Measurement of the SecVtx Mistag Asymmetry in 5.3.3

João Guimarães da Costa, Sal Rappoccio, Daniel Sherman

July 11, 2005

Abstract

We present a measurement of the mistag asymmetry for the tight and loose SecVtx taggers for Gen5 analyses using the pseudo-$c\tau$ fit method introduced in CDF6739. In the 50-GeV generic jet sample, we measure α (defined as $\frac{N^+_{light}}{N^-}$) to be 1.27 $\pm$ 0.13 for the standard tagger and 1.26 $\pm$ 0.11 for the loose tagger. A further 8% correction (β) common to both taggers is extracted by extending these fit results to the pretag sample. **We have performed equivalent measurements in each jet sample, and we present a ΣE_T-dependent correction to the mistag matrix that includes both α and β based on these new measurements.**

1 Introduction

In any b-tagging analysis, it is important to understand and quantify the contamination of the tagged sample with light-flavor tags, or mistags. These are tags arising from limited detector resolution, long-lived light particle decays (K_s and Λ), and interactions in the beampipe/detector material. Since resolution effects are expected to be symmetric about the primary vertex, the negative side of the vertex L_{2d} distribution has been used as a simple model of the positive light flavor contribution. More recently, it has been shown that the overall distribution of light-flavor tags is biased positive, since the other sources are much more likely to have a positive flight distance.

Two independent measurements of this asymmetry in 4.11.2 suggested that the negative tags be scaled up by an additional 20%. The first of these fit pseudo-$c\tau$ distributions in generic jet data to MC templates to quantify the light-flavor component [1]. The second took advantage of the beam offset (displacement from the origin) to measure the material tag contribution as a function of ϕ [2]. Since the beam is no longer offset, and the SecVtx tagger has been made more resistant to material tags, we focus on the first method in this note. A more complete justification of the method can be found there.

2 Method

The approach here is procedurally similar to that in CDF6739 [1], with only minor modifications. We use the 50 GeV generic jet sample (data and MC) to plot the pseudo-$c\tau$ for all tagged jets, making separate Monte Carlo templates for jets matched to b quarks, c quarks, and light quarks/gluons. We use Pythia Monte Carlo and TopNtuples processed in offline release `5.3.3_nt`. Since the MC is overly optimistic about the detector's resolution (and hence underestimates these tags), we make templates for the tag excess, the positive side of the $c\tau$ distribution minus the negative side. The templates used are shown in Figures 1 and 2.

With a simple χ^2 fit, we estimate the fraction of the tag excess in data that is attributable to each of the three sources. The results of the fits for the two taggers are shown in Figures 3 and 4. In both cases, the fit quality is reasonable (χ^2/dof of 1.38 for the tight tagger, 1.71 for the loose tagger), and the fractions have relative uncertainties provided by Minuit on the order of 1%.

We assume that the data-MC discrepancy in the size of the negative tags can be accounted for with a single scale factor, independent of the heavy flavor content of the jet; this same assumption was made in previous measurements. However, by independently constraining the positive and negative tags in Monte Carlo to add up to the data, we can solve for this scale factor directly in this sample. More specifically, if we instead assumed that the Monte Carlo correctly determines the ratio of positive to negative tags, then the fit values of the Monte Carlo excesses would only correspond to the correct *excess* in data, not the proper numbers of positive and negative tags. We take the ratio of the negative tags in data to the sum of the negative tags in Monte Carlo (corresponding to the fit excesses) as the *Negative Scale Factor*.

We define α to be the nominal mistag asymmetry, the ratio $\frac{N^+_{light}}{N^-_{light}+N^-_b+N^-_c}$, where the numbers come from the final fit after scaling the negative tags (and correcting the positive tags for this change). In principle, then, muliplying the total number of negative tags by α is exactly the light-flavor contribution to the tagged sample.

3 Results

The results are summarized in Tables 1 and 2. The final fits to the total $c\tau$ distribution are shown in Figures 5 and 6. In both cases, the Negative Scale Factor is consistent with the value 1.6 $\pm$ 0.3 measured in [1]. The measured values of α are 1.27 and 1.26 for the tight and loose taggers, respectively, a 1σ excursion from the current Gen5 estimate, 1.1 $\pm$ 0.1. (The estimate is just that; α has not yet been measured in 5.3.3.) Since the Negative Scale Factor was measured independently in Gen4, we also measure α using the exact method in [1], constraining it to be 1.6. If we assume this value for the negative scale factor and correct the positive side for this rescaling as well, we get 1.32 and 1.26 for the tight and loose taggers, respectively, already a slight excursion from 1.1 $\pm$ 0.1. However, if we scale only the negative side, as was done in Gen4, α is measured to be 1.08 and 1.02 for the two taggers.

To attach a systematic uncertainty to α, we consider its dependence on the heavy flavor fraction of the sample. In generic jets, the b and c content is restricted to a few percent of the data. In the W+jets sample, for instance, the heavy flavor content can be twice as large (see CDF7007 for estimates of the heavy flavor content in this sample). We take our systematic from doubling the b and c contribution in the tagged sample. For the tight tagger, this yields a new value of 1.14, and for the loose tagger, 1.15. We therefore adopt measured values of 1.27 $\pm$ 0.13 for tight SecVtx and 1.26 $\pm$ 0.11 for loose SecVtx .

4 Cross Checks

While the pseudo-$c\tau$ is the optimal choice for minimizing data-MC effects, the method can be performed on other vertex variables; all that is required is some discriminating power between the various jet types. We perform an identical analysis on the vertex mass and tag momentum for the tight tagger only as a cross check on the method. The fit quality is not as good (χ^2/dof of 2.40 for the momentum, 2.37 for the mass), but the results are sufficient to support the method. Using these two models and re-deriving α, we get values of 1.44 (mass) and 1.30 (momentum), consistent with the estimate using $c\tau$. A summary of the cross checks is shown in Table 5 and the tag excess fits are shown in Figures 7 and 8. Note that while the fits return largely different estimates of the charm and bottom fractions, they agree relatively well on the light flavor content, which dominates the measurement of α.

