Top Quark Decay Kinematics in Fully Reconstructed $t\bar{t}$ Events in the e or $\mu + \not\!\!E_T + \geq 4$ Jet Decay Channel

by

Eugene Hall Guillian

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Physics)
in The University of Michigan
1999

Doctoral Committee:

Professor Myron K. Campbell, Chair
Professor Frederick D. Becchetti, Jr.
Professor Richard B. Brown
Professor Gordon L. Kane
Professor Gregory Tarle

Acknowledgements

When I entered graduate school in the Fall of 1992, I never imagined how tough my life was going to be for the next seven years. On more occasions than I care to remember, I considered quitting graduate school and occupy myself with something more down-to-earth than the study of fundamental particles. I have, however, survived the ordeal. Moreover, I have written a thesis which — in spite of all its flaws — I can begin to believe in. I could not have come where I am today without the help of many. I would like to take this opportunity to thank them all.

Fist of all, I want to thank everybody at Fermilab and the CDF Collaboration who helped make Tevatron Run I a great success. The hard work and dedication of these talented scientists and engineers resulted in, among other things, a first glimpse of the nature of the top quark, on which my thesis is based.

I also want to thank all of my colleagues in the CDF group in the University of Michigan Physics Department. They have all been positive and supportive, and helped to create a productive environment.

Thanks are also due to Tina Wells and everybody at the Physics Department Student Office. They were always kind and helpful, and their support allowed me to focus on my graduate work.

The members of the Top Group at CDF have provided numerous technical assistance in analyzing the top candidate sample. I thank them all for their help. Special thanks to Nathan Eddy and Rob Roser, to whom I often turned to for information about data samples and analysis code.

I thank Kevin Burkett, Soo-Bong Kim, Steve Vejcik, and Myron Campbell for all their patience while I was stumbling my way to become the Level-2 Trigger Expert. I never

succeeded in completely taking over the awesome responsibility shouldered by Kevin alone for so long. But I am glad I was able to share some of the burden, and to work with such an awesome guy.

Special thanks to Rudy Thun who, early on in graduate school, assigned a "homework" problem in which I was to perform a matrix element calculation of the top quark decay process. His challenge made me think carefully about how the top quark decays. As a result, I was able to obtain a clear and thorough understanding of this process. I also want to thank Soo-Bong Kim for lively discussions about the top quark decay process. His tough, down-to-earth questions inspired me to find ways to clearly describe this process, and made me realize the important role that interference effects play. Discussions with Dave Winn and Dan Amidei were also very helpful in making me think clearly about top quark decay and issues regarding measurement. I thank them both for these stimulating and thought-provoking discussions.

I owe a debt of gratitude to Rob Roser, who gave me support in a time of need. I hope to repay this debt by becoming a great physicist.

I cannot thank my parents enough for all the sacrifices they made in bringing me up and providing me with a good education. But, even more, I thank them and my sister Ako for the less tangible gifts they have given me as I was growing up. I thank my father for having had so much trust in me, and for having given me so much freedom since early in my childhood. I thank my mother for having always been around, for having supported me in all of my endeavors, and for having been so flexible and open-minded. And I thank Ako for teaching me by example the importance of being tough and resilient. I believe their influence on me is responsible, in large part, for having given me the strength and courage to start and complete my graduate studies.

Finally, I thank my advisor Myron Campbell for having given me so much freedom in my research. This made my work riskier than if he had called the shots. Because of it, however, I was able to make many mistakes and learn from them. I truly believe that the best way to learn is by making mistakes. So by entrusting me with freedom, Myron has allowed me to make my thesis not only a stepping stone in my professional career, but to use it as an opportunity to grow as a physicist and as a thinker.

Preface

The analysis presented in this thesis began as an exploration of CP violation in the production and decay of the top quark. This was back in 1994, when the first evidence for the existence of the top quark was just submitted by the CDF collaboration to Physical Review Letters. The fact that I was focusing on CP violation in the top quark while everybody else was arguing about its very existence may give the impression that I was very forward-looking; in fact, I was simply clueless, groping in the dark trying to find a thesis topic while I spent the lion's share of my time looking after the Level-2 Trigger.

Before I started my analysis, I rummaged through many theory-oriented papers on CP violation in the top quark. I was quickly discouraged by the near-incomprehensibility of most of the papers, and I felt, at the time, that my survey of the literature was an exercise in futility. In retrospect, however, I believe that this exercise served a useful purpose: I realized that top quark physics was in its early infancy, and that the study of CP violation was premature. There were much more basic issues about top quark physics that neither I nor my colleagues had a firm grip on.

I, therefore, decided to focus on basic questions. Early in graduate school, Rudy Thun, a senior colleague of mine at the University of Michigan, suggested that we (I and other young graduate students) use Feynman diagrams to obtain the differential decay distribution of the top quark decay product momenta. I used this "homework" problem as a starting point to explore in detail how the top quark decays. This exploration opened my eyes to fascinating aspects of this process. I realized that the top quark decay was much more than a simple, two-step sequential process $t \to b + W$ followed by $W \to \ell + \nu$ or $q + \overline{q}'$; quantum interference effects play an interesting and essential role in providing beautiful simplicity to the decay process. Over the years, I extended my exploration to other aspects

of top quark physics, as well as experimental issues about what measurements to make and how they should be made. I made many fascinating discoveries, and, in spite of being a little tired and depressed after seven, very trying years in graduate school, I feel glad that I went on this exploration. The tangible result of my exploration is this thesis.

I am proud and gratified with my thesis. However, I do feel a need for an apologia about its flaws. First, the thesis is very long — almost 400 pages. My defense for this is that there really are a lot of interesting things to discuss, and it took that many pages to do this. Also, I wanted my thesis to be useful to others, not just a ream of papers that show that I did this measurement. So I included a lot of details, used a lot of illustrations, and explained ideas from various points of view. I am hoping that a young graduate student who reads this will be as excited about top quark physics as I am.

Another flaw of this thesis is that I ignored some important issues that would have made the measurement process more correct, but also more complicated. For instance, the non-W background to the $t\bar{t}$ sample has, in some respects, significantly different kinematic distributions compared to the W + jets background. Thus, an adequate measurement should treat these to processes separately. In my thesis, however, I simplified the analysis by using the W + jets monte carlo to model both types of background processes. Another oversimplification I made was to assign zero error to the estimated background fractions in various subsamples of the $t\bar{t}$ candidate events. A list of other oversimplifications and inadequacies in my analysis is given in chapter 5. I defend these oversimplifications on the following grounds. What I did, no doubt, causes errors to be underestimated. In spite of this, I show in chapter 8 that the errors are so large that no meaningful limits can be set with Run I statistics. Because of this, I felt that, at this stage, it made no sense to split hairs trying to finesse the measurement process; that is something I leave to those who attempt similar measurements in Run II. I feel that, right now, it is more important to grasp the big picture. For this, I believe that it is useful to make a simplified (if inadequate) measurement in order to clearly expose the factors that are most important in setting the scale of the measurement errors. Any finessing of the measurement will only obscure the big picture.

In spite of these flaws, I believe my thesis succeeds in laying out a map of the terrain

of top quark physics. I hope it will serve as a guide for those who perform measurements on the properties of the top quark in Run II.

Gene Guillian

May, 1999

List of Tables

List of Figures

Table of Contents

List of Appendices

Chapter 1

Introduction

1.1 About the Top Quark

According to the standard model, matter, at its most fundamental level, is formed of two types of fermions: leptons and quarks. These fermions interact with each other by exchanging fundamental spin-1 particles called gauge bosons. Also, it is believed that the lepton, quark, and gauge boson masses are generated by the coupling of these particles to fundamental scalar particles called the Higgs bosons. To date, no substantial disagreement between the standard model and experimental data exist[23],[24]. See [1], [2], and [3] for basics on the standard model.

The top quark is one of six quarks in the standard model. It is a 3^{rd} generation quark. In its left-handed helicity state, it is a weak isospin partner of the the left-handed helicity state of b, the other 3^{rd} generation quark. Table 1.1 summarizes the hierarchical structure of quarks and leptons in the standard model.

As of this writing, strong evidence for the existence of the top quark has been reported by the CDF and D0 collaborations ([5] through [17]). If it is assumed that the top quark candidate events observed by CDF and D0 indeed originate from processes involving the production and decay of top quarks, then the mass of the top quark is measured to be about 175 GeV; this would make it by far the most massive fundamental particle ever observed.

As suggested above, there is, as yet, some possibility that the top quark candidate events observed by CDF and D0 are not, or are not entirely, due to the production and

	GENERATION		
	1^{st}	2^{nd}	3^{rd}
QUARKS	$\begin{pmatrix} d \\ u \end{pmatrix}_L, d_R, u_R$	$\begin{pmatrix} s \\ c \end{pmatrix}_L, s_R, c_R$	$\begin{pmatrix} b \\ t \end{pmatrix}_L, b_R, t_R$
LEPTONS	$\begin{pmatrix} e \\ \nu_e \end{pmatrix}_L, e_R, (\nu_e)_R$	$\begin{pmatrix} \mu \\ \nu_\mu \end{pmatrix}_L, \mu_R, (\nu_\mu)_R$	$\begin{pmatrix} \tau \\ \nu_\tau \end{pmatrix}_L, \tau_R, (\nu_\tau)_R$

Table 1.1: A summary of the hierarchical structure of quarks and leptons in the standard model. The subscripts L and R indicate the left- and right-handed helicity state of a given fermion. The grouping of the L-state fermions into a column vector indicates that they form a weak $SU(2)$ doublet; the R-state fermions are weak $SU(2)$ singlets, so they are not grouped.

decay of top quarks. For instance, the top quark is a spin-$\frac{1}{2}$ particle; to date, however, no measurements have been made to demonstrate that this is so. Likewise, the weak isospin of the top quark is $T_3 = +\frac{1}{2}$, but no measurements have been made to support this fact. In spite of such missing pieces of information, however, there are compelling theoretical and experimental reasons why the most plausible source of the candidate events is the top quark.

1.2 Why the Top Quark Should Exist

References [19] and [3] have excellent discussions about why the top quark should exist. In this section, the arguments in these references are summarized.

One reason for the existence of the top quark is based on the renormalizability of the standard model. Before 1975 when the first 3^{rd} generation particle — the τ lepton — was discovered [20], the standard model was a highly successful theory of fundamental particle interactions. Its success hinged on the fact that: (1) it gave good quantitative predictions for all observed processes; (2) it did not predict processes that were not observed; and (3) the theory was renormalizable. (See [22] for an excellent account of the state of the standard

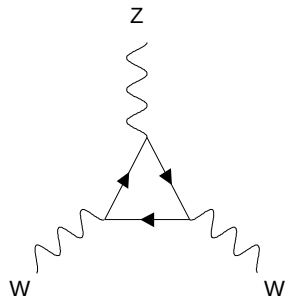


Figure 1.1: An example of a triangle anomaly diagram. These diagrams cause the standard model to be unrenormalizable unless the cancellation equation 1.1 is valid.

model before the discovery of τ). This final feature of the theory, however, depends crucially on the following fact:

> Within a given generation, the sum of the charge of all leptons and quarks weighted by the color factor must be zero.

This statement in equation form is given as follows:

$$Q(\ell) + Q(\nu_\ell) + 3 \cdot [Q(q_d) + Q(q_u)] = 0 \tag{1.1}$$

$Q(\ell)$, $Q(\nu_\ell)$, $Q(q_d)$, and $Q(q_u)$ are the charge of the charged lepton ℓ, the neutrino partner ν_ℓ of ℓ, the down-type quark q_d, and the up-type quark q_u. The factor 3 in front of the quark terms accounts for the fact that quarks come in three color states. This equality is necessary in order to cancel out infinite terms arising from Feynman diagrams of the form shown in figure 1.1. Without this cancellation, the standard model is not renormalizable, and thus the theory is incomplete. For the first two generations, $Q(\ell) = -1$, $Q(\nu_\ell) = 0$, $Q(q_d) = -1/3$, and $Q(q_u) = +2/3$. These values satisfy the sum in equation 1.1, so the standard model with two fermion generations is renormalizable.

This state of affairs, however, ended with the discovery of the τ lepton. Detailed studies of the production and decay of τ showed that, except for its mass, it had exactly the same properties as the electron and muon. In other words, it is a third generation lepton.

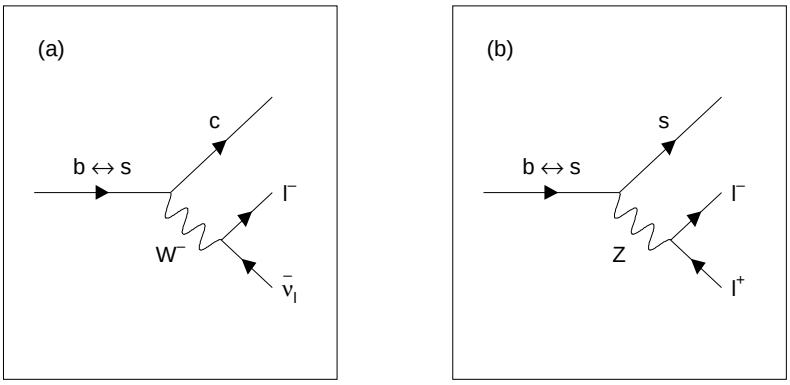


Figure 1.2: The decay of a weak isosinglet b-quark via mixing with the strange (or down) quark. (a) Charged current decay; (b) neutral current decay.

The renormalizability of the standard model, therefore, requires that third generation quarks exist. One of these — the b-quark — was discovered in 1977 [102]. The other 3^{rd} generation quark is, by definition, the top quark, and it ought to exist.

Experiments on the properties of the b-quark provide further support for the existence of the top quark. If the top quark does not exist, then the b-quark must be in a weak isosinglet state. This means that the b-quark cannot decay through the channels $b \rightarrow c + W^-$ and $b \rightarrow u + W^-$ via CKM mixing because this can happen only if b is a member of an isodoublet. Instead, it must decay by mixing with 1^{st} and 2^{nd} generation d-type quarks, as shown in figure 1.2(a). But if the process in this figure is possible, then so must that shown in figure 1.2(b). The relative rates for the processes $b \rightarrow \ell^+\ell^- X$ and $b \rightarrow \ell^-\nu_\ell X$ is:

$$\frac{\Gamma(b \rightarrow \ell^+\ell^- X)}{\Gamma(b \rightarrow \ell^-\nu_\ell X)} \approx 0.11 \tag{1.2}$$

Experimentally, the decay $b \rightarrow \ell^+\ell^- X$ does not occur at a detectable level, while $b \rightarrow \ell^-\nu_\ell X$ occurs frequently, so the ratio is measured to be much smaller than 0.11. Also, the diagrams in figure 1.2 imply a much greater degree of B_d^0-$\overline{B_d^0}$ mixing than is observed experimentally. For these reasons, the isosinglet hypothesis is incorrect.

More quantitative evidence for the existence of the top quark is obtained by examining $Z \rightarrow b\bar{b}$. An isosinglet b-quark gives $\Gamma(Z \rightarrow b\bar{b})$ which is about thirteen times

smaller than that for an isodoublet b. Experimentally, LEP has measured $R_b = \Gamma_b/\Gamma_{had} = 0.2170 \pm 0.0009$; this is in excellent agreement with the standard model with isodoublet b, giving the value 0.2158 [25]. Also, the forward-backward asymmetry of the b-quark in $Z \rightarrow b\overline{b}$ in the Z resonance region shows that the third component of the weak isospin of b is $T_3 = -0.490^{+0.015}_{-0.012}$ [26]. These two results convincingly demonstrate the fact that the b-quark is a member of a weak isodoublet having $T_3 = -1/2$. Thus a partner with $T_3 = +1/2$ must exist: this is the top quark.

1.3 Direct Evidence for the Existence of the Top Quark: Run 1a

In 1994, the CDF collaboration reported direct evidence for the existence of the top quark [17], [16]. The evidence was based on data collected during Tevatron Run 1a in 1992-1993, with integrated luminosity of 19 pb^{-1}. At the Tevatron, where p and $\overline{p}$ are collided at a center-of-mass energy of $\sqrt{s} = 1.8\ TeV$, the processes with the greatest production cross section for top quarks are $q\overline{q} \rightarrow t\overline{t}$ and $gg \rightarrow t\overline{t}$. The produced t and $\overline{t}$ decay to the following intermediate states:

$$\begin{aligned} t &\rightarrow b + W^+ \\ \overline{t} &\rightarrow \overline{b} + W^- \end{aligned}$$

The decay of the $t\overline{t}$ system is categorized according to how W^+ and W^- decay. If both decay to $\ell + \nu_\ell$, where $\ell = e$ or μ, the decay channel is call *dilepton*; if one decays to $\ell + \nu_\ell$ while the other decays to a quark pair, the decay channel is called *lepton + jets*; if both decay to quark pairs, then the decay channel is called *all-hadronic*. These are the three main decay channels of the top quark. In addition to the above, there is a decay channel where one or both W decay to $\tau + \nu_\tau$; this decay channel is referred to as *tau*. The decay channels are summarized in table 1.2.

