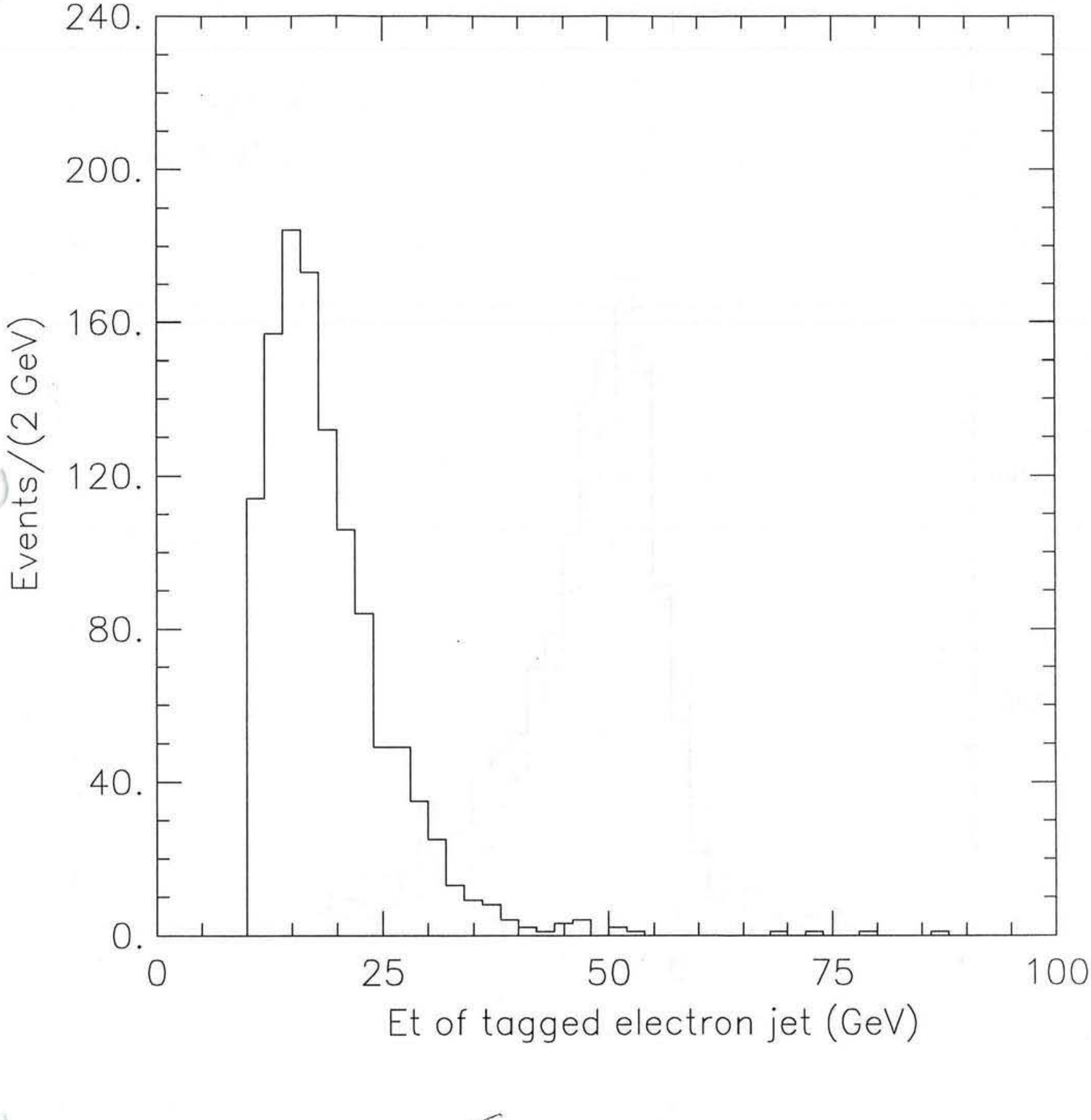