	Positive	Negative	Excess
Data Actual	192,908	53,203	139,705
MC B Actual	56,762	1,053	55,709
MC C Actual	23,855	1,289	22,566
MC Light Actual	19,235	10,732	8,503
Negative Scale Factor		1.815	
B Fit	91,674.6	3,040.1	88,634.5
C Fit	33,717.7	3,166.7	30,551
Light Fit	67,515.6	46,996.1	20,519.5
α		1.269	

Table 1: Summary of fit results for the tight tagger. Fit errors are negligible compared to other systematics and have been suppressed here.

	Positive	Negative	Excess
Data Actual	326,564	133,327	193,237
MC B Actual	65,343	2,318	63,025
MC C Actual	32,209	3,137	29,072
MC Light Actual	49,981	30,389	19,592
Negative Scale Factor		1.576	
B Fit	102,043	5,592.2	96,451.2
C Fit	56,021.9	8,144.2	47,877.7
Light Fit	168,498	119,590	48,908.1
α		1.264	

Table 2: Summary of fit results for the loose tagger. Fit errors are negligible compared to other systematics and have been suppressed here.

5 Further Correction

Correcting the mistag matrix prediction by α eliminates the heavy flavor term in the *numerator* of the tag rate (i.e., the number of tags), but the denominator still includes the number of heavy flavor jets in the sample. We define a quantity β which performs this correction, essentially a pretag equivalent of α. More exactly:

$$R^{-}_{mistag} = \frac{N^{-}_{light} + N^{-}_{heavy}}{N^{pre}_{light} + N^{pre}_{heavy}} \tag{1}$$

$$\alpha = \frac{N^{+}_{light}}{N^{-}_{light} + N^{-}_{heavy}} \tag{2}$$

$$\beta = \frac{N^{pre}_{light} + N^{pre}_{heavy}}{N^{pre}_{light}} \tag{3}$$

The average light flavor tag rate is then $R^{-}_{mistag}\alpha\beta$, with R^{-}_{mistag} being the average per-jet mistag probability. The quantity β can be measured either by looking directly at the Monte Carlo or by applying the inverse tagging efficiency to the heavy flavor fits from the α measurement. The former measurement simply utilizes the same jet matching scheme used to get the $c\tau$ templates on the pretag sample, and we find 92.8% of our jet sample to be light flavor, corresponding to a β of 1.077.

The latter method requires that we measure the b and c efficiencies in Monte Carlo, multiply by the data-to-Monte Carlo scale factor (0.909 $\pm$ 0.060 for the tight tagger, 0.927 $\pm$ 0.066 for the loose tagger, with double uncertainty for charm jets), and divide the fit results by this data tag rate to determine the heavy flavor content of the pretag sample. A summary of the numbers used in Tables 3 and 4. In our sample of 11.2 million jets, we derive β corrections of 1.08 $\pm$ 0.01 from the tight tagger and 1.09 $\pm$ 0.01 from the loose tagger. We average these results with the Monte Carlo value and assume a value of 1.08 $\pm$ 0.04, to be conservative in including a systematic for the $c\tau$ templates.

Type	N^{+}_{fit}	ϵ_{MC} (%)	SF	N_{pre} (kJets)
b Jets	91670	33.3	0.909 $\pm$ 0.060	302.8 $\pm$ 20.0
c Jets	33720	7.4	0.909 $\pm$ 0.120	501.3 $\pm$ 66.2
Total	804.1 $\pm$ 86.2 kJets			

Table 3: Summary of numbers used in calculation of β for the tight tagger.

Type	N^{+}_{fit}	ϵ_{MC} (%)	SF	N_{pre} (kJets)
b Jets	102040	38.4	0.927 $\pm$ 0.066	286.7 $\pm$ 20.4
c Jets	56020	9.9	0.927 $\pm$ 0.132	610.4 $\pm$ 86.9
Total	897.1 $\pm$ 107.3 kJets			

Table 4: Summary of numbers used in calculation of β for the loose tagger.

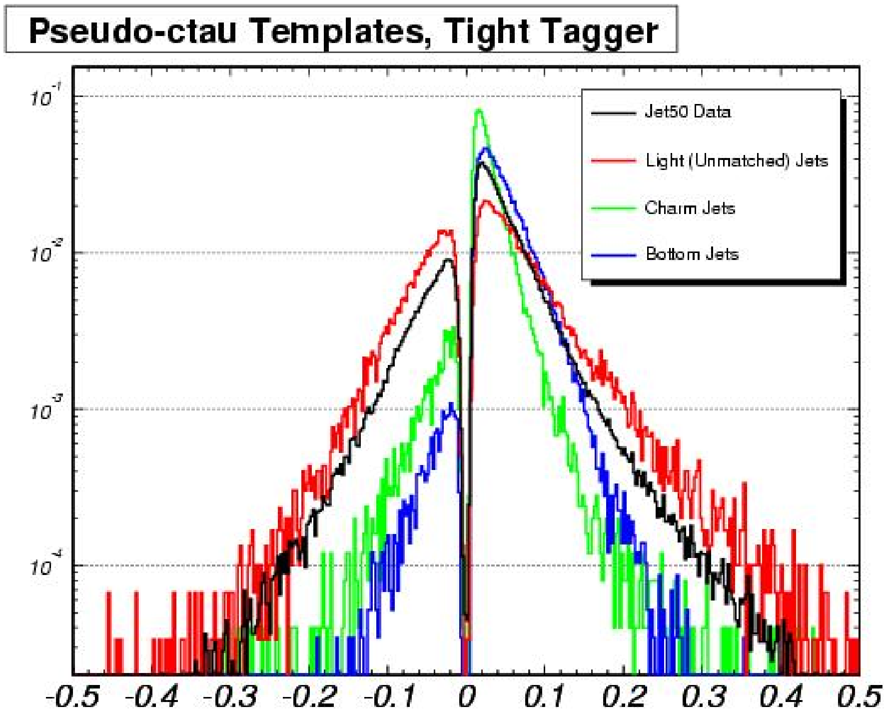

Figure 1: Normalized b, c, and light $c\tau$ templates for the tight tagger.