The Run 1a CDF search for the top quark examined the dilepton and lepton + jets channels. The details of the event selection criteria for these channels can be found in [17],

DECAY CHANNEL	DESCRIPTION	BRANCHING FRACTION
dilepton	Both W's decay to ℓ + ν_ℓ, ℓ = e or μ.	4/81 (5%)
lepton + jets	One W decays to ℓ + ν_ℓ, the other decays to a quark pair.	24/81 (30%)
all-hadronic	Both W's decay to a quark pair.	36/81 (44%)
tau	One or both W's decay to τ + ν_τ.	17/81 (21%)

Table 1.2: The decay channels of the $t\bar{t}$ system, and their branching fractions.

	DILEPTON	SVX	SLT
BACKGROUND	$0.56^{+0.25}_{-0.13}$	2.3 ± 0.3	3.1 ± 0.3
OBSERVED	2	6	7
PROBABILITY	12%	3.2%	3.8%

Table 1.3: A summary of CDF's Run 1a top search. The column under SVX (SLT) refer to the lepton + jets events with one or more jets tagged by the the secondary-vertex (soft-lepton) b-tagging algorithm. Row 1: Expected number of background events. Row 2: Number of observed events. Row 3: Probability that background events alone can account for the observed excess of events. The SVX and SLT channels have three events in common.

[16]; here, they will be sketched. For the dilepton channel, two high-P_T e or μ, large missing transverse energy ($\not\!\!E_T$), and two or more jets are required; for the lepton + jets channel, one high-P_T e or μ, large $\not\!\!E_T$, and three or more jets are required. In order to increase the signal-to-background ratio in the lepton + jets channel, at least one jet was also required to have a b-tag. An excess of events was found in both decay channels (table 1.3). When the statistics in both channels are combined, the probability that the number of background events fluctuated to give the observed number of events is 0.26%, which corresponds to a 2.8 σ excess.

1.4 Direct Evidence for the Existence of the Top Quark: Run 1b

The Run 1a result presented in the last section provides only circumstantial evidence for the existence of the top quark: the analysis indicates an excess of events passing cuts that are relatively efficient for $t\bar{t}$ and inefficient for background. It, however, provides no evidence that the excess is due to, or due only to, $t\bar{t}$ production and decay. In order to establish this, the experimental data must be examined for consistency with the $t\bar{t}$ hypothesis. The following is a list of consistency requirements:

$t\bar{t}$ Production Cross Section

If the events in data come from, and only from, the top quark, then: (1) the measured values of the $t\bar{t}$ production cross section should agree with the theoretical calculation evaluated at the experimentally measured top quark mass; and (2) the cross section measurement in orthogonal decay channels should all agree with each other.

Top and W Mass Spectrum

The top quark mass spectrum should cluster around a fixed value. Also, the measured mass in different decay channels should agree with each other. Finally, in decay channels with at least one hadronically decaying intermediate W, the mass spectrum of the hadronic W should cluster around $M_W \approx 80\ GeV$.

Kinematic Features of $t\bar{t}$ Events

The fact that the top quark decays by the sequential, 3-body process $t \to b + W$ followed by $W \to \ell + \nu_\ell$ or $q + \overline{q}$ suggests that it might have kinematic features that distinguish it from: (1) known background processes; and (2) possible new physics processes that mimic the production and decay of $t\bar{t}$. Assuming that signal and background monte carlos adequately model the processes of interest, all kinematic distributions in $t\bar{t}$ candidate events should be described by a fixed combination of signal and background monte carlo distributions.

During Tevatron Run 1b (1994 to 1996), the CDF and D0 collaborations collected about four to five times as much data as they did in Run 1a. With this level of statistics,

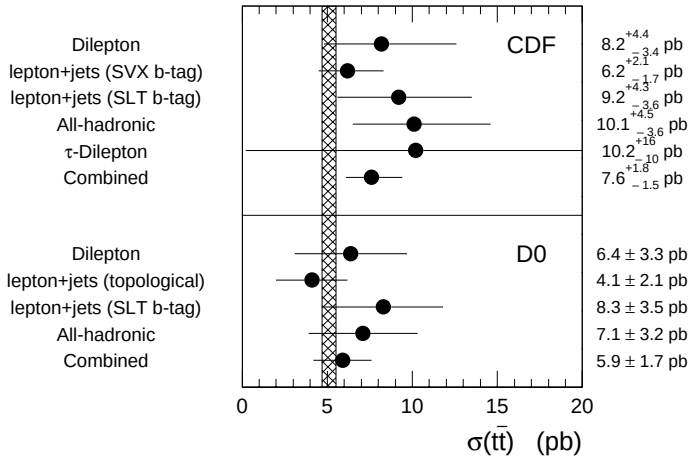


Figure 1.3: $p\overline{p} \rightarrow t\overline{t}$ cross section measurements. This figure is reproduced from figure 33 in [19]. The CDF combined cross section excludes the tau-dilepton decay channel; the D0 combined cross section excludes the all-hadronic decay channel. The hatched band represents the theoretical $t\overline{t}$ cross section, taking $m_{top} = 175\ GeV$. The band spans the region 4.7-5.5 pb.

they were able to check these consistency requirements. The following sections summarize their findings.

1.4.1 $t\overline{t}$ Production Cross Section

Figure 1.3 shows the Run 1 (*i.e.* combined Run 1a and Run 1b) $p\overline{p} \rightarrow t\overline{t}$ production cross section for various decay channels from the CDF and D0 collaborations. Within errors, the measured values of the cross section agree with the theoretical values (evaluated at m_{top} = 175 GeV). Also, within errors, the cross section in the various decay channels agree with one another. Thus, within the precision attainable in Run 1, the experimental data are consistent with the $t\overline{t}$ hypothesis.

1.4.2 Top and W Mass Spectrum

The cross section measurements make clear the fact that known background processes are extremely unlikely to account for the excess of events passing the top candidate selection cuts. They do not, however, take into account the possibility that some or all of

the excess may be originating from new physics processes that do not involve the top quark. One way to check whether or not the source of the excess is due solely to the production and decay of the top quark is to examine the reconstructed top mass distribution.

At CDF and D0, the excess top candidate events are presumed to originate from the production and decay of a $t\bar{t}$ pair. If one were able to reconstruct the t and $\bar{t}$ mass with perfect precision, the mass distribution should be narrowly peaked about a mean value (assuming m_{top} is not too large). If, however, the excess top candidate events do not originate from a pair of particles of equal mass with narrow width, and one were to reconstruct such events as if they did, the resulting mass distribution would (even at parton-level) most likely be of a smeared-out, continuum form. Therefore, one way to check if the excess top candidate events originate from a $t\bar{t}$ pair is to see if the top mass distribution clusters about some fixed value.

In an experimental setting, this is more easily said than done. In order to measure the top quark mass on an event-by-event basis, it is necessary to deduce the quark and neutrino momenta from the objects found in the detector. This process is referred to as *event reconstruction*. Attempts have been made by CDF and D0 to perform event reconstruction in the dilepton, lepton + jets, and the all-hadronic channels. Event reconstruction in the dilepton and all-hadronic channels is quite unreliable, while it is fairly reliable in the lepton + jets channel.[1] Because of this, the resonance in the top mass distribution is smeared out and difficult to detect in the dilepton and all-hadronic channels, while it is easier to detect in the lepton + jets channel.

Figure 1.4 shows the reconstructed top mass distribution in the lepton + jets channel from CDF.[8] There does appear to be a clustering around 170-190 GeV. However, there is apparently not to enough statistics to rule out a continuum distribution that is harder than that from the known background. For the sake of completeness, the top mass distribution in the all-hadronic channel is shown in figure 1.5. The distribution does seem to agree well with signal + background monte carlo (signal has $m_{top} = 175\ GeV$). However, as expected,

[1] The unreliability in the former channels is due to high degree of information loss from: (1) the large amount of energy carried away by the two neutrinos in the dilepton channel; and (2) the large combinatoric background in assigning jets to the correct quarks in the all-hadronic channel. The lepton + jets channel, an hybrid of the other two channels, suffer from both of these problems, but to lesser degrees.

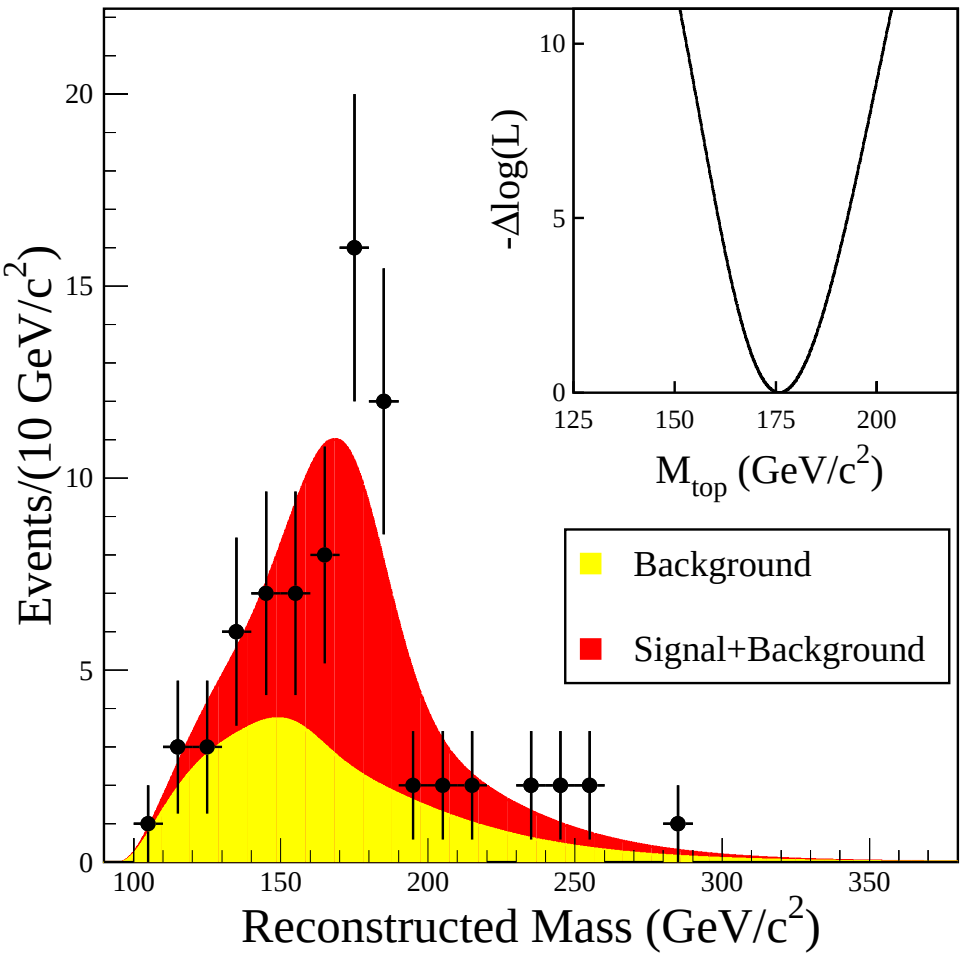


Figure 1.4: The reconstructed top mass distribution in the lepton + jets channel. From Tevatron Run 1, CDF. See [8].

no clear resonance structure is seen (nor is it expected by the monte carlo). No mass plot is available for the dilepton channel, mainly on account of the large degree of ambiguity in defining a unique reconstructed mass on an event-by-event basis.

Another test of the consistency of the experimental data with the $t\bar{t}$ hypothesis is to see if the top mass measured in different decay channels agree with each other. Figure 1.6 shows the results of top mass measurements from CDF and D0, in various decay channels. It is seen that, within the precisions attainable in Run 1, all of the decay channels have

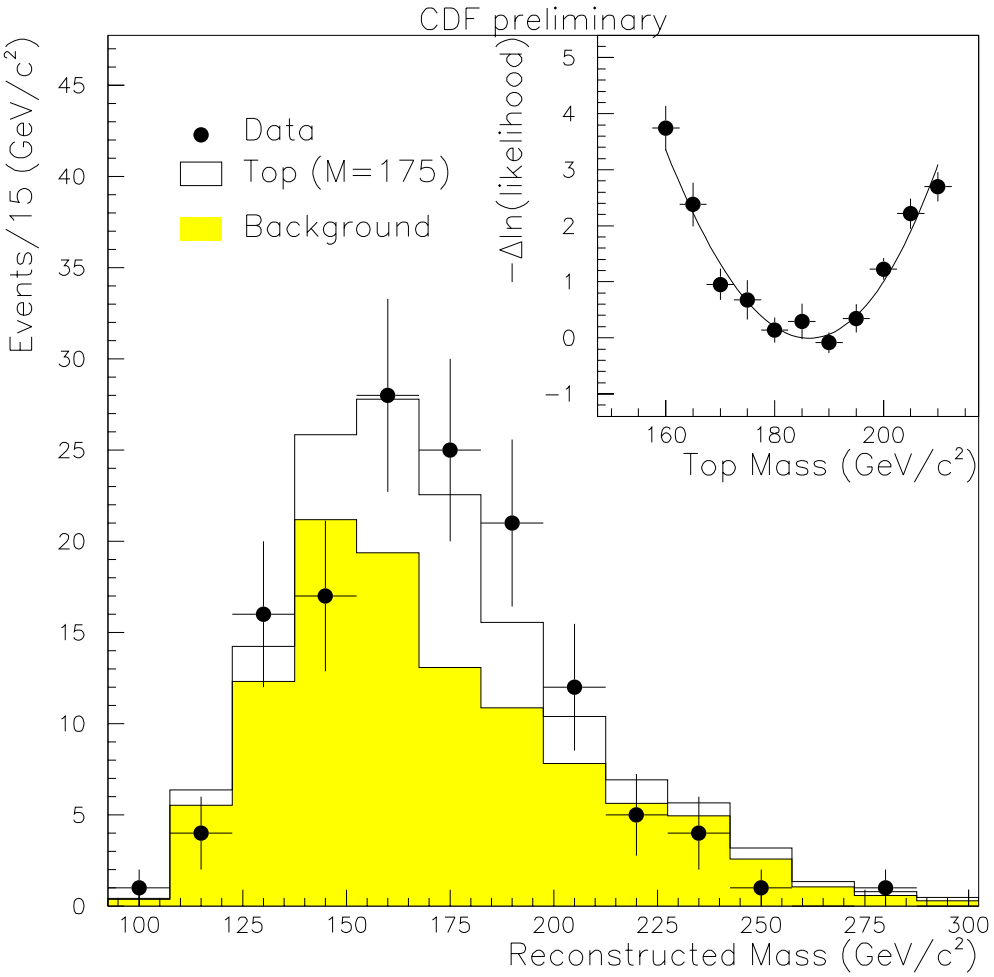


Figure 1.5: The reconstructed top mass distribution in the all-hadronic channel. From Tevatron Run 1, CDF. See [10].

compatible masses. The conclusion, therefore, is that the Run 1 data are consistent with the $t\bar{t}$ hypothesis.

Yet another test of the consistency of the experimental data with the $t\bar{t}$ hypothesis makes use of the fact that two on-shell W's exist in $t\bar{t}$ events. In cases where W decays hadronically, one can measure its mass. If the event reconstruction process is sufficiently reliable, one ought to see a resonance around 80 GeV in the W mass distribution. A search for this resonance in the lepton + jets channel has been carried out by CDF.[5] The results are summarized in figures 1.7 to 1.9. These figures make a compelling case that hadronically decaying W bosons do seem to be present in the top candidate events. This is yet another piece of evidence supporting the $t\bar{t}$ hypothesis.

1.4.3 Kinematic Features of $t\bar{t}$ Events

One can imagine performing two types of tests on the kinematic features of $t\bar{t}$ candidate events to verify that the excess events are consistent with the $t\bar{t}$ hypothesis. In the first type of test, no attempt is made to reconstruct the momenta of the decay products of $t\bar{t}$, while in the second type, this reconstruction is performed.

In the first type of test, the quantities to be examined are: (1) the charged lepton and jet momentum, and missing transverse energy; and (2) various combinations of items in the first category. An example of a variable in the second category is H_T, the scalar sum of the transverse momentum of all objects in an event.