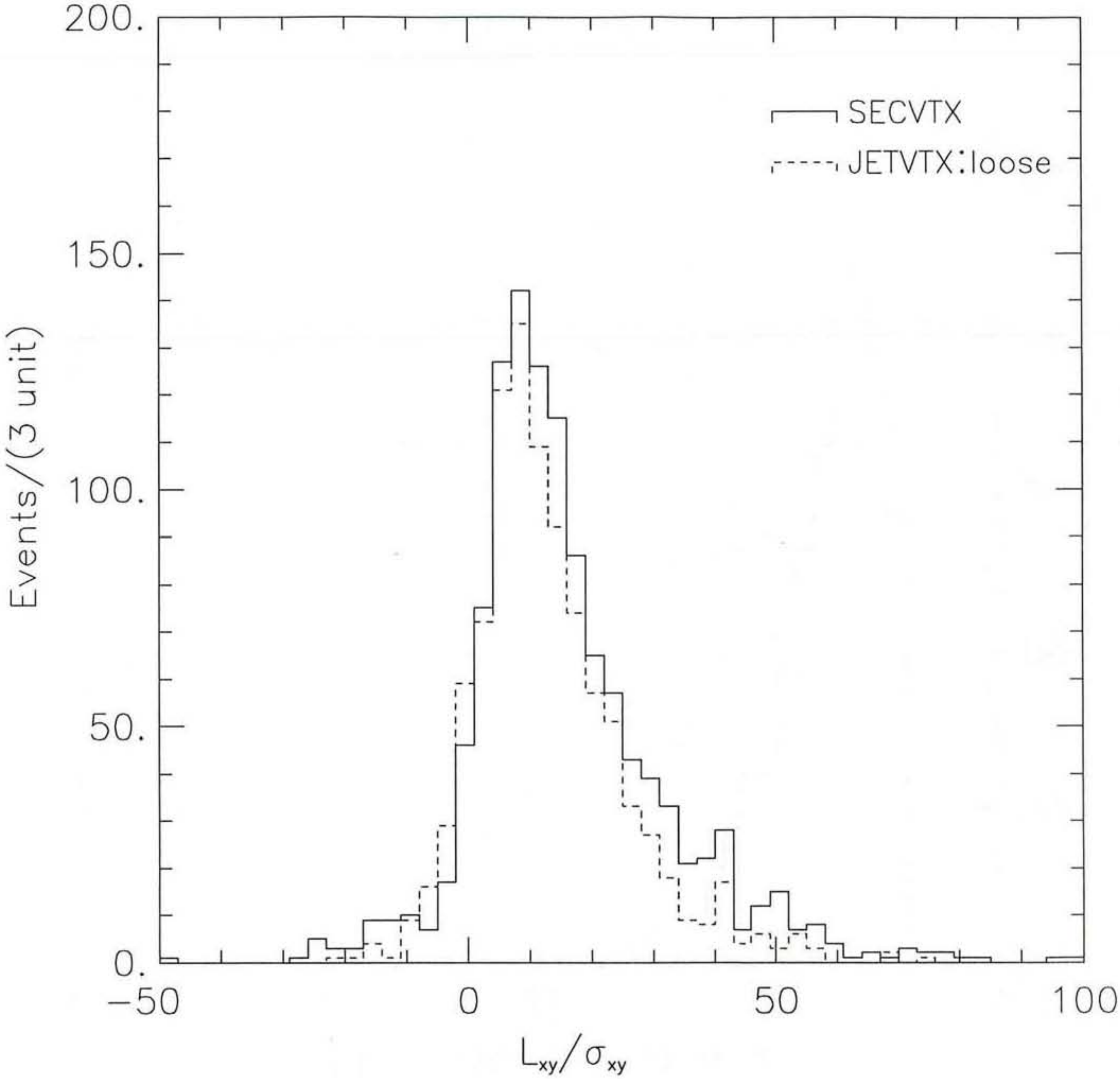