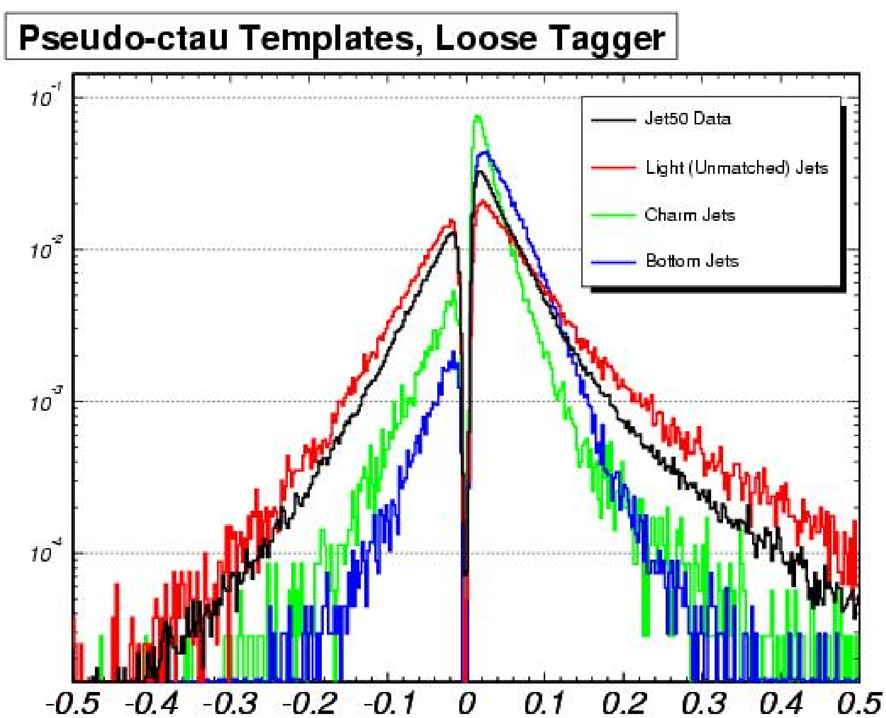

Figure 2: Normalized b, c, and light $c\tau$ templates for the loose tagger.

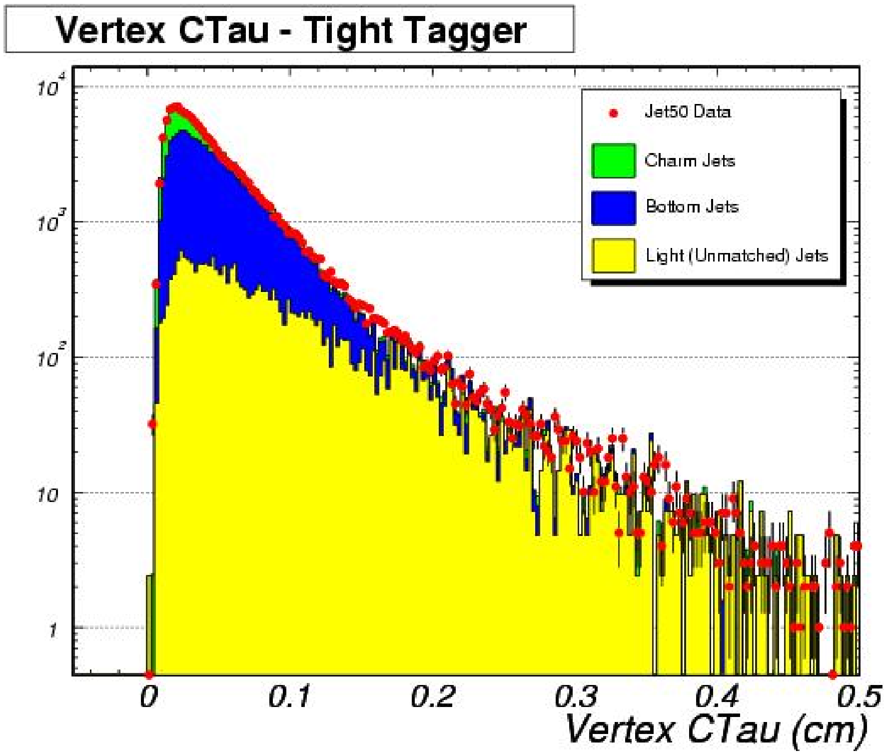

Figure 3: Fitted fractions of b, c, and light jets in the positive tag excess for the tight tagger.

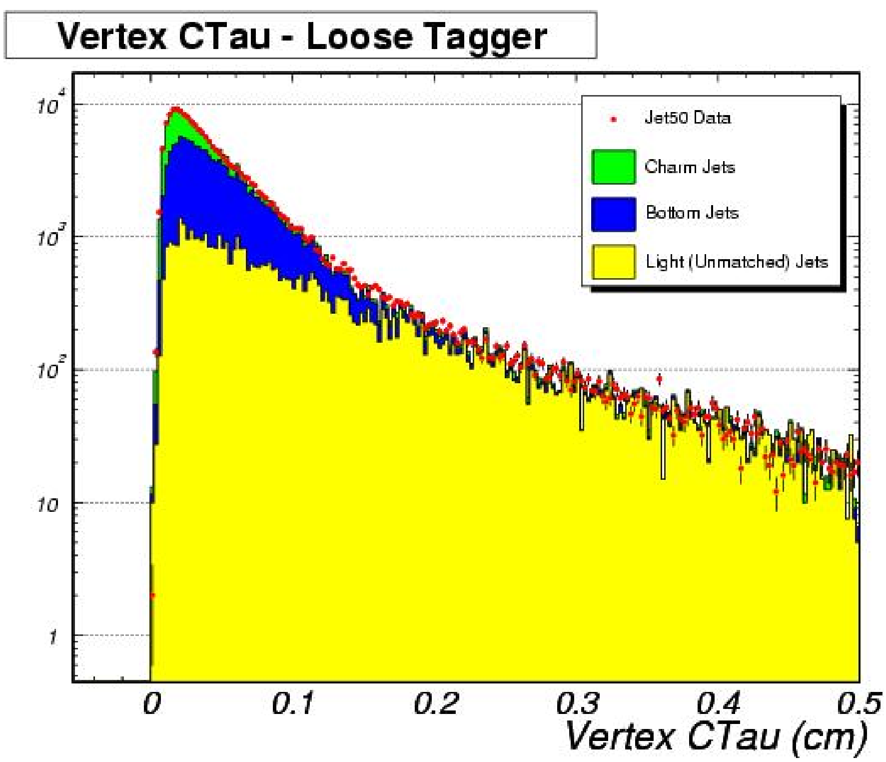

Figure 4: Fitted fractions of b, c, and light jets in the positive tag excess for the loose tagger.