Let us denote any given kinematic quantity as x. Also, let $< x >$ be the average value of x over a given sample of events. Typically, $< x >$ in background processes is different from that in $t\bar{t}$ events. (For instance, taking $x = H_T$, $< H_T >$ in $t\bar{t}$ events is much larger than in background events.) Thus, if the $t\bar{t}$ candidate events really consist of background + signal events with signal being only $t\bar{t}$, then $< x >$ evaluated over the candidate events should, on average, lie part way between $< x >$ for the background and signal. More specifically, if the estimated background fraction in the candidate event sample is β, then the most likely value of $< x >$ in this sample is:

$$< x >_{exp} = \beta \cdot < x >_{back} + (1 - \beta) \cdot < x >_{sig}, \tag{1.3}$$

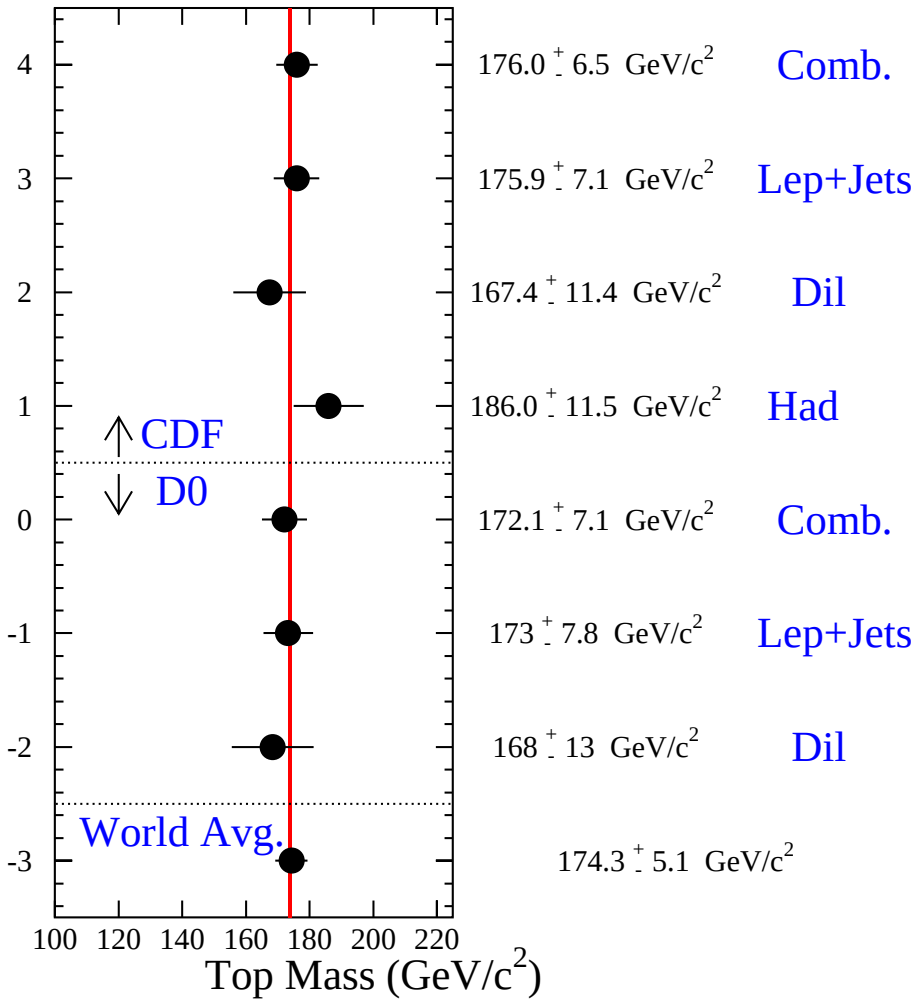


Figure 1.6: A summary of top mass measurements in various decay channels of $t\bar{t}$. The data are from Tevatron Run 1, CDF and D0.

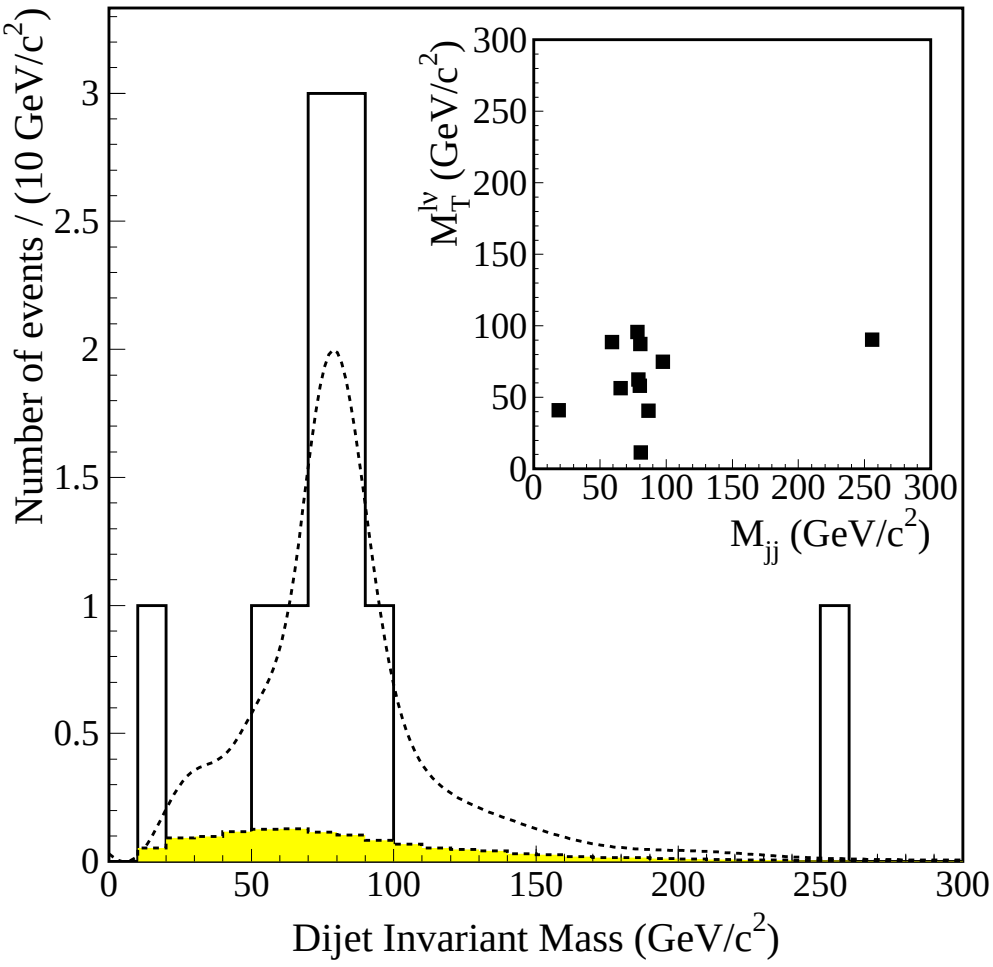


Figure 1.7: The invariant mass distribution of the jet pair presumed to originate from the hadronic decay of W in double b-tagged lepton + jets events. The inset shows the correlation between the transverse mass of the ℓ-ν_ℓ system and the dijet mass. See [5].

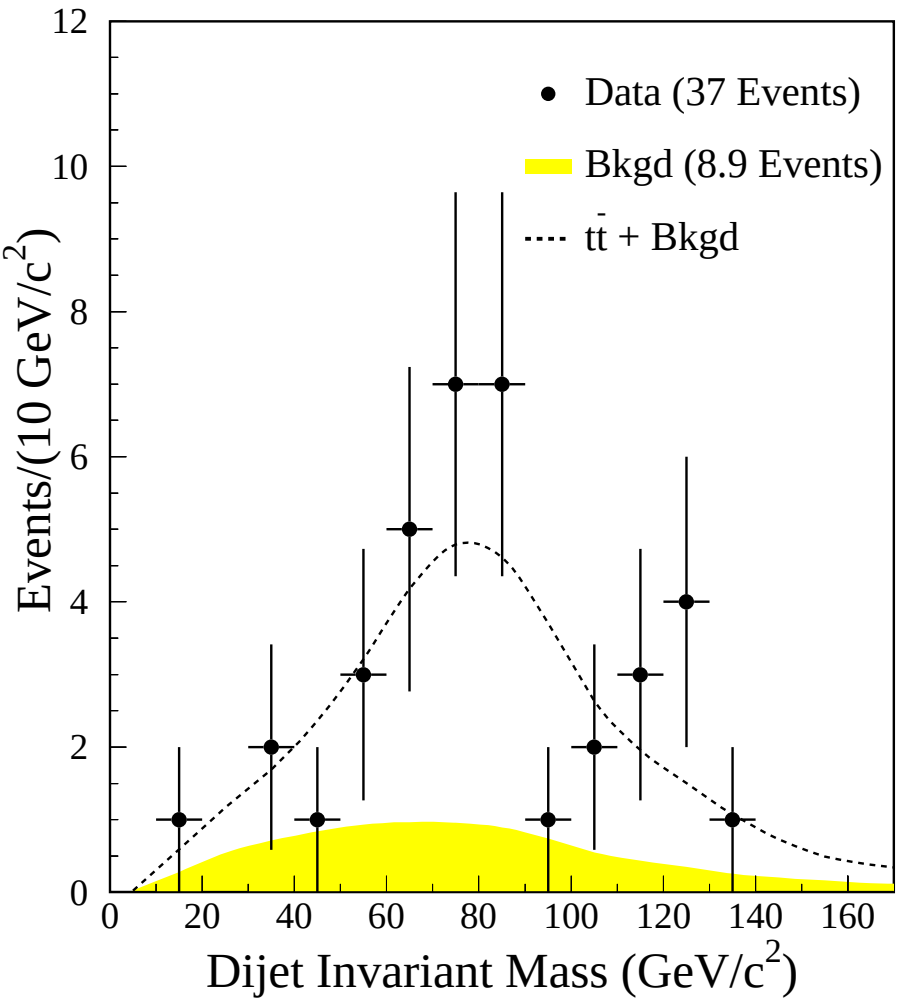


Figure 1.8: The invariant mass distribution of the jet pair presumed to originate from the hadronic decay of W in lepton + jets events with ≥ 1 b-tag. See [5].

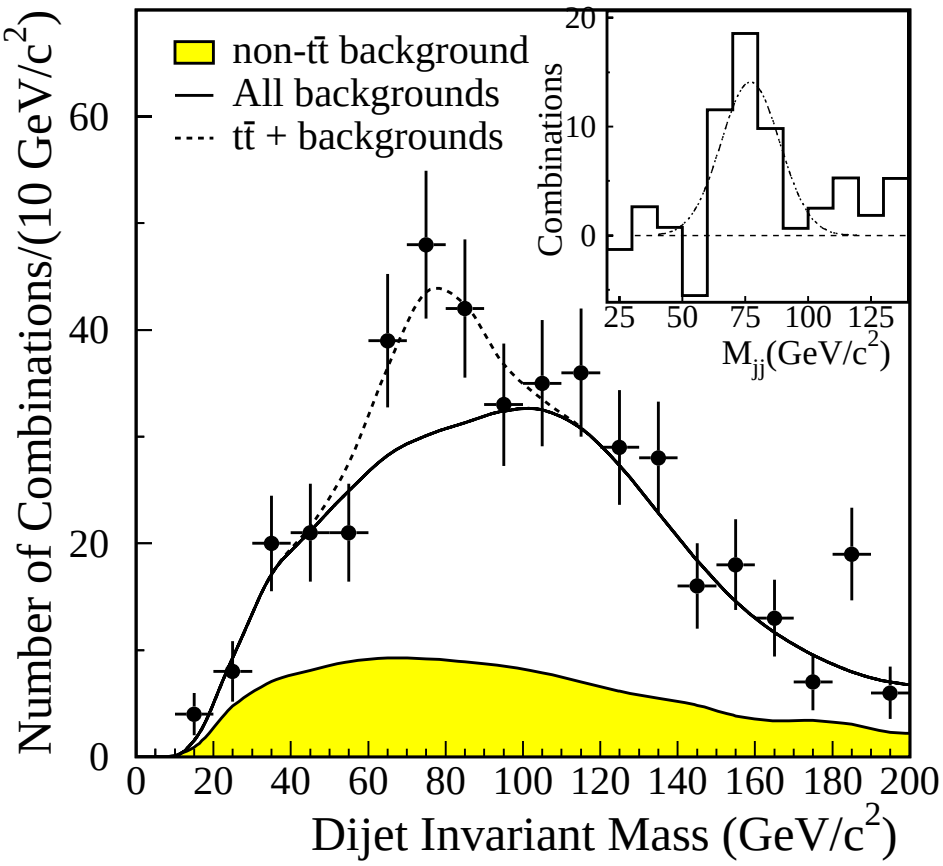


Figure 1.9: The invariant mass distribution of the jet pair presumed to originate from the hadronic decay of W in lepton + jets events with no b-tag requirement, but passing the cut $H_T > 310\ GeV$. The distribution plots all possible *combinations* in an event, so that a single event makes several entries in the distribution. The shaded region is the monte carlo distribution from background processes. The curve labeled "All backgrounds" combines to the shaded region the monte carlo distribution of top events with incorrect jet-parton matching. The dashed curve adds to the last curve monte carlo distribution of top events with correct jet-parton matching. The inset shows the background-subtracted dijet mass distribution. See [5].

where $< x >_{exp}$, $< x >_{back}$, and $< x >_{sig}$ are the average of x evaluated over the experimental data, background, and signal. An important point to note about this equation is the fact that it is valid *for all choices of* x, assuming that the $t\bar{t}$ hypothesis is valid.

A check of this sort has been performed in the lepton + jets channel by the CDF collaboration.[18] In this study, two samples of top candidate events were examined. The first sample is the so-called "$W + \geq 3$ jets" sample, which consists of one high-P_T e or μ, large $\not\!\!E_T$, and three or more jets. The second sample is a subset of the first, where one or more SVX b-tagged jet is required. The first sample contains $N_1 = 322$ events, with estimated background fraction $\beta_1 = (80 \pm 4)\%$; in the second sample, the corresponding numbers are $N_2 = 34$ and $\beta_2 = (25 \pm 5)\%$. Figures 1.10 and 1.11 show the result of comparing $< x >_{exp}$ with $< x >_{back}$ and $< x >_{sig}$, where, in the signal events, m_{top} was taken as 175 GeV. See [18] for details about all of the kinematic variables tested, and about how to read this plot. These plots show that, for all choice of x, $< x >_{exp}$ lie more or less where they are expected. Thus, within the precision attainable in Run 1, the kinematic features of the $t\bar{t}$ candidate events are consistent with the $t\bar{t}$ hypothesis.

The second type of test of the kinematic features of the $t\bar{t}$ candidate events requires the use of the $t\bar{t}$ reconstruction algorithm. An implementation of this algorithm for $t\bar{t}$ in the lepton + jets decay channel is described in detail in chapter 6. The algorithm operates, roughly, in the following manner. One first starts with a sample of $t\bar{t}$ candidate events. These events consist of one high-P_T isolated e or μ, large $\not\!\!E_T$, and four or more jets. The signal portion of these events are presumed to originate from the decay $t_\ell \rightarrow \ell + \nu_\ell + b_\ell$ and $t_h \rightarrow W_d + W_u + b_h$; these symbols are described in the following chart:

t_ℓ	Semileptonically decaying top quark
b_ℓ	b-quark from t_ℓ decay
ℓ	$e^\pm$ or $\mu^\pm$
ν_ℓ	Neutrino
t_h	Hadronically decaying top quark
b_h	b-quark from t_h decay
W_d	Down-type quark from hadronic W decay
W_u	Up-type quark from hadronic W decay

The reconstruction algorithm makes use of the fact that (1) the t and $\bar{t}$ masses are the same, and (2) $M_W \approx 80\ GeV$, in order to accomplish the following:

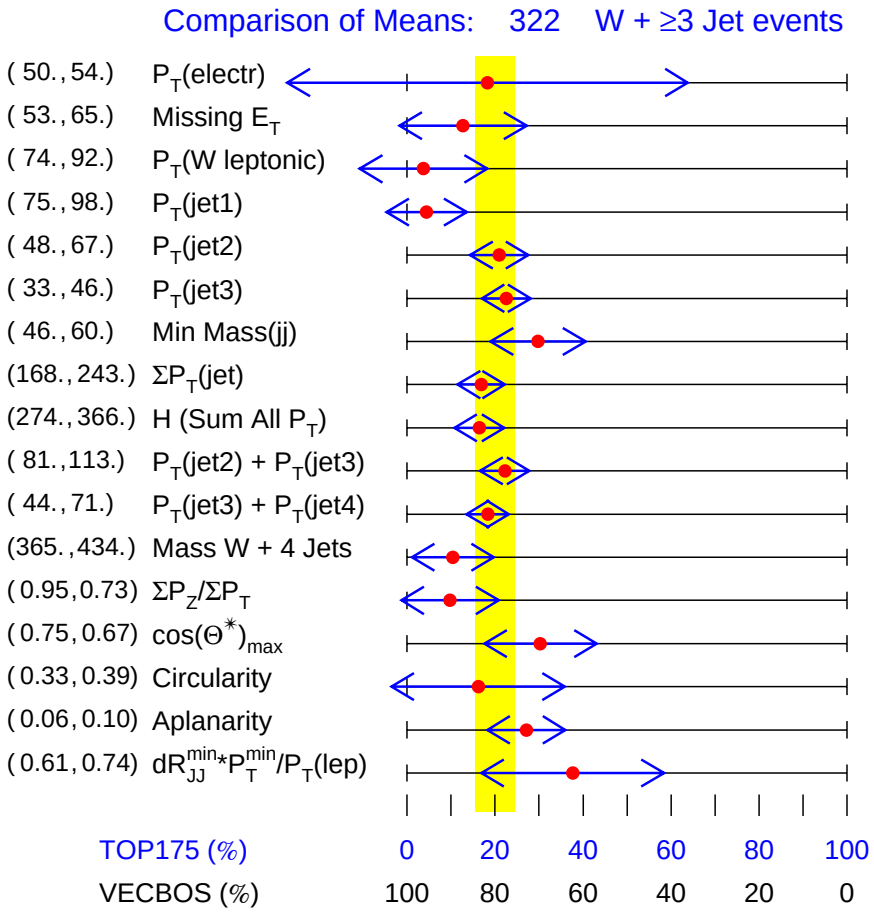


Figure 1.10: The mean of various kinematic quantities in the $W \geq 3$ jets data sample relative to the mean in the background (VECBOS) and signal (HERWIG $t\bar{t}$, $m_{top} = 175\ GeV$). The points indicate the mean in the experimental data, while the left (right) edge indicates the mean in the background (signal). The arrow in each row indicates the error of the mean. The shaded vertical strip indicates the most likely position of the mean, given the estimated background fraction $\beta = (80 \pm 4)\%$. See [18] for the definition of the kinematic variables and more on how to read this chart.