Figure 1



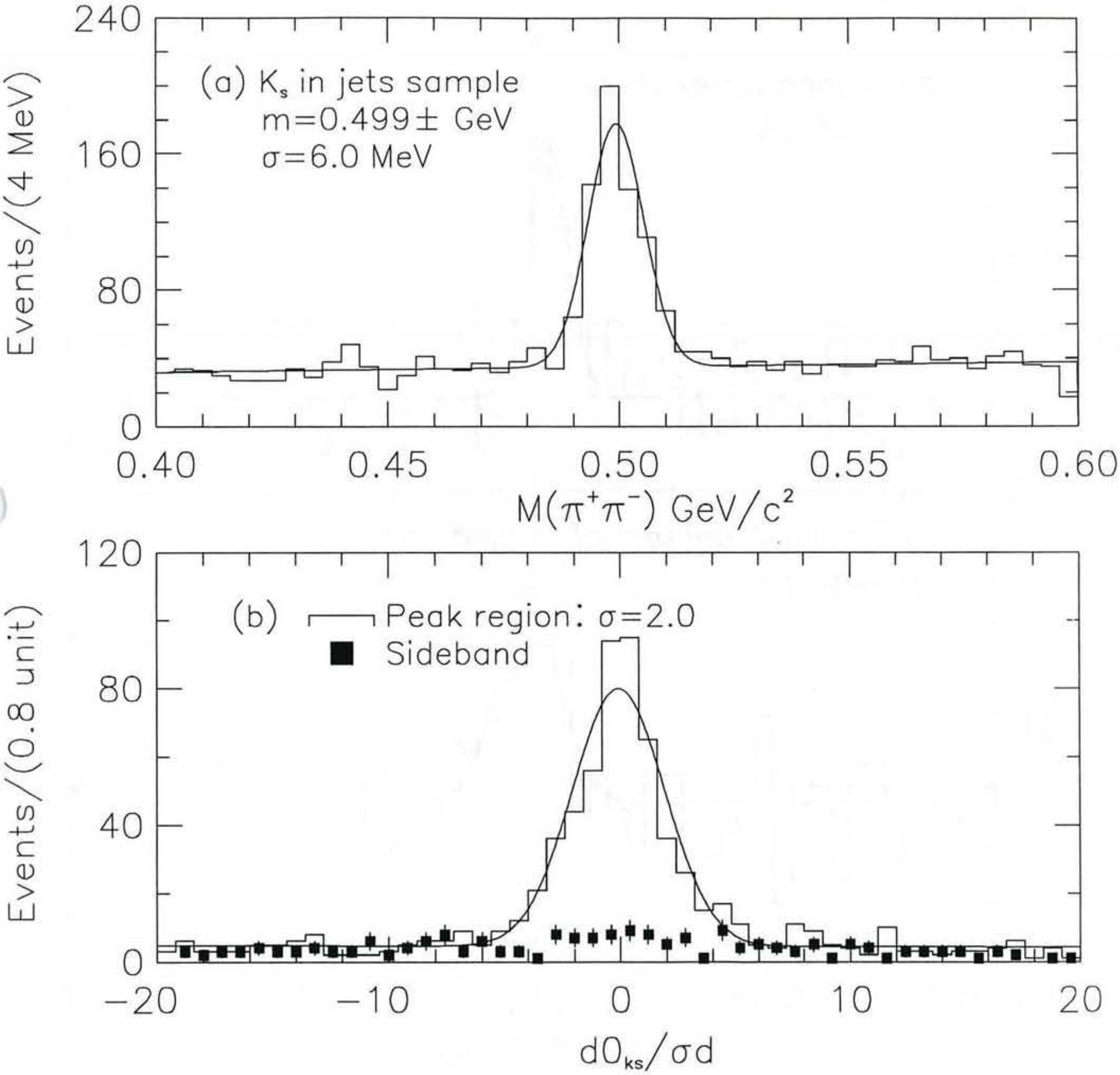
240.
200.
160.
120.
80.
40.
0.
Events/(2 GeV)
0
25
50
75
100
Et of tagged electron jet (GeV)