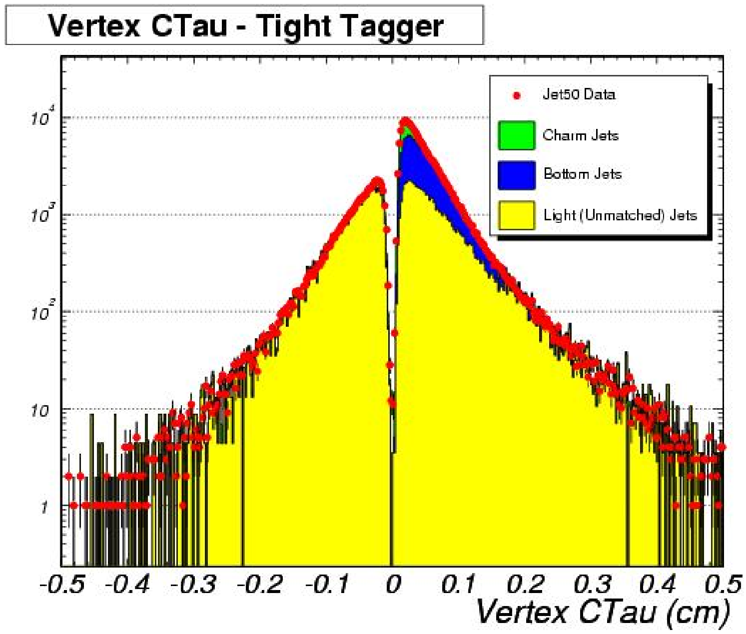

Figure 5: Fitted fractions of b, c, and light jets in all tight tagged jets. The negative tails have been scaled up by 1.81, and the positive side has been corrected accordingly.

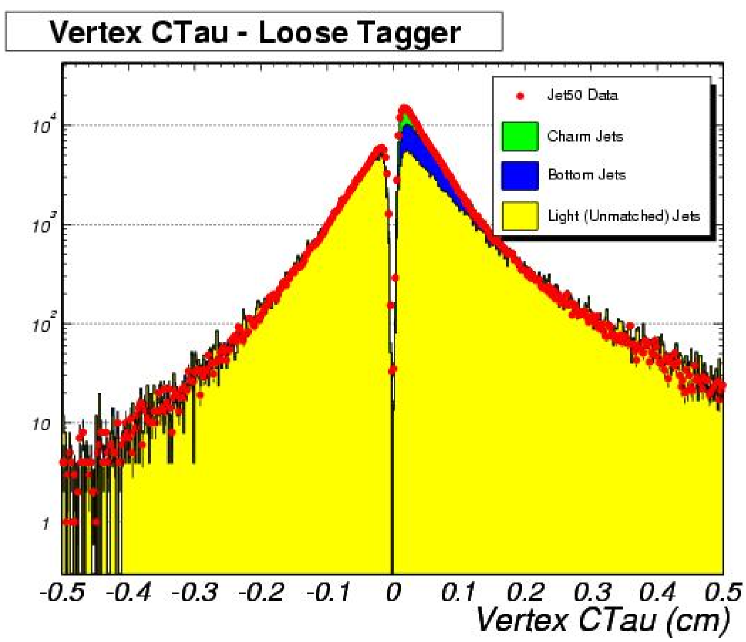

Figure 6: Fitted fractions of b, c, and light jets in all loose tagged jets. The negative tails have been scaled up by 1.52, and the positive side has been corrected accordingly.

	$c\tau$	Vertex Mass	Vertex Momentum
Data Excess	139,705	139,575	123,627
Negative Scale Factor	1.815	1.357	1.573
B Fit Excess	88,634.5	70,463.1	45,641.6
C Fit Excess	30,551	40,991.2	55,667.3
Light Fit Excess	20,519.5	28,120.8	22,308.1
χ^2/dof	1.38	2.40	2.37
α	1.269	1.436	1.299

Table 5: Summary of fit results for the tight tagger cross checks.

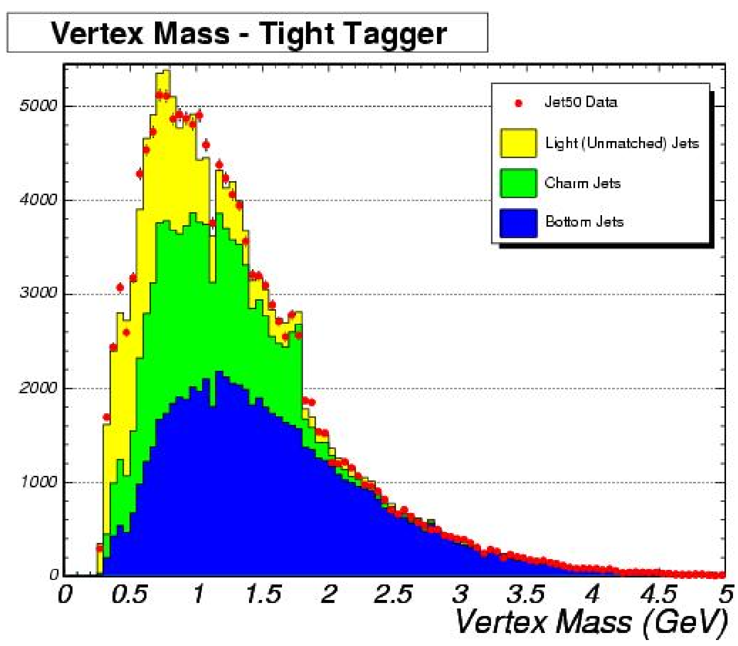

Figure 7: Tag excess fits to the Jet50 vertex mass distribution for b, c, and light jets.