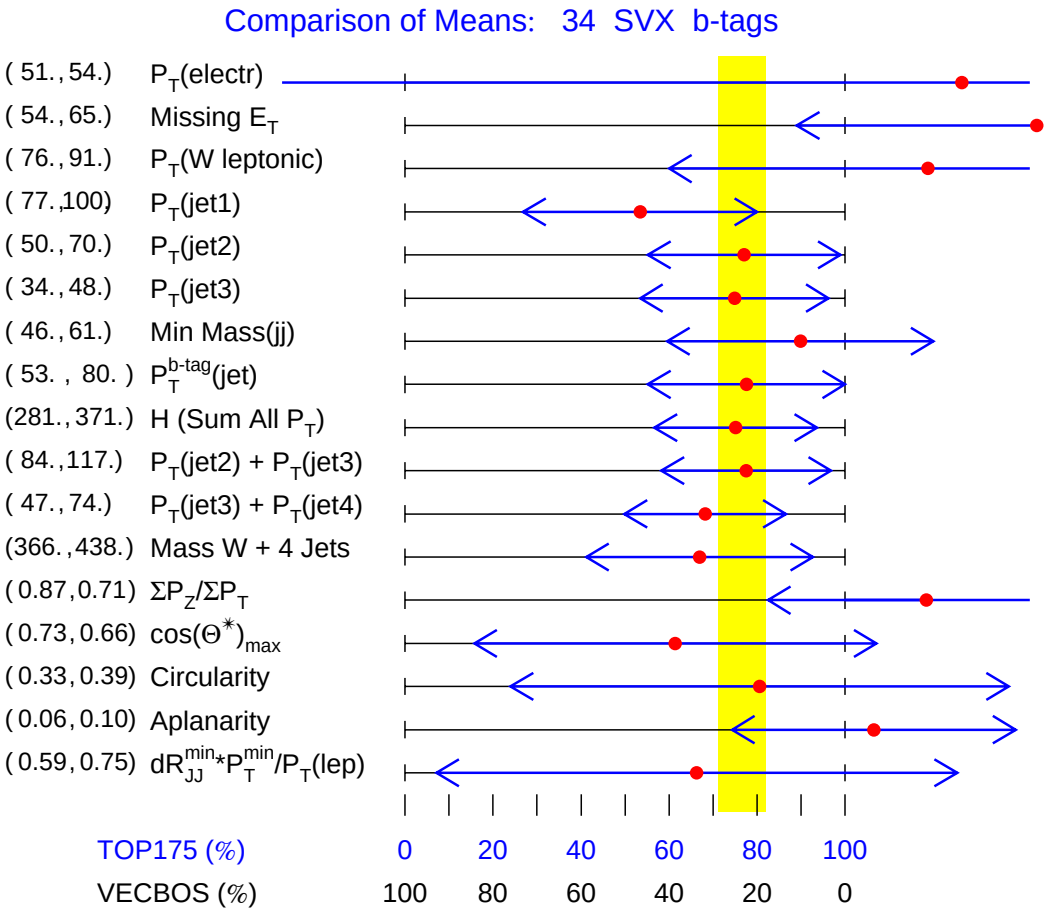


Figure 1.11: Same as figure 1.10, but taking the subsample of $W + \geq 3$ jets events with at least on SVX b-tagged jet. The background fraction in this case is $\beta = (25 \pm 5)\%$.

- Match the correct jets to the quarks that gave rise to them
- Reconstruct the jet energy
- Reconstruct the neutrino momentum

If the signal portion of the $t\bar{t}$ candidate events originate from $t\bar{t}$ alone, then the result of the application of this algorithm on the experimental data should be similar to an appropriate combination of results from the signal and background monte carlo. On the other hand, if the signal portion of the candidate events do not, or only partly, contain $t\bar{t}$ events, then this fact may show up as disagreements between the results from the experimental data and monte carlo.

Figures 1.12 to 1.16 show distributions of some variables from the output of the $t\bar{t}$ reconstruction algorithm. All except one variable appear consistent with signal + background monte carlo, with signal being $t\bar{t}$ with $m_{top} = 175\ GeV$. The one exception is the rapidity of the $t\bar{t}$ system, $y(t\bar{t})$ (figure 1.16). The rapidity distribution is somewhat depleted around $y = 0$. This problem, in fact, is more clearly manifest in the distribution of the longitudinal momentum of the $t\bar{t}$ system, $P_z(t\bar{t})$ (figure 1.17). A Kolmogorov-Smirnov test comparing the distribution of $|P_z(t\bar{t})|$ in the experimental data with that of signal + background monte carlo shows that the probability that the distribution in the experimental data originates from the signal + background monte carlo is 0.1%. As of this writing, this appears to be the only aspect of $t\bar{t}$ candidate events that shows statistically significant discrepancy with monte carlo. It is believed, therefore, that this anomaly is due to statistical fluctuation.

1.5 Confirming the Existence of the Top Quark

The Run 1 results make a strong case for the existence of the top quark. To be sure, test-by-test, the statistics are too limited to rule out the possibility that all or part of the signal portion of the $t\bar{t}$ candidate events originate from physics processes other than the production and decay of top quarks. However, it was seen in the last section that, test after test, the experimental data are consistent with the $t\bar{t}$ hypothesis. Although it may be

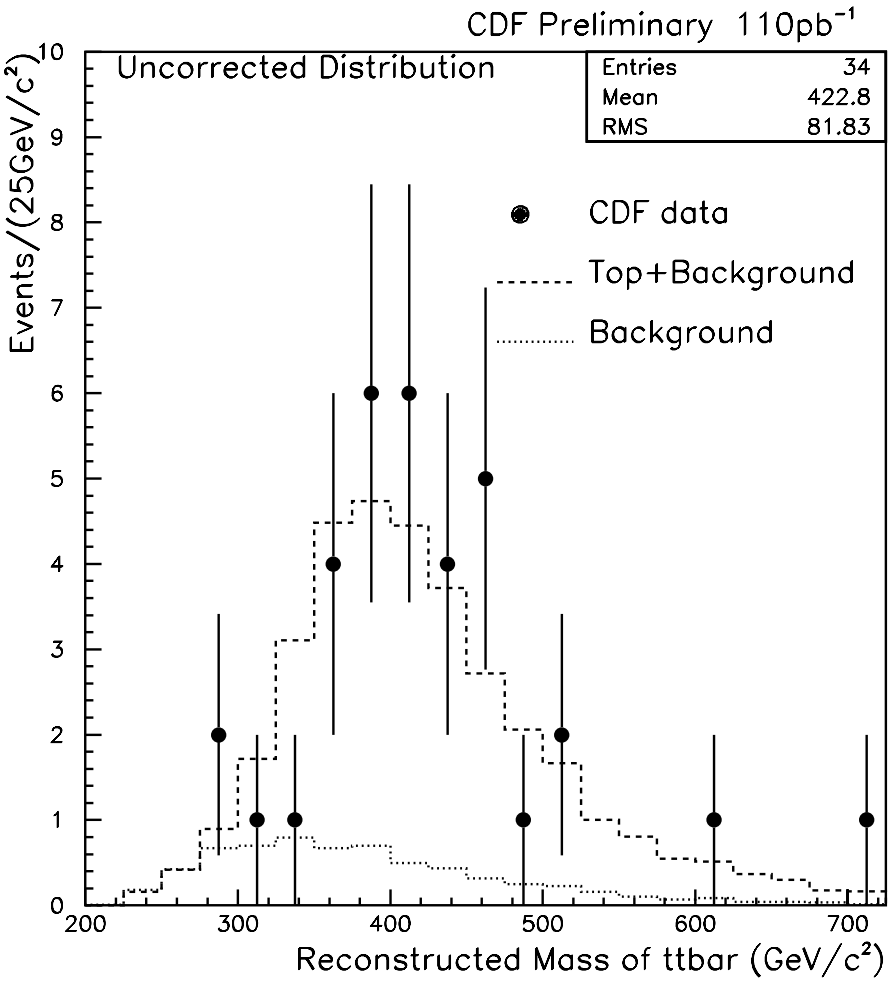


Figure 1.12: The $t\bar{t}$ invariant mass distribution in reconstructed $t\bar{t}$ candidate events in the lepton + jets channel with $\geq$ b-tagged jet.

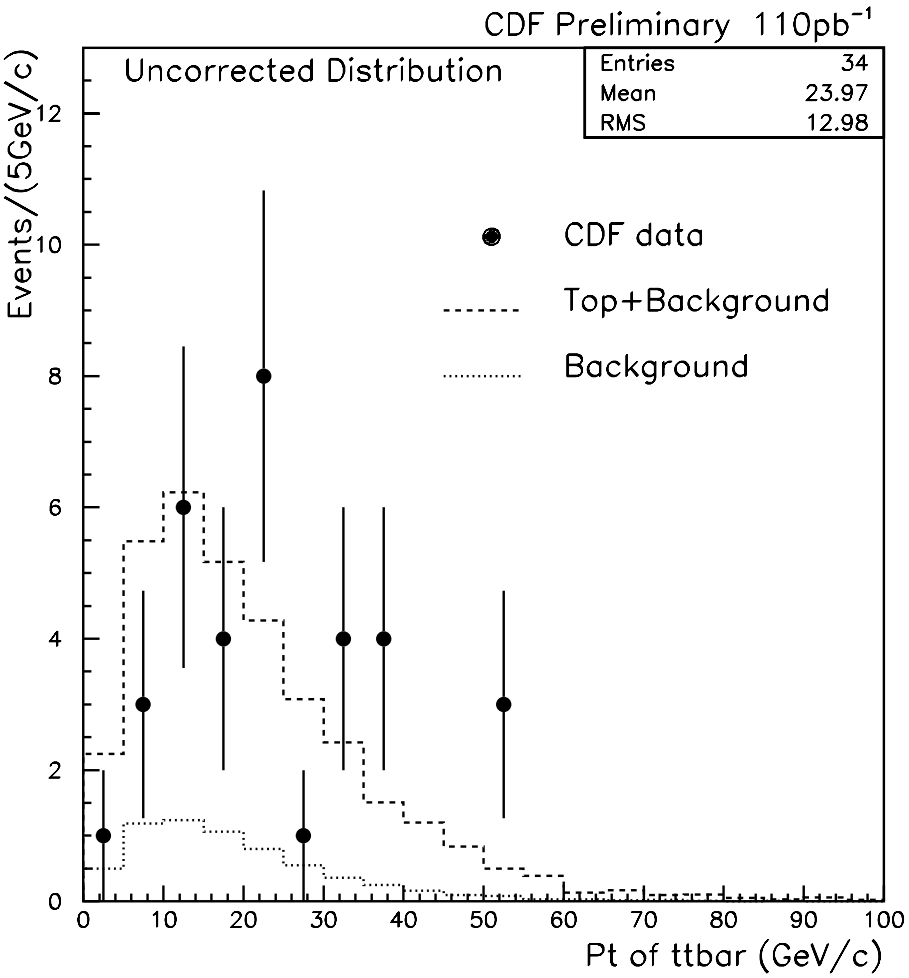


Figure 1.13: The distribution of the recoil transverse momentum of the $t\bar{t}$ system, $P_T(t\bar{t})$, in reconstructed $t\bar{t}$ candidate events in the lepton + jets channel with $\geq$ b-tagged jet.

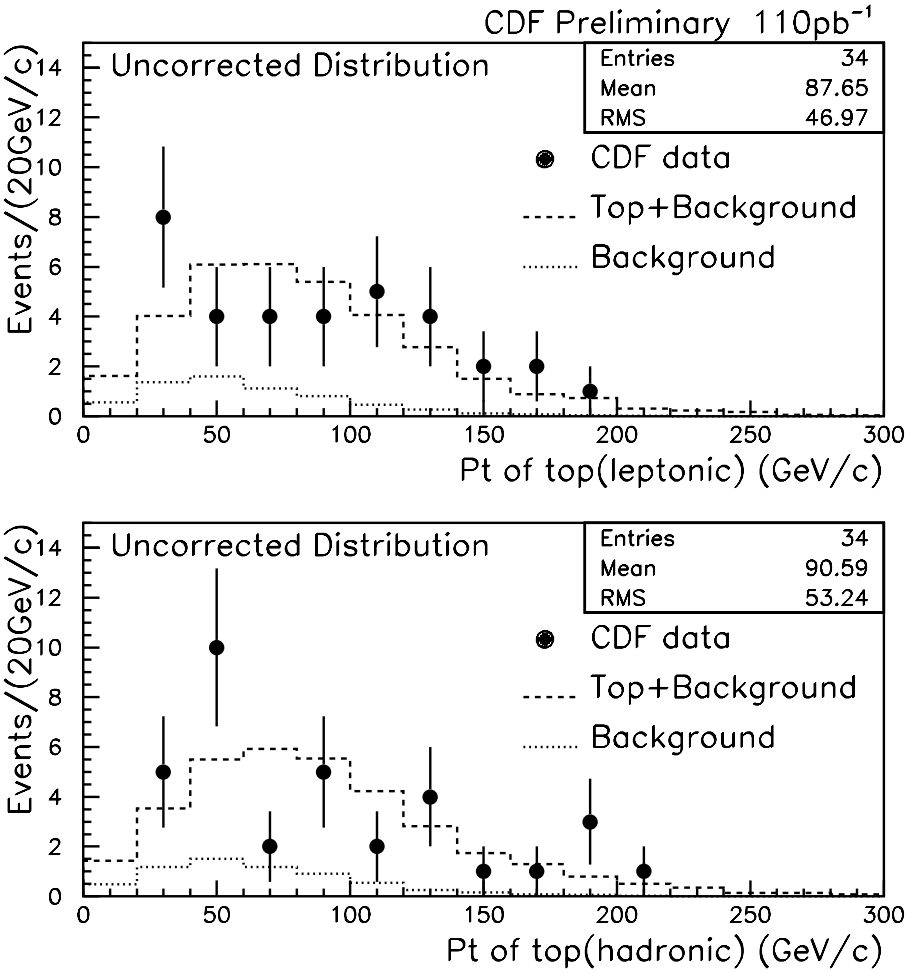


Figure 1.14: The distribution of P_T of the semileptonically and hadronically decaying top quark in reconstructed $t\bar{t}$ candidate events in the lepton + jets channel with $\geq$ b-tagged jet.