Figure 2



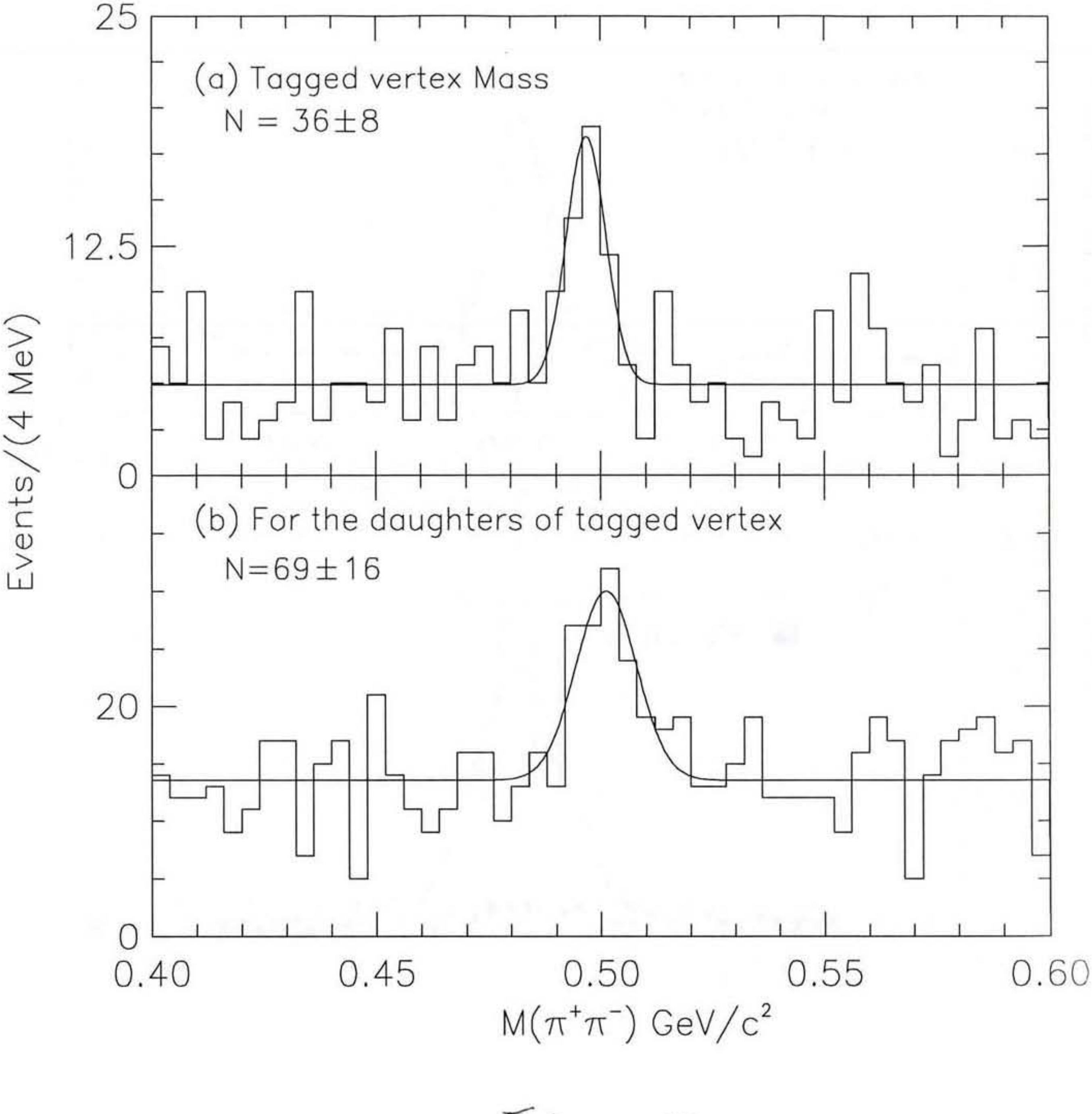
200.
150.
100.
50.
0.
-50
0
50
100
Events/(3 unit)
L_{xy}/σ_{xy}
SECVTX
JETVTX:loose

Figure 3.