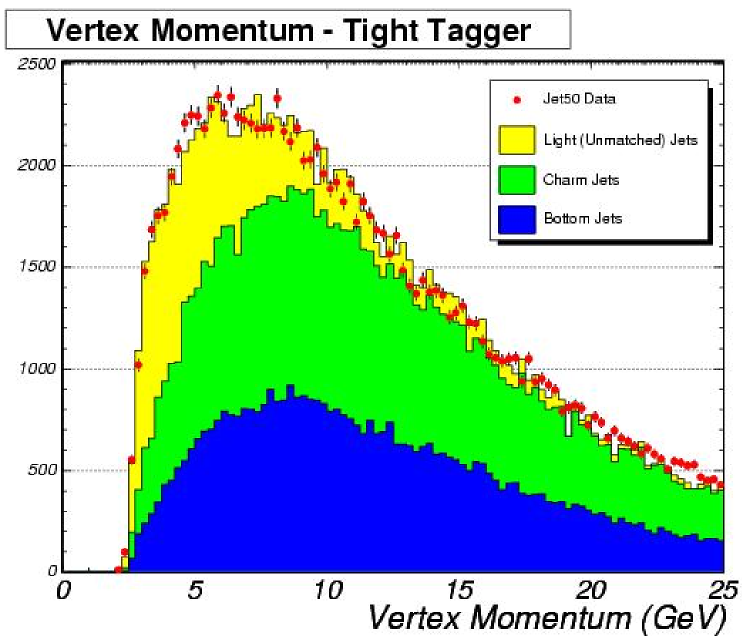

Figure 8: Tag excess fits to the Jet50 vertex mass distribution for b, c, and light jets.

6 Combining Samples

The `Jet50` measurement has been used thusfar to represent the correction in the signal region of the W+jets sample; jet energies there are typically somewhat lower than 50 GeV on average, but the ΣE_T variable in the matrix was beaing sampled appropriately. To properly correct the mistag prediction, we should make an effort to be sensitive to the sample-dependence of the matrix. We have repeated our fit method in `Jet20`, `Jet70`, and `Jet100` for the tight and loose taggers, and by weighting these measurements according to the relative number of data events from each sample in ΣE_T bins, we determine a more precise correction to the matrix prediction. Plots of the templates and fit results follow for both taggers and all samples, and the total correction is plotted as a function of ΣE_T in Figure 17. The correction there includes both α and β. A summary of the fit results is in Table 14, and the total correction used is in Table 15.

Negative SF: 2.60704			
	Positive	Negative	Excess
Data	175590	33129	142461
Fit B	104142	3434.34	100708
Fit C	37240.1	2857.4	34382.7
Fit P	34207.6	26837.3	7370.32
$\alpha = 1.03256$			
MC $\beta = 1.05787$ Tag $\beta = 1.06623$			

Table 6: Summary of fit results for the tight tagger in Jet20.

Negative SF: 2.03484			
	Positive	Negative	Excess
Data	277312	88093	189219
Fit B	113466	5840.03	107626
Fit C	64846.6	8108.48	56738.1
Fit P	98999.7	74144.5	24855.2
$\alpha = 1.12381$			
MC $\beta = 1.05787$ Tag $\beta = 1.07853$			

Table 7: Summary of fit results for the loose tagger in Jet20.

Negative SF: 1.81282			
	Positive	Negative	Excess
Data	192908	53203	139705
Fit B	91636.8	3035.96	88600.8
Fit C	33726.2	3164.67	30561.5
Fit P	67545	47002.4	20542.7
$\alpha = 1.26957$			
MC $\beta = 1.07712$			
Tag $\beta = 1.07763$			

Table 8: Summary of fit results for the tight tagger in Jet50.

Negative SF: 1.57574			
	Positive	Negative	Excess
Data	326564	133327	193237
Fit B	101999	5587.46	96411.3
Fit C	56036.9	8143.34	47893.6
Fit P	168528	119596	48932.1
$\alpha = 1.26402$			
MC $\beta = 1.07712$			
Tag $\beta = 1.0869$			

Table 9: Summary of fit results for the loose tagger in Jet50.

Negative SF: 1.89454			
	Positive	Negative	Excess
Data	121633	38544	83089
Fit B	52433	2668.3	49764.7
Fit C	21841.1	3150.47	18690.7
Fit P	47358.8	32725.2	14633.6
$\alpha = 1.22869$			
MC $\beta = 1.06804$			
Tag $\beta = 1.08223$			

Table 10: Summary of fit results for the tight tagger in Jet70.

Negative SF: 1.58394			
	Positive	Negative	Excess
Data	212506	94504	118002
Fit B	54449.3	4070.83	50378.4
Fit C	38628.1	7661.75	30966.4
Fit P	119429	82771.4	36657.2
$\alpha = 1.26374$			
MC $\beta = 1.06804$			
Tag $\beta = 1.09193$			

Table 11: Summary of fit results for the loose tagger in Jet70.

Negative SF: 1.54387			
	Positive	Negative	Excess
Data	177963	62841	115122
Fit B	68254.2	3874.92	64379.3
Fit C	24613.8	4203.14	20410.7
Fit P	85095	54762.9	30332
$\alpha = 1.35413$			
MC $\beta = 1.06733$ Tag $\beta = 1.07429$			

Table 12: Summary of fit results for the tight tagger in Jet100.

Negative SF: 1.43246			
	Positive	Negative	Excess
Data	321420	151277	170143
Fit B	75269.9	6630.7	68639.2
Fit C	46525.9	11029.9	35496.1
Fit P	199624	133616	66007.7
$\alpha = 1.31959$			
MC $\beta = 1.06733$ Tag $\beta = 1.08426$			

Table 13: Summary of fit results for the loose tagger in Jet100.