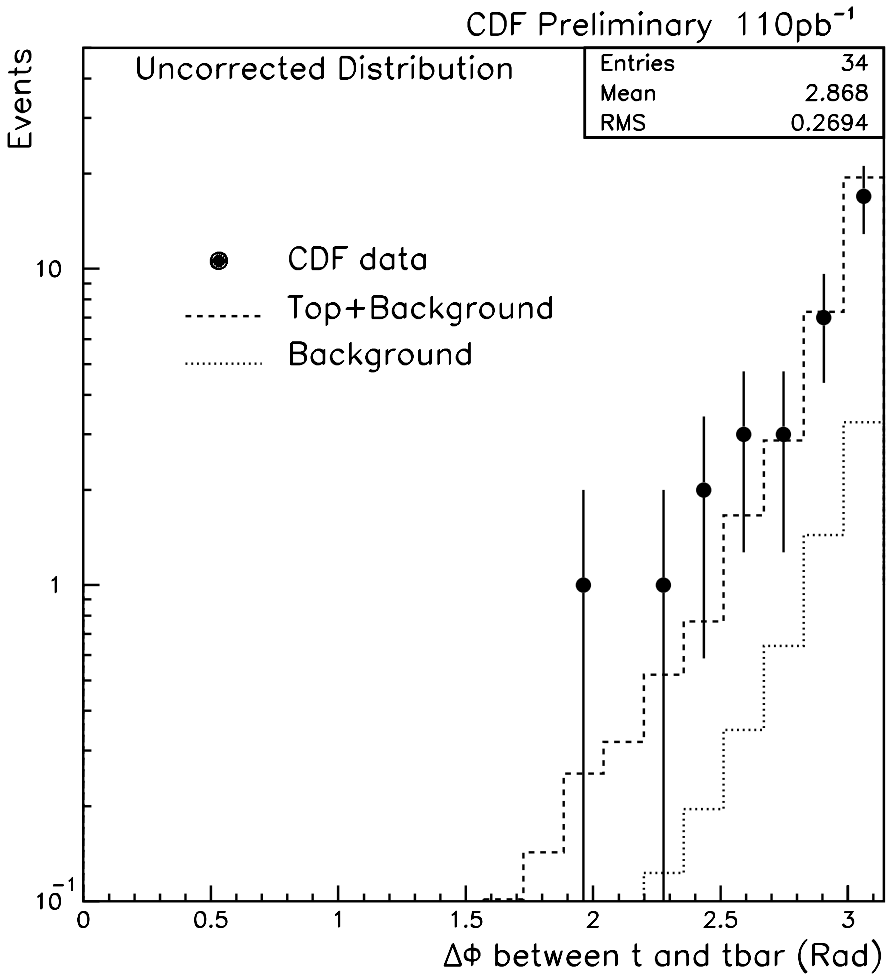


Figure 1.15: The azimuthal angular separation between t and $\bar{t}$ in reconstructed $t\bar{t}$ candidate events in the lepton + jets channel with $\geq$ b-tagged jet.

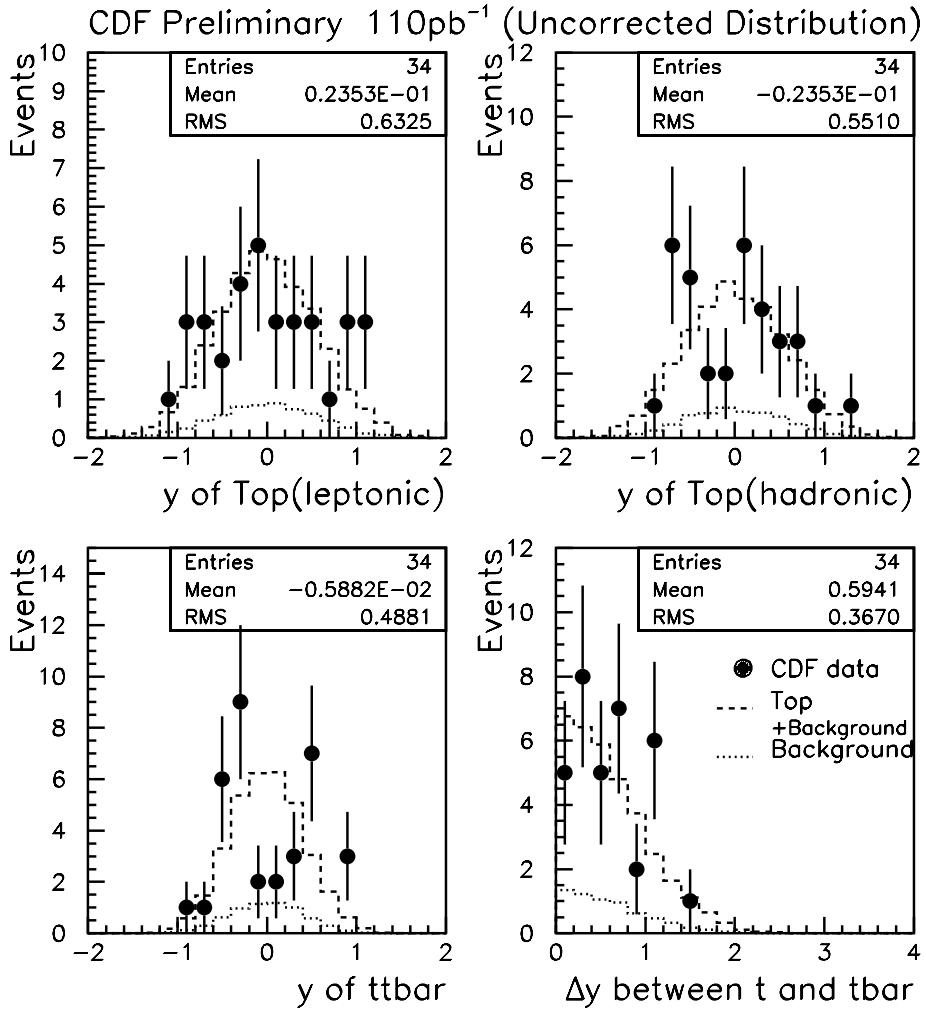


Figure 1.16: The rapidity of the semileptonically and hadronically decaying top quark, the rapidity of the $t\bar{t}$ system, and the rapidity separation between t and $\bar{t}$, in reconstructed $t\bar{t}$ candidate events in the lepton + jets channel with $\geq$ b-tagged jet.

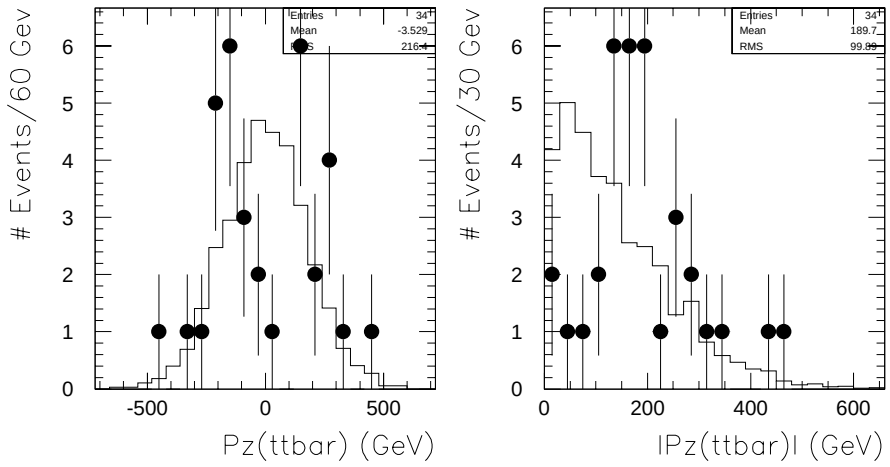


Figure 1.17: Distribution of $P_z(t\bar{t})$ and $|P_z(t\bar{t})|$ in $t\bar{t}$ in the lepton + jets events with ≥ 1 b-tag. The points show the distribution in the Run 1 experimental data; the histogram shows the signal + background monte carlo, with the signal being $t\bar{t}$ with $m_{top} = 175\ GeV$.

possible to think of new physics processes that could mimic the $t\bar{t}$ hypothesis at this level, it would seem to require a high degree of coincidence.

In Tevatron Run 2 and beyond, the question as to the existence of the top quark should be answered decisively. The following are some remaining tests that must be answered before one can conclude that the top quark does indeed exist:

Single-Top Production

In addition to the $p\bar{p} \rightarrow t\bar{t}$ production channel, top quarks can be produced singly at hadron colliders. Figure 1.18 shows the main diagrams responsible for single-top production. Single-top production was not observable during Tevatron Run 1 because of its small production cross section, small acceptance, and large background contamination. It is believed, however, that the integrated luminosity in Tevatron Run 2 will be large enough to overcome these difficulties. Observation of statistically significant excess beyond the expected contributions from known background processes will provide further evidence of the existence of the top quark. See [27], [30], and [64] to [70] for more on single top production.

The Spin of the Top Quark

The top quark is a spin-$\frac{1}{2}$ particle. Experimental confirmation of this fact is important in firmly establishing the existence of the top quark. The most direct way of testing the spin of the top quark is to measure the $t\bar{t}$ production cross section as a function of $\sqrt{s}$ in e^+e^- collisions. Such a test, however, will not take place in the near future. In the near-term, only indirect tests can be made. Two such tests probe: (1) the property of the t-b-W decay vertex; and (2) spin-correlation in $t\bar{t}$ decay. The first property can be probed by examining the distribution of the charged lepton energy; the second by examining the distribution of the angle between the two charged leptons in the dilepton decay channel. The details of both properties depend crucially on the fact that the top quark is a spin-$\frac{1}{2}$ particle. Negative results (*i.e.* the experimental data being inconsistent with monte carlo predictions) will not necessarily rule out the existence of the top quark, because they can also be produced by unexpected types of couplings/production mechanisms; they, however, will cast some doubt as to whether the particles presumed to be t and $\bar{t}$ really have spin-$\frac{1}{2}$, and, therefore, cast doubt as to the very existence of the top quark. These two properties are discussed in detail later in this thesis.

The Weak-isospin and Electric Charge of the Top Quark

The top quark has weak-isospin $T_3 = -1/2$ and electric charge $Q = +2/3$. Direct experimental confirmation of these facts will unequivocally establish the existence of the top quark. These properties, however, cannot be directly studied in any straightforward way other than in e^+e^- colliders. Thus these tests cannot be tested in the near future.

1.6 About this Thesis

In the last several sections, evidence for the existence of the top quark, and the remaining pieces of information needed to confirm its existence, have been shown. The work presented in this thesis also addresses the issue of the existence of the top quark. Specifically, it is addressed by examining the decay kinematics of the semileptonically decaying top quark in the top rest frame.

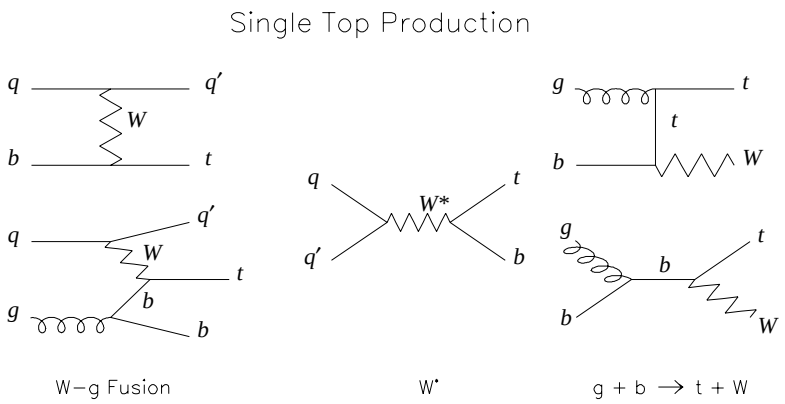


Figure 1.18: Main Feynman diagrams for single-top production.

It will be shown in chapter 7 that the standard model makes two non-trivial predictions concerning the center-of-mass kinematics of the semileptonically decaying top quark. They are the following:

Charged Lepton Energy Distribution

The distribution of the charged lepton energy in the top rest frame contains information about the nature of the t-b-W decay vertex. In particular, the standard model predicts that the decay vertex has $V - A$ coupling. One implication of this form of coupling is that the top quark almost completely decouples from the right-handed helicity state of the W boson, while its normalized coupling to the left-handed and longitudinal helicity states of W is $h_- = 0.30$ and $h_0 = 0.70$, respectively (taking $m_{top} = 175\ GeV$).

Charged Lepton Angular Distribution

The angular distribution of the charged lepton in the top rest frame contains information about top quark spin polarization. According to the standard model, the top quarks produced in $p\overline{p} \rightarrow t\overline{t}$ at $\sqrt{s} = 1.8\ TeV$ should be almost completely unpolarized. This means that the charged lepton angular distribution in the top rest frame is isotropic.

The work presented in this thesis examines the experimental data in light of these predictions. If the signal portion of the $t\overline{t}$ candidate events all originate from the production

and decay of $t\bar{t}$ pairs with $m_{top} \approx 175\ GeV$, then, within statistical error, the experimental data should be consistent with these predictions. If, on the other hand, part or all of the signal portion of the $t\bar{t}$ events originate from something other than $t\bar{t}$, then this fact could show up as energy and/or angular distributions that are inconsistent with the standard model predictions. (Of course, such inconsistencies can also indicate $t\bar{t}$ with anomalous couplings and/or production mechanisms.)

The comparison of the experimental data with standard model predictions is the ultimate goal of this thesis. Before this goal can be achieved, however, a lot of technical issues must be dealt with. The following is a list of the most important such issues:

$t\bar{t}$ Event Reconstruction

In order to examine the charged lepton energy and angular distribution in the top rest frame, the top quark momentum in the lab frame must be known. Obtaining this from the experimental data are not straight-forward, and requires a complicated algorithm, referred to as the *$t\bar{t}$ Reconstruction Algorithm*. Because the analysis depends crucially on this algorithm, it is important to understand how it works in detail (chapter 6). It is also important to understand what aspects of the algorithm are most responsible for the degradation of the measurement resolution (chapter 9).

Choice of Observables

In the semileptonic decay channel, the top quark decays to $\ell + \nu_\ell + b$, where ℓ is a charged lepton, ν_ℓ is the neutrino partner of ℓ, and b is the b-quark. At first thought, it is not clear why one should examine only the energy and angular distribution of ℓ. A careful examination of the standard model prediction of this process makes clear that: (1) the energy and angle of b and ν_ℓ provide no information independent of that provided by the energy and angle of ℓ; and (2) the energy and angle of ℓ have properties that make them the most desirable quantities for examining the top center-of-mass decay kinematics. This sort of question must be answered before the experimental data can be compared to theoretical predictions in an efficient manner. These issues are discussed in chapter 7.

Measurement Method and Resolution

A measurement method must be chosen before one can quantify the degree of agreement between the experimental data and standard model prediction. Once a method is chosen, it is useful to know what factors affect the measurement resolution. These issues are dealt with in chapter 8.

It is shown in chapter 8 that, with Run 1 statistics, the statistical errors are too large to draw any meaningful conclusion from the comparison of the experimental data with standard model predictions. Thus, in this thesis, the technical issues are emphasized. This thesis is intended mainly as reference material for those who might perform related analyses in Run 2.

Chapter 2

Experimental Apparatus

The two main experimental apparatus in a collider experiment are: (1) a collider; and (2) a detector. In this chapter, the Run 1 Tevatron Collider and the CDF detector are described.

2.1 The Run I Tevatron Collider

The Tevatron collider is located at the Fermi National Accelerator Laboratory (FNAL) in Batavia, Illinois (west of Chicago). It has a radius of 1 km, and uses superconducting magnets to collide protons and antiprotons at a center-of-mass energy of $\sqrt{s}$ = 1.8 TeV. A schematic diagram of the Tevatron collider complex is shown in figure 2.1. In this section, the operation of the collider is sketched, and the parts that make up the collider complex is described.

2.1.1 A Sketch of the Collider Operation

The Tevatron collider has two main modes of operation: the collider mode and fixed-target mode. The collider mode is of interest in this discussion. In the collider mode, the Tevatron collider complex is comprised of six parts:

1. The Pre-acc
2. The Linac
3. The Booster

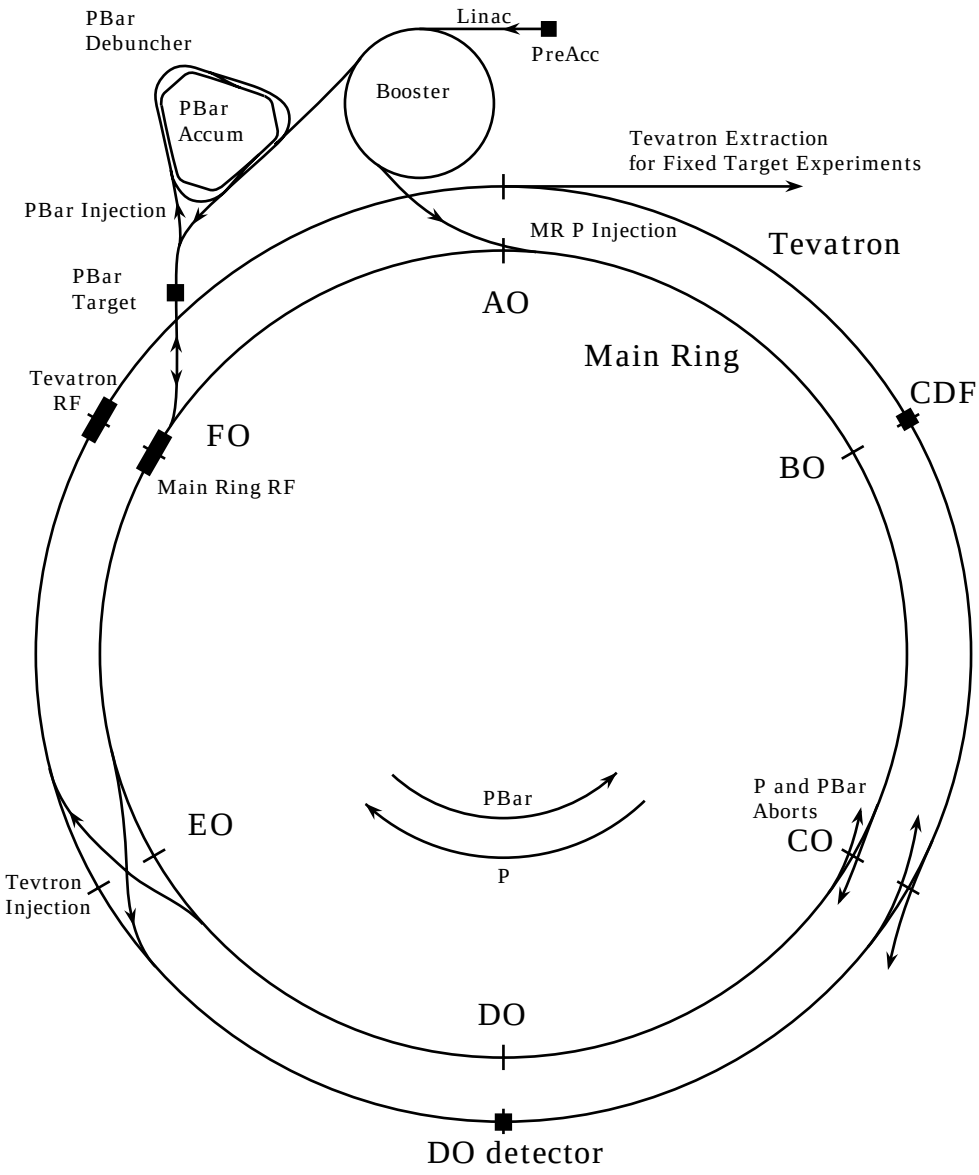


Figure 2.1: A schematic diagram of the Tevatron collider complex. This figure is adopted from [19].