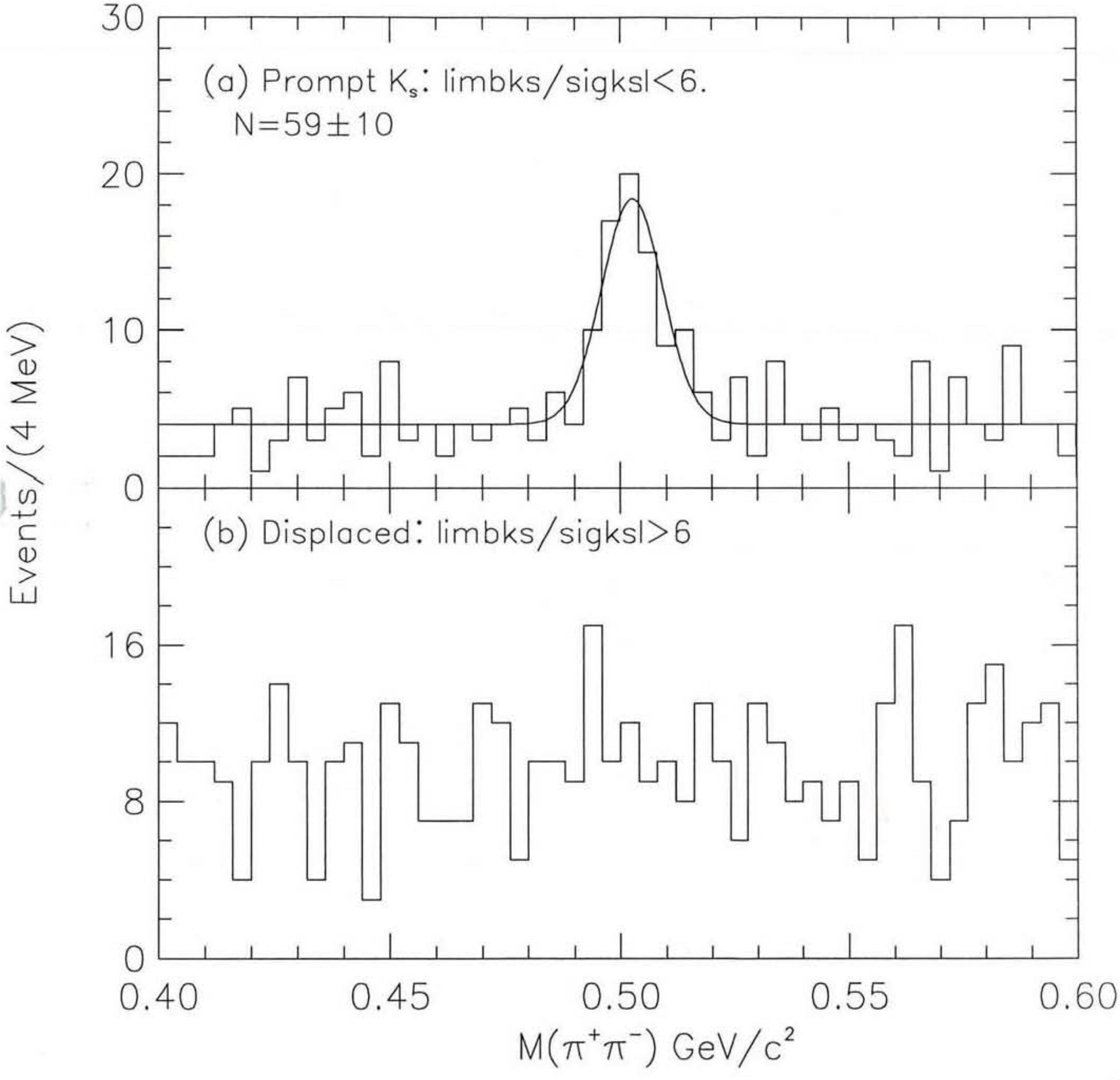
Events/(4 MeV)
(a) K_s in jets sample
m=0.499± GeV
σ=6.0 MeV
240
160
80
0
0.40
0.45
0.50
0.55
0.60
M(π⁺π⁻) GeV/c²
Events/(0.8 unit)
(b) Peak region: σ=2.0
Sideband
120
80
40
0
−20
−10
0
10
20
dO_ks/σd

Figure 4



(a) Tagged vertex Mass
N = 36±8
(b) For the daughters of tagged vertex
N=69±16
Events/(4 MeV)
25
12.5
0
20
0
0.40
0.45
0.50
0.55
0.60
M(π⁺π⁻) GeV/c²

Figure 5



(a) Prompt K_s: limbks/sigksl<6.
N=59±10
(b) Displaced: limbks/sigksl>6
Events/(4 MeV)
$M(\pi^+\pi^-)$ GeV/c^2
30
20
10
0
16
8
0
0.40
0.45
0.50
0.55
0.60

Figure 6