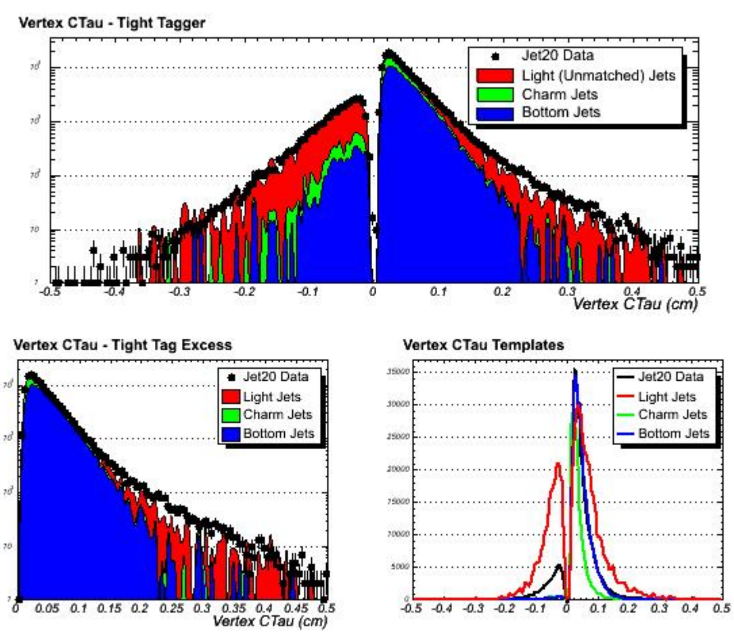

Figure 9: Fit results for the total distribution (top) and the tag excess (bottom left) with the templates used (bottom right) for Jet20 with the tight tagger.

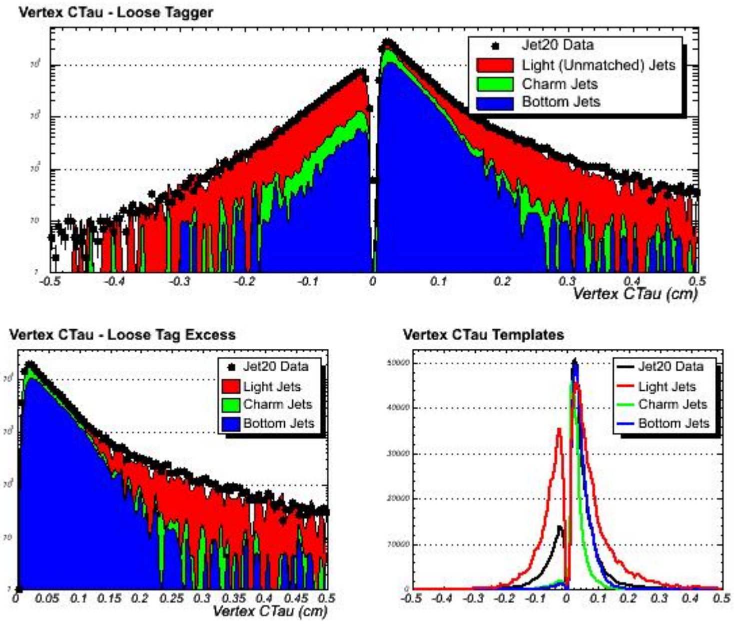

Figure 10: Fit results for the total distribution (top) and the tag excess (bottom left) with the templates used (bottom right) for Jet20 with the loose tagger.

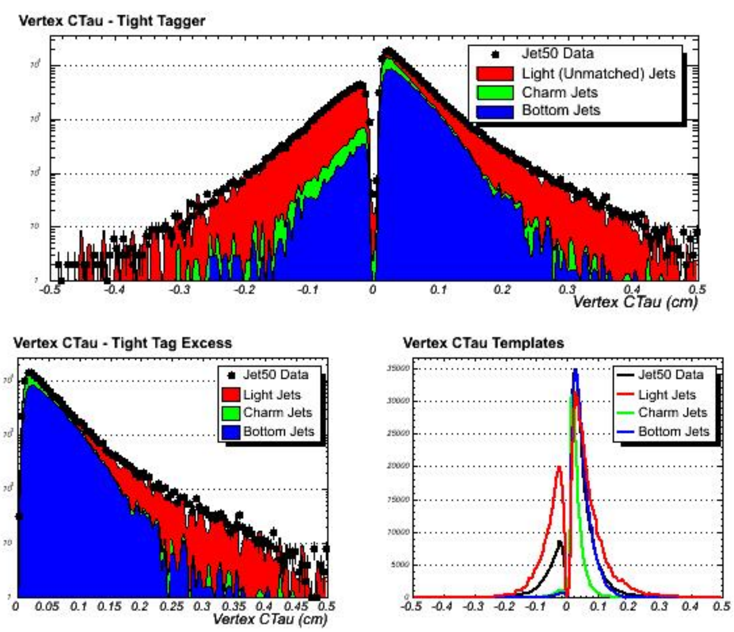

Figure 11: Fit results for the total distribution (top) and the tag excess (bottom left) with the templates used (bottom right) for Jet50 with the tight tagger.

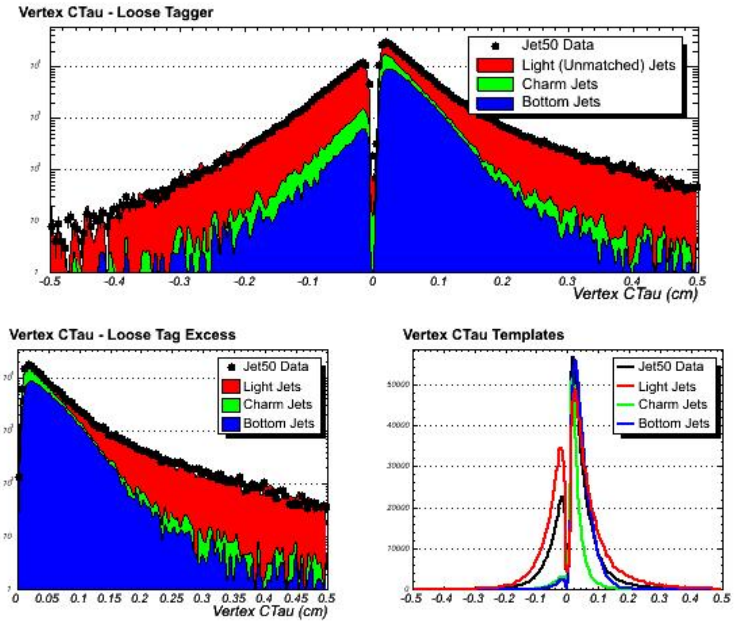

Figure 12: Fit results for the total distribution (top) and the tag excess (bottom left) with the templates used (bottom right) for Jet50 with the loose tagger.