4. The Main Ring

5. The Antiproton Complex

6. The Tevatron

These parts will be described in detail in the next section. In this section, the roles played by the parts are sketched.

The ultimate goal of the Tevatron collider operating in the collider mode is to get protons and antiprotons to collide at $\sqrt{s} = 1.8\ TeV$. This goal is achieved in three steps: (1) create and accumulate $\overline{p}$; (2) shot-setup (preparation for collisions); and (3) collisions. The following happen in each of the steps:

Antiproton Creation and Accumulation

Antiprotons are created by smashing protons into a nickel target. The source of the protons used for this purpose is ionized hydrogen gas, H^-. The H^- are extracted from the source and accelerated to 750 keV by the "Pre-acc", which is a Cockroft-Walton electrostatic accelerator. These H^- ions are then fed to the "Linac", which is a collection of radio-frequency (RF) chambers that are strategically placed to accelerate the incoming H^- ions to 400 MeV. The outgoing H^- ions are then forced through a carbon foil, which removes the electrons from H^- and passes through only the proton in the core. These protons are then fed into the Booster, which is a synchrotron of diameter of about 150 m that accelerates the protons to 8 GeV. The protons are then fed into the Main Ring, which is also a synchrotron, but a much larger one with radius of 1 km. The Main Ring accelerates the protons to 120 GeV. They are then extracted and brought into collision with a nickel target. In each collision, there are about 10^{12} protons impinging upon the target, and about 10^7 antiprotons coming out of the target. The outgoing $\overline{p}$ are then cooled in the debuncher and stored in the accumulator. Using this procedure, about 10^{10} $\overline{p}$ can be produced per hour. A viable rate of collisions requires about 10^{11} $\overline{p}$, so the accumulation process typically takes about half a day to complete.

Shot-setup

During collision, six bunches of protons and antiprotons are accelerated in opposite directions in the Tevatron collider, each p and $\overline{p}$ carrying about 900 GeV of energy. The process of getting the p and $\overline{p}$ into this state is referred to as shot-setup. The first step in shot-setup is the transfer of 150 GeV protons from the Main Ring to the Tevatron. The source of the protons is the same as for $\overline{p}$ creation described above. The protons are injected into the Tevatron bunch-by-bunch. Once injected, each bunch typically contain about 2×10^{11} protons.[1] Typically, it takes about 30 sec to inject a proton bunch. Once six bunches are injected into the Tevatron, the process is repeated with $\overline{p}$ bunches. The process for $\overline{p}$ proceeds in very much the same way as for p. One important difference between p and $\overline{p}$ is the bunch size: $\overline{p}$ bunches typically contain only about 6×10^{10} antiprotons. Once the Tevatron is injected with six bunches each of p and $\overline{p}$, they are "ramped up" to 900 GeV. This process usually takes tens of seconds. At this point, very few collisions are taking place because the transverse size of the beam is relatively large. The beam is "squeezed" to a diameter of roughly 100 micron, at which point there are typically about 2.5 interactions per beam crossing. The final step in shot-setup is "scraping", whereby p and $\overline{p}$ in the tail of the bunch distribution are removed; this step is taken in order to protect sensitive equipment in the detector from being damaged by excessive bombardment by particles. Typically, a trouble-free shot-setup takes about an hour to complete; complications in any of the above steps could delay collisions for many hours, or longer.

Collisions

Once the beam is scraped, the detectors are activated. This activation process takes a minute or two, and principally involves turning on high-voltage power sources for various detector components. Once the detector is up and running, the triggers in the detector are activated, and data-taking proceeds. A collision session is referred to as a store. A typical store lasts about 10 to 20 hours. During the course of a store, the beam intensity steadily decreases because of emittance effects. Thus the rate at which data are collected also steadily decreases. Normally, a store is ended when the

[1] See [105] for a table describing beam parameters. The numbers quoted here are for Run 1b.

beam luminosity gets below a certain threshold. On occasion, however, malfunction in some component in the accelerator could force an end to a store. It should be noted that during a store, the Main Ring is often operating in parallel, creating and accumulating $\overline{p}$ in preparation for the next store.

2.1.2 The Parts of the Collider

In the collider mode, the Tevatron collider complex consists of the Pre-acc, Linac, Booster, Main Ring, Antiproton Complex, and the Tevatron. These components are described in this section. (The sources for the material presented in this section are from [102], [103], [104], [106], [107], [108], [109], [110].)

2.1.2.1 The Pre-acc

The Pre-acc system consists of an H^- ion source and a Cockroft-Walton electrostatic accelerator that accelerates H^- to 750 keV.

Gaseous hydrogen is first extracted from a small tank and injected into the ion source, out of which emerges H^- ions (figure 2.2). These ions are extracted from the source at 18 keV and transferred to a Cockroft-Walton electrostatic pre-accelerator (the Pre-acc), which accelerates H^- to 750 keV (figure 2.3). The acceleration process can be thought of roughly in the following terms: the dome containing H^- ions is held at an electric potential of -750 kV; a column connects the dome to a ground potential, and the ions rush toward through the column to achieve the final energy of 750 keV.

According to nominal specifications [103], the source-accelerator system gives rise to pulses of H^- of current of $\approx$ 50 mA and pulse length of 30 μs. These pulses can be created at a rate of 15 Hz, or $\approx 1.4 \times 10^{14}$ H^- ions/sec.

2.1.2.2 The Linac

The Linac takes as input H^- ions at 750 keV and accelerates them to 400 MeV. The Linac consists of fourteen cylindrical radio-frequency (RF) accelerating cavities arranged collinearly. The first five cavities are drift-tube linacs operating at 201 MHz; they accelerate H^- to 116 MeV. The last nine cavities are side-coupled cavity linacs operating at 805 MHz;

Figure 2.2: The H^- ion source. The small tank in the upper-left contains hydrogen gas. The machinery to the right of center is the ion source. The photograph is from the Fermilab Accelerator Division/Proton Source Department World Wide Web page [107].

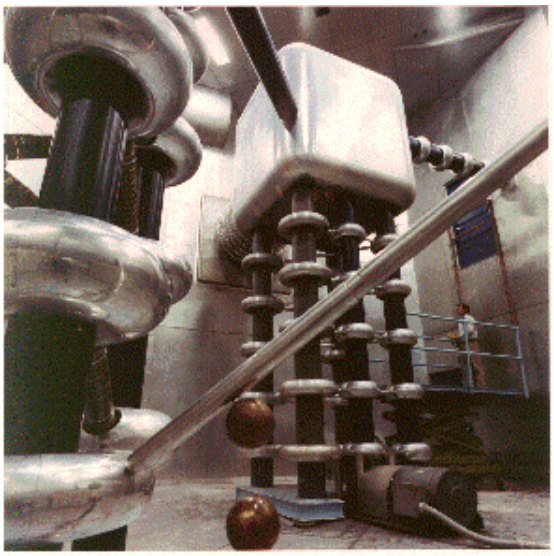


Figure 2.3: The Cockroft-Walton 750 kV dome. This is where the H^- ions are accelerated to 750 keV. The photograph is from the Fermilab Accelerator Division/Proton Source Department World Wide Web page [107].

Figure 2.4: The inside of one of the Drift Tube Linac at Fermilab. The photograph is from the Fermilab Accelerator Division/Proton Source Department World Wide Web site [107].

they accelerate H^- to about 400 MeV. The linac system can accelerate a beam at a rate of about 15 Hz (the same rate as the H^- source).

Figure 2.4 shows a photograph of the inside of a drift-tube linac. The operation of a drift-tube linac goes along the following lines. Amplified RF is applied to the drift-tubes in the linac cavities. This sets up an alternating, high-gradient ($\sim$ 2 MV/m) electric field between the drift tubes, while the interior of the drift-tubes are almost free of electric field. The drift-tubes are strategically placed so that the H^- ions are outside of a drift-tube when the electric field is along the desired direction of motion, and lies in the drift tube when the electric field is in the wrong direction.

Figure 2.5 shows a photograph of the side-coupled linac. The operation of a side-coupled linac goes along the following lines. Imagine a cylindrical pipe partitioned along its length with walls (somewhat like a stack of cans) with a hole in the center, through which H^- ions travel. A pipe is connected to amplified RF generated by klystrons (basically, a very powerful microwave generator). The applied RF creates a standing electromagnetic wave patterns in the cavities of the pipe. The geometry of the pipe and the cavities therein are designed in such a way that H^- ions travelling along it are accelerated by the electric field set up in the cavities.

Figure 2.5: One of the Side-Coupled Linacs at Fermilab. The photograph is from the Fermilab Accelerator Division/Proton Source Department World Wide Web site [107].

2.1.2.3 The Booster

The Booster (figure 2.6) is a circular accelerator of radius of about 75 m. It is an alternate gradient synchrotron that accelerates protons from 400 MeV to 8 GeV. The acceleration is performed by 17 RF cavities (similar to the drift-tube linacs described earlier) placed in nine straight sections of the Booster (two per straight section, except for one of the straight sections, which has room for only one). The protons reach their peak energy after about 20,000 rotations. The Booster is capable of accomplishing this at a rate of 15 Hz (the same rate as the H^- source and the Linac).

An important auxiliary mechanism of the Booster is the carbon foil that strips off the electrons from the incoming 400 MeV H^- ions. The following is a description of how the stripper works. First, a bunch of H^- ions from the linac is injected parallel to the Booster orbit; the ions are about 8 cm outside of this orbit. The H^- ions and the protons travel in a parallel path, until their paths are bent by a dipole magnet. The paths of the H^- ions are bent inward and that of the protons are bent outwards (relative to the center of the Booster). Their paths are bent so that the H^- and proton bunch merge. The merged beam is then sent through the carbon foil. About 98 to 99 % of the the H^- come out as

Figure 2.6: A photograph of the Booster and other structures at the Fermilab $p\overline{p}$ collider. The Booster is located under the small circular structure shown toward the left-center. The building behind the Booster is the 15-story Hi-Rise. Underneath the triangular structure in front of the Booster is where the antiproton Debuncher and Accumulator are. The underneath the large circular structure to the right is where the Main Ring and the Tevatron are. This photograph is from the Fermilab Accelerator Division/Proton Source Department World Wide Web site [107].

protons, while most of the incoming protons pass right through. The outgoing protons are then bent back to the original Booster orbit.

2.1.2.4 The Main Ring

The Main Ring is a circular accelerator that accelerates protons from the Booster and antiprotons from the Accumulator. It is an alternate gradient synchrotron of radius 1000 m, and is located in a tunnel, which has a diameter of about 3 m (10 ft) and is 6 m (20 ft) underground. The initial particle energy is 8 GeV; the Main Ring is capable of increasing that to 120 GeV or 150 GeV, depending on the beam's destination. This can be accomplished at a rate of once every 2.4 sec.

The Main Ring is made of 774 bending (dipole) magnets, 240 quadrupole (focusing) magnets, and 18 RF cavities for particle acceleration. The RF cavities are all located in a single straight section. All of the magnets are conventional, copper-coiled magnets.

The Main Ring is divided into six sectors, labeled A through F, each subtending an angle of 60^o (figure 2.1). Each sector is further divided into five parts, labeled 0 through 4. Part 0 is a straight section; parts 1 through 4 contain bending and focusing magnets. The six straight sections ($A0$ to $F0$) each serve special functions:

A0

This is the injection point for protons from the booster.

B0

This is where the CDF detector is.

C0

This is where p and $\overline{p}$ are dumped when the beam is aborted.

D0

This is where the D0 detector is.

E0

This is where p and $\overline{p}$ are injected from the Main Ring to the Tevatron.

F0

This is where the 18 RF cavities are. This is also where protons destined for the antiproton targets are extracted, and where $\overline{p}$ from the Accumulator is injected into the Main Ring.

The Main Ring operates in two modes: one for creating $\overline{p}$, and one for injection into the Tevatron. In the first mode, the Main Ring accepts 8 GeV protons from the Booster and accelerates them to 120 GeV. The resulting beam is extracted at $F0$ and sent to the $\overline{p}$ target. In the latter mode, the Main Ring extracts 8 GeV protons from the Booster, or 8 GeV antiprotons from the Accumulator, and accelerates them to 150 GeV. The 150 GeV p or $\overline{p}$ are then injected into the Tevatron at $E0$.

2.1.2.5 The Antiproton Complex

The antiproton complex consists of three main parts: the target, the debuncher, and the accumulator. The basic operating principle of this complex is as follows. First, protons accelerated to 120 GeV in the Main Ring are extracted and are brought into collision with the target. Many different types of particles are created by this collision, among which are antiprotons. The cone of produced particles go through a lithium lens, which renders the particle trajectories nearly parallel. These trajectories are then bent by a dipole magnet in order to select $\overline{p}$ with approximately 8 GeV of energy. The extracted $\overline{p}$ are then placed in the Debuncher, in which the momentum spread of the 8 GeV beam is reduced. The debunching process is continued until just before the next batch of protons is extracted from the Main Ring. At that time, the $\overline{p}$ are transferred from the debuncher to the Accumulator, where the $\overline{p}$ are stored at 8 GeV. This process is continued until sufficient $\overline{p}$ are collected for shot-setup. During shot-setup, $\overline{p}$ in the accumulator are extracted and injected into the Main Ring at 8 GeV.

The target in the antiproton complex is a nickel disk of diameter $\approx$ 10 cm and thickness $\approx$ 2 cm. A typical target assembly is shown in figure 2.7. The beam hits the target sidewise, as shown in figure 2.8.

The lithium lens is a cylinder of lithium 15 cm in length and 1 cm in diameter. It is placed in a toroidal transformer that produced magnetic field inside of the lens. Lithium was

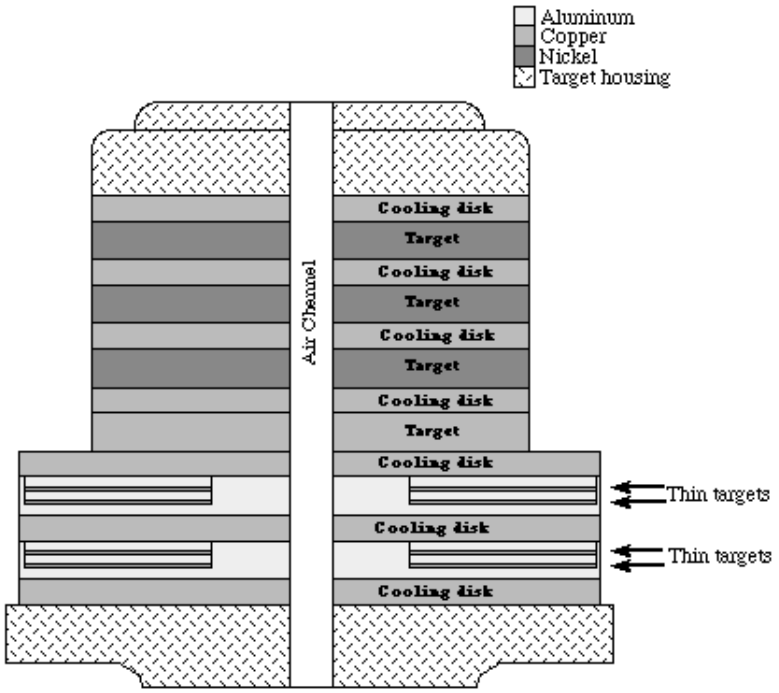


Figure 2.2 Typical target assembly

Figure 2.7: Beam's view of the $\overline{p}$ target assembly. This figure is from the "Antiproton Rookie Book" at the Fermilab Accelerator Division/Antiproton Group World Wide Web site [110].

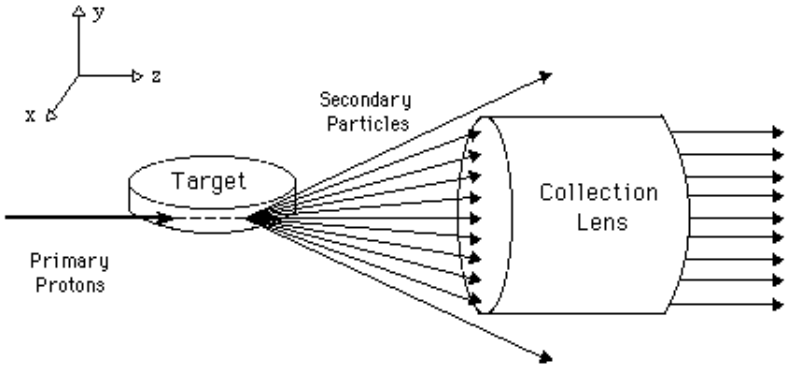


Figure 2.3 Focusing of secondary particles by the Collection Lens

Figure 2.8: The proton beam hits the nickel target disk sidewise and off-center to create antiprotons. Also shown is the lithium lens used to make the secondary particle trajectories parallel to each other. This figure is from the "Antiproton Rookie Book" at the Fermilab Accelerator Division/Antiproton Group World Wide Web site [110].

chosen for the lens because it is the least dense solid conductor; this minimizes scattering and absorption of the traversing particles.

The dipole magnet downstream of the lithium lens selects $\overline{p}$ having momentum of about 8 GeV. Various factors (both physics and engineering) are involved in the choice of this momentum. One of the physics-based factors is the fact that the $\overline{p}$ momentum spectrum peaks at about 8 GeV, so that this choice of momentum optimizes the collection rate of $\overline{p}$.

The debuncher is rounded-triangular shaped synchrotron with a mean radius of about 90 m (figure 2.6). The accumulator is somewhat smaller than the debuncher (mean radius $\approx$ 75 m), and is located within the perimeter of the debuncher (figure 2.9). The debunching process, whereby the momentum spread of the $\overline{p}$ is reduced, is referred to as "cooling". Various cooling methods are used to reduce both the longitudinal and transverse momentum spread; they are based on electronic feedback circuits set up in the ring. The $\overline{p}$ beam in the Debuncher is cooled for two reasons: (1) to make cooling in the Accumulator more efficient; and (2) to maximize the efficiency in transferring $\overline{p}$ from the Debuncher to the Accumulator. The beam is cooled in the Accumulator in order to: (1) efficiently store the beam for many hours; (2) efficiently transfer low-emittance (*i.e.* narrow and dense) beam to the Main Ring; and (3) to achieve high $p\overline{p}$ luminostiy.

Figure 2.9: Inside the tunnel containing the Debuncher (left) and the Accumulator (right). This photograph is from the Official Fermilab World Wide Web site [106].

2.1.2.6 The Tevatron

The Tevatron is an alternate gradient synchrotron of radius 1 km. It is housed in the same tunnel as the Main Ring (it lies just beneath the Main Ring, see figure 2.10). It has eight RF cavities at $F0$ that are used to accelerate p and $\overline{p}$ from 150 GeV to 900 GeV. The Tevatron has about 1000 superconducting magnets that generate strong enough magnetic field to keep such high energy particles in a circular orbit. In figure 2.10, the Tevatron magnets are contained in long, rectangular housing cases. Inside the housing are concentric cylinders containing, from outside in: (1) a vacuum shell; (2) a liquid nitrogen pipe; (3) another vacuum pipe; (4) a liquid helium pipe; (5) superconducting magnetic coils; and, finally, (6) the beam pipe.

Figure 2.10: The tunnel containing the Main Ring and the Tevatron. The tunnel is about 3 m wide and 2.4 m high. The Main Ring is on top of the Tevatron. This photograph is from the Official Fermilab World Wide Web site [106].

2.2 The Run I CDF Detector

2.2.1 General Description

The Collider Detector at Fermilab (CDF) is one of two general-purpose cylindrical detectors used to detect particles produced in $p\overline{p}$ collisions at the Tevatron Collider (the other detector is the D0). References [78] and [79] describes CDF in detail. Here, a brief discussion about it will be given for reference.

The complete detector complex — the detector itself, the readout electronics, the trigger, the data acquisition system, and computers for system monitoring — is contained in a three story hangar-like structure with a basement that reaches about 30 ft underground (figure 2.11). The detector proper is located in the underground portion of the complex, so that the Tevatron beam line (which is located about 6 m (20 ft) underground) goes through the detector's axis.

In order to describe the detector components, it is useful to first define coordinate axes. Figure 2.12 shows both cartesian and cylindrical coordinate systems defined relative to the detector. In the cartesian system, the z direction points along the proton beam (to the east through the detector axis, into the photo in figure 2.11), the x direction points north (to the left of the photo in figure 2.11), and the y direction points upwards. In the cylindrical system, the z axis is the same as before, the radial axis points outwards from the detector axis, and the azimuthal angle is defined with 0^o pointing along the x axis, and azimuth increasing toward the $+y$ axis.

The CDF detector is segmented into five parts along the z direction: east forward, east plug, central, west plug, and west forward. The most critical portion of the detector for top quark physics is the central detector, which is needed to identify energetic electrons and muons. The central detector is segmented into different parts going along the radial direction. In increasing radius, one has the beam pipe, the tracking detectors, a superconducting magnetic solenoid coil, the central calorimeters, and the muon detectors.

The tracking detectors, going from small to large radii, are the Silicon Vertex Detector (SVX), the Vertex Time Projection Chamber (VTX), and the Central Tracking Chamber (CTC). These tracking elements are contained in a nearly constant axial magnetic field of

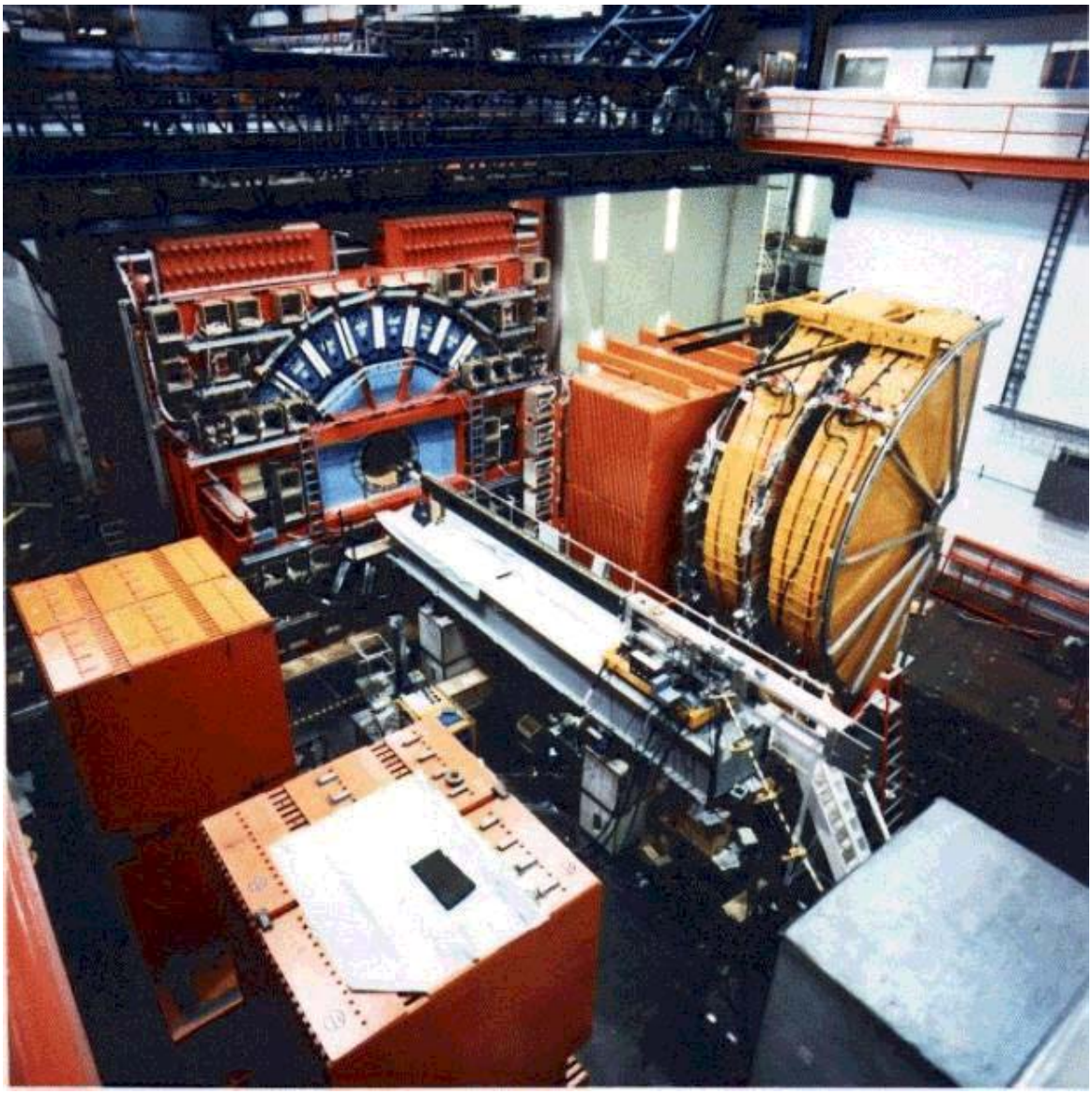


Figure 2.11: A photograph of the CDF detector. A crane (upper portion of photo) holds the central portion of the detector; in the foreground are the forward detectors. In the photo, the detector components are pulled out from their normal positions. During normal operating conditions, the central portion of the detector is carried by the crane into the basement, and the forward detectors are placed close in front of the central detectors. Photo from the CDF World Wide Web site [111].

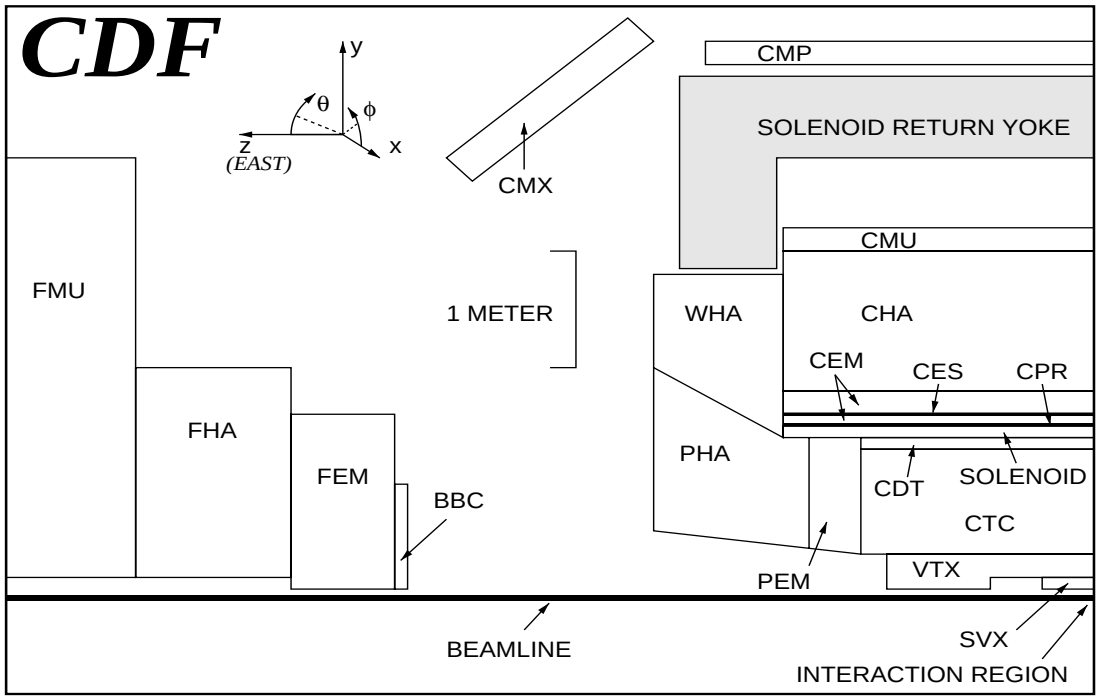


Figure 2.12: A one-quarter view of the CDF detector. In order to picture the complete detector, one should imagine rotating the picture about the beam axis, and reflect it about the vertical plane through the nominal interaction point ($z = 0$). This figure is adopted from [16].

about 1.4 T generated by a superconducting magnetic solenoid. The CTC is basically a magnetic spectrometer that measures the momenta of charged particle tracks by measuring their curvature in the magnetic field.

The central calorimeters are divided between the electromagnetic (CEM) and the hadronic (CHA). The CHA is placed radially outside of the CEM. This separation between electromagnetic and hadronic calorimeters allows one to distinguish electrons and photons from hadronic jets.

The muon detectors consist of the Central Muon Detector (CMU), the Central Muon Upgrade Detector (CMP), and the Central Muon Extension Detector (CMX). The CMU is located immediately outside of the hadronic calorimeter, while the CMP is located beyond CMU, with 0.6 m of steel in between. The steel layer absorbs most of the hadronic "punch-through" particles from the CHA while allowing most muons through. The CMX is located somewhat to the side of the central portion of the detector and detects muons that escape CMU and CMP coverage.

In addition to the central detectors, there are the plug and forward detectors. Specifically, there are the plug and forward calorimeters and the Forward Muon detector (FMU). The FMU is not used in the analysis of top quark events. The plug and forward calorimeters are segmented, like the central calorimeters, into electromagnetic (PEM and FEM) and hadronic portions (PHA and FHA). These calorimeters are used only to detect jets and contribute to the calculation of the missing transverse energy ($\not\!\!E_T$) and the unclustered energy.

Finally, there are the trigger and data acquisition systems in the three story B0 detector building. The trigger has three levels and is used to select subsets of $p\overline{p}$ collision events that are likely to have originated from physics processes of interest. The data acquisition system, working in unison with the trigger, collects, organizes, and stores information from all detector components.

2.2.2 The Parts of CDF

In this section, the parts of the CDF detector that are most important in the detection of the top quark are described. The operating principle of the detector components is beyond the scope of the present discussion; for that information, see [4].

2.2.2.1 Tracking Detectors

The SVX, VTX, and the CTC comprise the tracking system.

The CTC [80], the outermost tracking detector, is a wire drift chamber with 84 concentric cylindrical layers of sense wires. It is 3.2 m long, centered at $z = 0$, and extends from a radius of 31.0 cm to 132.5 cm. The sense wires are organized into nine "superlayers" (*i.e.* cluster of layers). There are five axial superlayers containing twelve wire layers each, and whose wires are parallel to the z axis. Interleaved between the axial superlayers are the four stereo superlayers containing six wire layers each, and whose wires are tilted by $\pm 3^o$ with respect to the beam axis. The axial superlayers play a key role in the reconstruction of track momentum in the r-ϕ plane, while the stereo layers allow one to reconstruct the z component of track momentum. The CTC covers the detector pseudorapidity region $|\eta_d| < 1$.[2] The CTC measures track P_T with a resolution of $\delta P_T/P_T = 0.0011 \cdot P_T$.

The VTX [81] is located between the CTC and SVX. Its main role is to provide r-z tracking up to radius 22 cm and $|\eta_d| < 3.25$. This information is used to determine the z coordinate of the event vertex. The VTX is capable of finding the event vertex with a precision of about 1 mm.

The SVX ([82], [83], [84], [85]) is the innermost tracking system, and consists of two barrels, one each on the east ($z > 0$) and west ($z < 0$) side of the detector. Each barrel is 25.5 cm long, and has four layers of single-sided silicon detectors located 3.0, 4.2, 6.8, and 7.9 cm from the beam line. The axial strips on the three innermost layers have 60 μm pitch (*i.e.* width); the strips on the outermost layer has 55 μm pitch. The SVX single-hit resolution is measured in data to be 13 μm. Track impact parameter relative to the beam position can be measured to a precision of 17 μm. The position of secondary vertices of

[2] Pseudorapidity is defined as $\eta = -\log\left(\tan\frac{\theta}{2}\right)$, where θ is the angle relative to the proton beam line. The detector pseudorapidity η_d is the pseudorapidity calculated taking the interaction vertex to be exactly at the center of the detector, $z = 0$.

the decay of long-lived particles is measured by matching SVX tracks with tracks found by the CTC. The SVX is precise enough to efficiently identify the secondary vertices from the decay of b-hadrons, which have typical decay lengths of about 400 μm. The momentum resolution of a combined SVX/CTC track is $\delta P_T/P_T = 0.0009 \cdot P_T \oplus 0.0066$ ($\oplus$ indicates addition in quadrature).