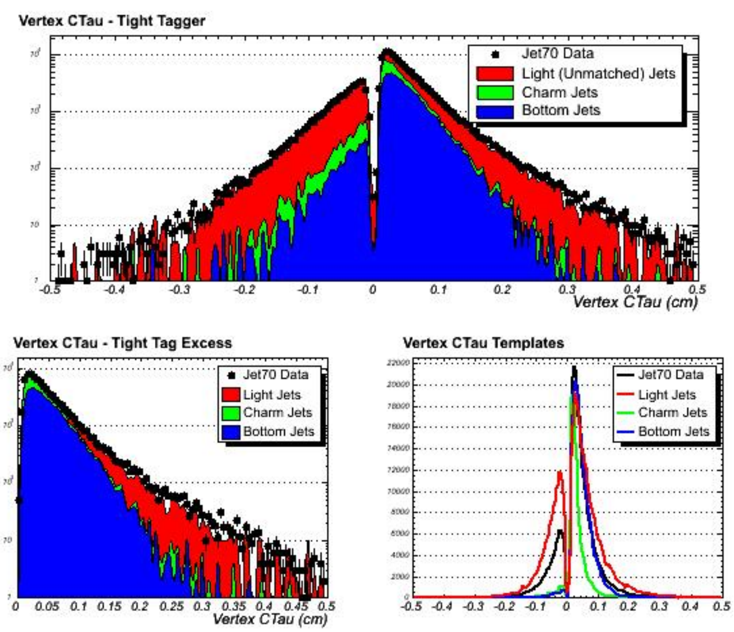

Figure 13: Fit results for the total distribution (top) and the tag excess (bottom left) with the templates used (bottom right) for Jet70 with the tight tagger.

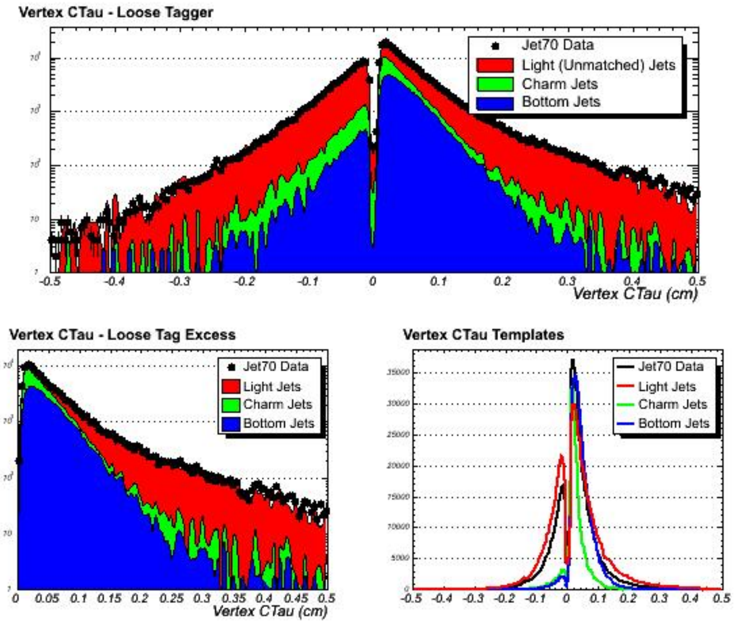

Figure 14: Fit results for the total distribution (top) and the tag excess (bottom left) with the templates used (bottom right) for Jet70 with the loose tagger.

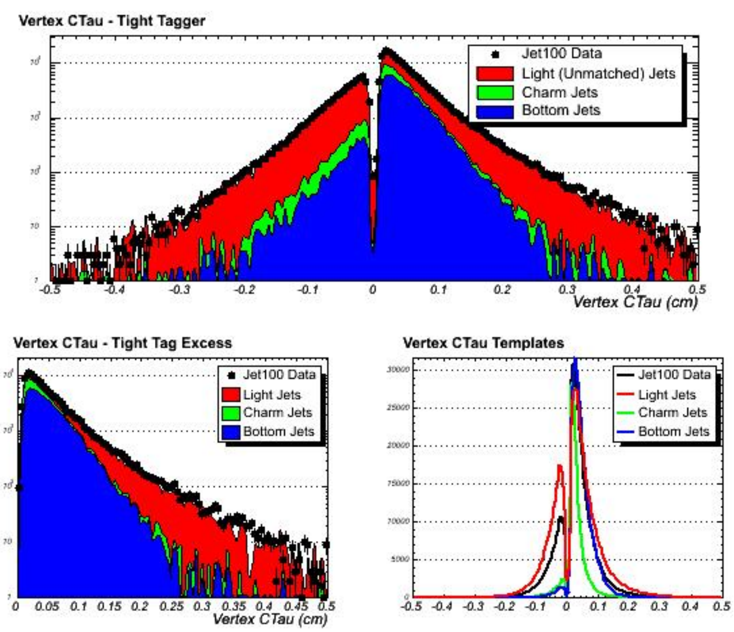

Figure 15: Fit results for the total distribution (top) and the tag excess (bottom left) with the templates used (bottom right) for Jet100 with the tight tagger.

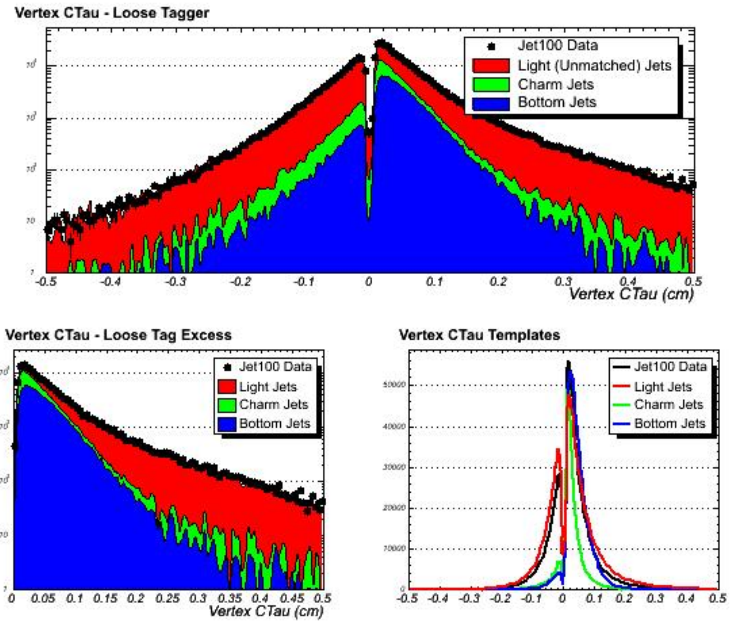

Figure 16: Fit results for the total distribution (top) and the tag excess (bottom left) with the templates used (bottom right) for Jet100 with the loose tagger.

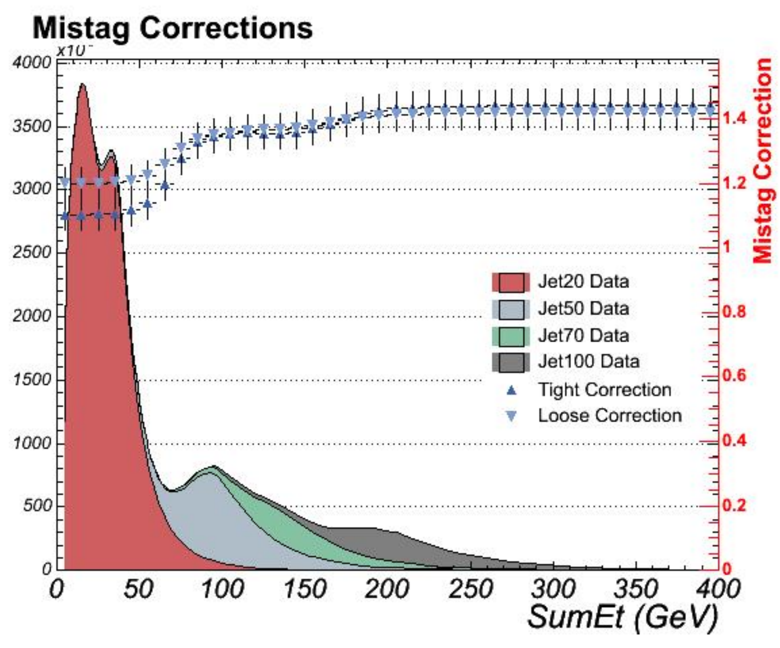

Figure 17: Total ΣE_T-dependent correction to the negative mistag matrix for both taggers. The correction includes α and β.

	Jet20	Jet50	Jet70	Jet100
Common				
N_{data}	23.0M	11.2M	5.6M	6.7M
N_{MC}	6.1M	6.7M	2.8M	6.1M
β_{MC}	1.058 ± 0.001	1.077 ± 0.001	1.068 ± 0.001	1.067 ± 0.001
Tight Tagger				
α	1.03	1.27	1.23	1.35
β_{tag}	1.07	1.08	1.08	1.07
Neg. SF	2.61	1.81	1.89	1.54
Fit χ^2	1.97	1.38	1.54	1.22
Tagged c/b	0.34	0.34	0.38	0.32
MC c/b	0.53	0.41	0.38	0.37
Loose Tagger				
α	1.12	1.26	1.26	1.32
β_{tag}	1.08	1.09	1.08	1.08
Neg. SF	2.03	1.58	1.58	1.43
Fit χ^2	3.06	1.71	1.39	1.02
Tagged c/b	0.53	0.50	0.61	0.52
MC c/b	0.57	0.46	0.0.44	0.44

Table 14: Summary of fit results for both taggers.

7 Conclusions

We have revisited the technique of pseudo-$c\tau$ fitting for determining the mistag asymmetry α in SecVtx tagged jets. In Jet50 data and MC, we measure asymmetries of 1.27 ± 0.13 for the default tagger and 1.26 ± 0.11 for the loose tagger, consistent with the Gen4 measurement of 1.2 ± 0.1 and the previous Gen5 estimate of 1.1 ± 0.1. The method has been checked using the vertex momentum and mass distributions as templates, yielding consistent results. We further derive a second correction, β, to correct for the heavy flavor jets in the denominator of the tag rates. In Jet50, this correction is determined to be 1.08 ± 0.04. By repeating the measurement in all jet samples, a more precise ΣE_T-dependent correction is derived.

References

[1] Bachacou, H., C. Ferretti, J. Nielsen, and W. Yao, 'SECVTX Tag Composition and Heavy Flavor Fraction Studies in QCD Jets' [CDF/ANAL/TOP/CDFR/6739].

[2] Foland, A., J. Guimarães da Costa and S. Rappoccio, 'Examination of Material Effects Using Jet Data In SecVtx For the Winter 2004 Conferences' [CDF/PUB/SEC_VTX/PUBLIC/6906].

ΣE_T Range (GeV)	Tight Correction	Loose Correction
0-10	1.104	1.199
10-20	1.103	1.199
20-30	1.105	1.200
30-40	1.107	1.201
40-50	1.116	1.207
50-60	1.140	1.221
60-70	1.199	1.258
70-80	1.279	1.306
80-90	1.327	1.335
90-100	1.347	1.348
100-110	1.355	1.356
110-120	1.357	1.361
120-130	1.356	1.364
130-140	1.356	1.368
140-150	1.359	1.372
150-160	1.368	1.378
160-170	1.384	1.386
170-180	1.401	1.396
180-190	1.416	1.404
190-200	1.426	1.409
200-210	1.432	1.412
210-220	1.435	1.414
220-230	1.437	1.415
230-240	1.438	1.415
240-250	1.439	1.416
250-260	1.439	1.416
260-270	1.440	1.416
270-280	1.440	1.416
280-290	1.440	1.416
290-300	1.441	1.417
300-310	1.441	1.417
310-320	1.441	1.417
320-330	1.441	1.417
330-340	1.441	1.417
340-350	1.441	1.417
350-360	1.441	1.417
360-370	1.441	1.417
370-380	1.441	1.417
380-390	1.441	1.417
390-400	1.441	1.417

Table 15: Total correction ($\alpha\beta$) for both taggers binned in ΣE_T.