2.2.2.2 Calorimeters

The calorimeters surround the tracking detectors and the solenoid. They cover 2π in azimuth and $|\eta_d| < 4.2$ in pseudorapidity. The calorimeter system is divided into three separate regions in η_d: central, end-plug, and forward. Each region has electromagnetic calorimeters (CEM [86], [87], [88], [89], [90], PEM [92], FEM [95]) in front of hadronic calorimeters (CHA/WHA [91], PHA [93], FHA [94]).[3] The calorimeters are segmented in azimuth and pseudorapidity to form a projective tower geometry — *i.e.* they point back to the nominal interaction point, $z = 0$. The central calorimeters have towers that are 15^o wide in azimuth and 0.1 units wide in η_d, while the tower size in the plug and forward calorimeters is 5^o wide in ϕ and 0.1 units wide in η_d. The physical size of the inner face of these towers varies from 24.1 cm (η) $\times$ 46.2 cm (ϕ) in the central region to 1.8 cm $\times$ 1.8 cm in the forward region. In all cases, the absorber in the electromagnetic calorimeter is lead, and, in the hadronic calorimeter, it is iron. The active sampling medium in the central calorimeters is scintillators, while it is gas proportional chambers in the plug and forward calorimeters. See figure 2.12 for the position of the calorimeters. Their coverage, thickness, and resolution are summarized in table 2.1.

The CEM has embedded in it proportional chambers with strip and wire readout (CES) located at the approximate shower maximum depth ($6 \cdot X_0$). The CES provides precise shower position measurements in both the z and r-ϕ views. In addition, proportional chambers (CPR) are located between the solenoid and the CEM; they sample the early development of electromagnetic showers in the solenoid coil. The CPR measures only the

[3] WHA is the "wall" hadron calorimeter, which is located between CHA and PHA. It is constructed of the same material as the CHA, but has significantly different shape because it covers the edge of the cylindrically shaped central region. Throughout this report, the term "CHA" will be used to refer to both CHA and WHA.

System	η_d Range	Energy Resolution	Thickness
CEM	$\lvert\eta_d\rvert < 1.1$	$13.7\%/\sqrt{E_T} \oplus 2\%$	18 X_0
PEM	$1.1 < \lvert\eta_d\rvert < 2.4$	$22\%/\sqrt{E_T} \oplus 2\%$	18-21 X_0
FEM	$2.2 < \lvert\eta_d\rvert < 4.2$	$26\%/\sqrt{E_T} \oplus 2\%$	25 X_0
CHA	$\lvert\eta_d\rvert < 0.9$	$50\%/\sqrt{E_T} \oplus 3\%$	4.5 λ_0
WHA	$0.7 < \lvert\eta_d\rvert < 1.3$	$75\%/\sqrt{E_T} \oplus 4\%$	4.5 λ_0
PHA	$1.3 < \lvert\eta_d\rvert < 2.4$	$106\%/\sqrt{E_T} \oplus 6\%$	5.7 λ_0
FHA	$2.4 < \lvert\eta_d\rvert < 4.2$	$137\%/\sqrt{E_T} \oplus 3\%$	7.7 λ_0

Table 2.1: Summary CDF calorimetry properties. The symbol $\oplus$ signifies addition in quadrature. Energy resolutions for the electromagnetic calorimeters are for incident electrons and photons; for the hadronic calorimeters, they are for incident isolated pions. Energy is in GeV. Thicknesses are given in radiation lengths (X_0) for electromagnetic calorimeters, and in interaction lengths (λ_0) for hadronic calorimeters.

r-ϕ position of showers. The CES and CPR aid in the precise reconstruction of electron and photon momenta.

2.2.2.3 Muon Detectors

The CDF muon system consists of the CMU, CMP, and the CMX. The CMU consists of four layers of drift chambers located directly outside (radially) of each 15^o wedge of the CHA. The CMU covers the region $|\eta_d| < 0.6$. The four drift tube layers are used to obtain trigger-level muon P_T; a more accurate measurement of P_T is obtained by matching CTC tracks to tracks in the CMU, and taking the CTC momentum as the muon P_T. The CHA acts as a hadron absorber for the CMU. However, it is not thick enough to prevent some degree of hadronic "punch-through". In order to reject these muon fakes, 60 cm of steel is placed radially outside of the CMU, and, beyond this, there is the CMP system, which also consists four-layer drift chambers. By requiring a CTC track to extrapolate to both a CMU and CMP track, one can significantly reduce fake muons. The CMU and CMP have significant gaps in coverage due to design and geometric constraints. Approximately 84% of the solid angle in the region $|\eta_d| < 0.6$ is covered by the CMU, 63% is covered by the CMP, and 53% is covered by both.

The CMX extends muon coverage to the region $0.6 < |\eta_d| < 1.0$. It consists of drift tubes for muon detection sandwiched between scintillator counters used for triggering.

TRIGGER LEVEL	ACCEPT RATE	DEAD TIME
LEVEL 1	1 kHz	0%
LEVEL 2	20 Hz	4%
LEVEL 3	5 Hz	10%

Table 2.2: Typical trigger accept rates and dead times at each trigger level at instantaneous luminosity $10 \times 10^{30}\ cm^{-2}\ s^{-1}$ (a typical value for Run 1b).

These detectors are placed on four, free-standing conical arches. The CMX covers about 71% of the solid angle of the region $0.6 < |\eta_d| < 1.0$. A significant portion of the loss in coverage is due to the fact that certain wedges cannot be covered due to obstacles in the detector system. The CMX, like the CMU and CMP, uses the four drift tubes to obtain trigger-level muon P_T, and relies on extrapolated CTC tracks to obtain accurate muon momentum measurement.

2.2.2.4 Trigger and Data Acquisition

The CDF uses a 3-level trigger system. The Level 1 trigger relies on "fast output" from detector preamplifiers to quickly determine whether to keep a given event for further examination. This decision is made within $3.5\mu s$, which is the time between beam crossings. The Level 1 trigger, therefore, incurs zero dead time.[4] The Level 1 triggers relevant for the detection of the top quark are the central muon and calorimeter triggers. The three Level 1 central muon triggers use fast outputs from the CMU, CMP, and the CMX. The calorimeter triggers use fast outputs from all calorimeters. In these triggers, the energy in the electromagnetic and hadronic calorimeter are separately summed into towers of $(\Delta\eta_d = 0.2) \times (\Delta\phi = 15^o)$. At a typical luminosity of $10 \times 10^{30}\ cm^{-2}\ s^{-1}$, the Level 1 trigger rate is approximately 1 kHz (table 2.2).

The Level 2 trigger is activated shortly after the Level 1 trigger accepts an event. The Level 2 trigger examines in greater detail trigger signals coming from the calorimeter, tracking, and muon systems. For the calorimeter system, a hardware calorimeter cluster

[4]Dead time is defined as the fraction of beam crossings during which the detector is inactive. Dead time can be due to equipment failure or to the trigger. While an event is being analyzed by the trigger, the detector is inactive. Thus if the trigger takes more than the beam crossing time to analyze an event, dead time is incurred.

finder looks for clusters of calorimeter towers containing energy above certain threshold values. These clusters are normally due to jets, electrons, and photons. These clusters are listed in memory, and, for each cluster, the E_T, mean ϕ, and mean η_d are calculated by dedicated hardware. Trigger information on tracks are provided by the Central Fast Tracker (CFT) [96]. The CFT uses fast timing information from the CTC to obtain the charge, P_T, and ϕ of track candidates. The CFT momentum resolution is $\delta P_T/P_T \approx 0.035 \cdot P_T$, and has an efficiency of $93.5 \pm 0.3\%$ for tracks with P_T above 10 GeV. Trigger information on muon candidates are available for the CMU, CMP, and CMX. Fast signals from these detectors are sent to dedicated hardware, which determine the azimuthal position and P_T. There is also dedicated hardware that extrapolates CFT tracks to the muon candidates, and those with matching tracks are treated as "golden" (*i.e.* good or likely) muon candidates.

The information on calorimeter clusters, CFT tracks, and muon candidates are sent to the Level 2 "crate", in which resides several hardware decision modules, one track memory module, and two processor boards. The hardware decision modules take the information about the clusters and summarize them (*e.g.* counting the number of electromagnetic clusters passing a certain E_T threshold, finding the index to the most energetic cluster, *etc.*). The track memory module stores information about CFT tracks and muon candidates. The information in the decision and memory modules are accessed by the two processor boards, which make more sophisticated decisions using microprocessors and code stored in on-board memory. The connection between the processor boards and the various modules is supplied by a custom-made back-plane bus (the "processor bus"). The core of the processor boards is the DEC Alpha 21064 processor, which is one of the first commercial versions of the Alpha processor. It has a clock speed of 300 MHz. On the board is 4 Mb of memory in which to store software trigger code and trigger data accessed from the other modules.[5]

The time required by the Level 2 trigger to process an event depends greatly on the complexity of an event. Most events are very simple, having few calorimeter clusters and CFT tracks. Such simple events can be analyzed in about 20 to 30 μs. Events with many

[5]The processor boards were upgraded to the Alpha processor-based ones early in Run I. Before the upgrade, the processor boards were based on Motorola MC 10900 chips with a clock speed of 50 MHz. The on-board memory was used only to store trigger code; the hardware decision and track memory modules served as memory for data storage.

calorimeter clusters and CFT tracks can take hundreds of microseconds to analyze. Most of the time in these events is spent downloading tracking information (there can be > 100 tracks) from the track memory module to one of the processor boards, and the subsequent execution of trigger code that use tracking information in complicated ways. For a typical instantaneous luminosity of $10 \times 10^{30}\ cm^{-2}\ s^{-1}$, the Level 2 trigger incurs a dead time of about 4%, and outputs events to the Level 3 trigger at a rate of about 20 events/sec (table 2.2).

The Level 3 trigger is a fully software-based trigger. The trigger hardware consists of a "farm" of eight Silicon Graphics multi-cpu Challenges and PowerServers. This farm of cpu's has a combined processing power of about one billion instructions/sec. Unlike the lower level triggers, the Level 3 trigger bases its decision on the complete detector signal. The detector signal is collected and organized by the data acquisition system (DAQ). The event data are then reconstructed and analyzed using standard algorithms. Most of the reconstruction time is used for three-dimensional tracking in the CTC. At an instantaneous luminosity of $10 \times 10^{30}\ cm^{-2}\ s^{-1}$, the Level 3 trigger and the DAQ incurs a dead time of about 10%. At this luminosity, the trigger outputs events at a rate of about 5 per second (table 2.2). The accepted events are stored on data-quality 8 mm magnetic tapes (similar to those used in video "cam-corders").

Chapter 3

Physics Objects

3.1 Introduction

This thesis deals with $p\overline{p} \rightarrow t\overline{t}$ in the lepton + jets decay channel (see table 1.2). In this decay channel, there are three basic types of *physics objects*: (1) a *primary* [1] charged lepton (electron or muon); (2) a neutrino; and (3) quarks. Primary leptons are identified directly — *i.e.* electrons are identified reliably as electrons through a combination of information from the tracking and calorimeter system, while muons are identified by combining information from the tracking and muon systems. The presence of a neutrino is inferred from the imbalance in the sum of the transverse component of the deposited calorimeter energy. Finally, a quark is identified as a jet, which is defined as a cluster of calorimeter towers with significant energy deposition.

The objectives of this chapter are: (1) to indicate how each physics object is identified; and (2) to indicate the method used to reconstruct the momentum of the physics objects. A complete discussion of physics objects can be found in [16]. Here, the information is summarized.

[1] The qualification *primary* is used to designate the charged lepton originating from the semileptonic decay of one of the intermediate W's in $t\overline{t} \rightarrow W^+ W^- b \overline{b}$. In the context of lepton + jets events, there is a second class of charged leptons, referred to as the *soft* leptons. These originate from the semileptonic decay of b- and c-quarks, and are used in the soft lepton tagging (SLT) algorithm. Soft leptons will be discussed later in the description of the SLT algorithm.

3.2 Primary Lepton

3.2.1 Electron

3.2.1.1 Identification

The signature of a primary electron is basically a high-P_T CTC track pointing to a high-E_T CEM cluster (figures 3.1 and 3.2). More specifically, a primary electron candidate has the following features:

- It must be in the central rapidity region $|\eta_d| < 1.0$.
- It must have a CTC track extrapolating to a CEM cluster, which consists of a seed tower with $E_T > 3\ GeV$ and two neighboring towers in η_d. The size of the CEM cluster, therefore, is 3 towers in η_d ($\Delta\eta_d = 0.3$) by 1 tower in azimuth ($\Delta\phi = 15^o$).
- The shower position, as measured by the CES, is required to be within the *fiducial* region. This means that it should be sufficiently far from the calorimeter boundary. This cut is applied so that the energy can be measured reliably. The fiducial region covers 84% of the solid angle in the region $|\eta_d| < 1.0$.

These selection requirements are rather loose, in that they accommodate a significant fraction of fake primary electrons. There are two important types of fake primary electrons: (1) photon conversion[2]; and (2) charged hadrons. An electron from photon conversion is removed through the following requirements:

- The CTC track of the electron candidate must extrapolate to a *good* VTX track, where good means that the VTX occupancy is ≥ 0.2.
- There must *not* be a CTC track with charge opposite that of the primary electron candidate, and such that the invariant mass of the two tracks is small.

Charged hadron fakes are removed through the following seven requirements:

[2]The term *photon conversion* denotes a photon produced in the primary event that, upon traversal through the detector, hits some detector material and produces an e^+e^- pair.

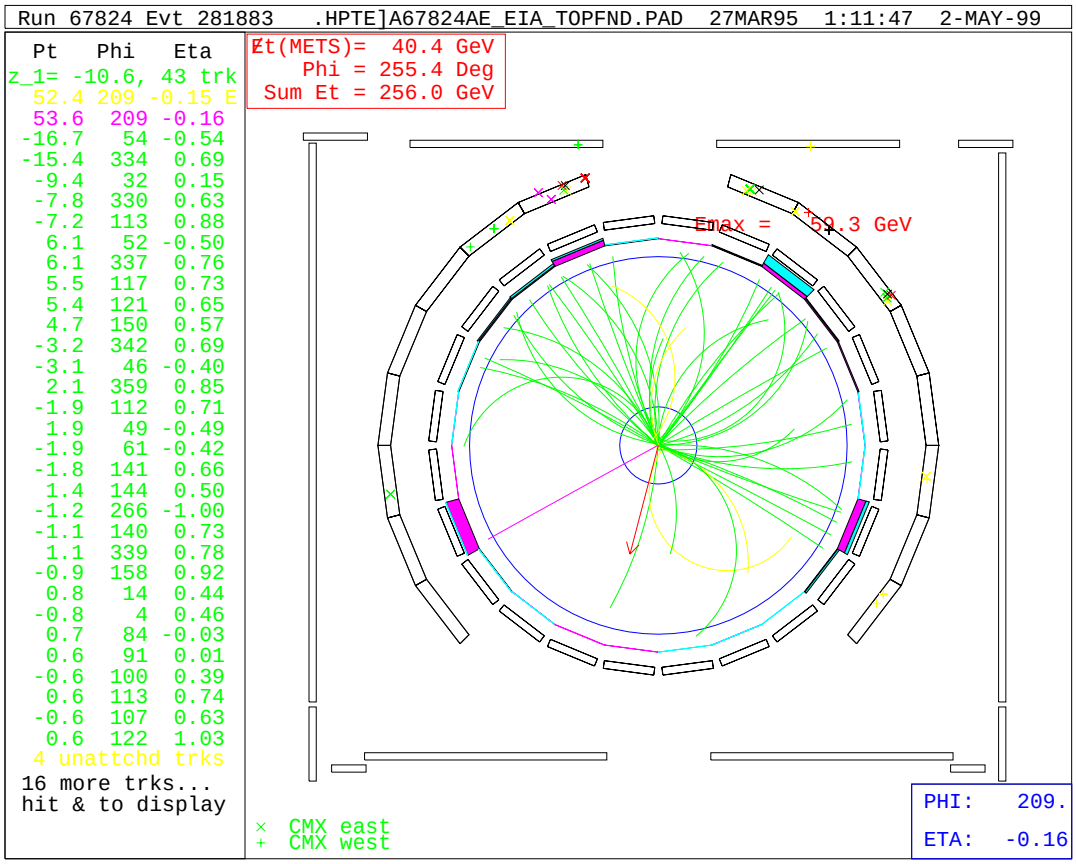


Figure 3.1: Event display of the CTC from a top candidate event (Run 67824, Event 281883). The primary electron is indicated by the stiff track pointing to the lower left of the figure and extrapolating to significant energy deposition in the electromagnetic calorimeter (the dark cell just outside of the tracking volume).

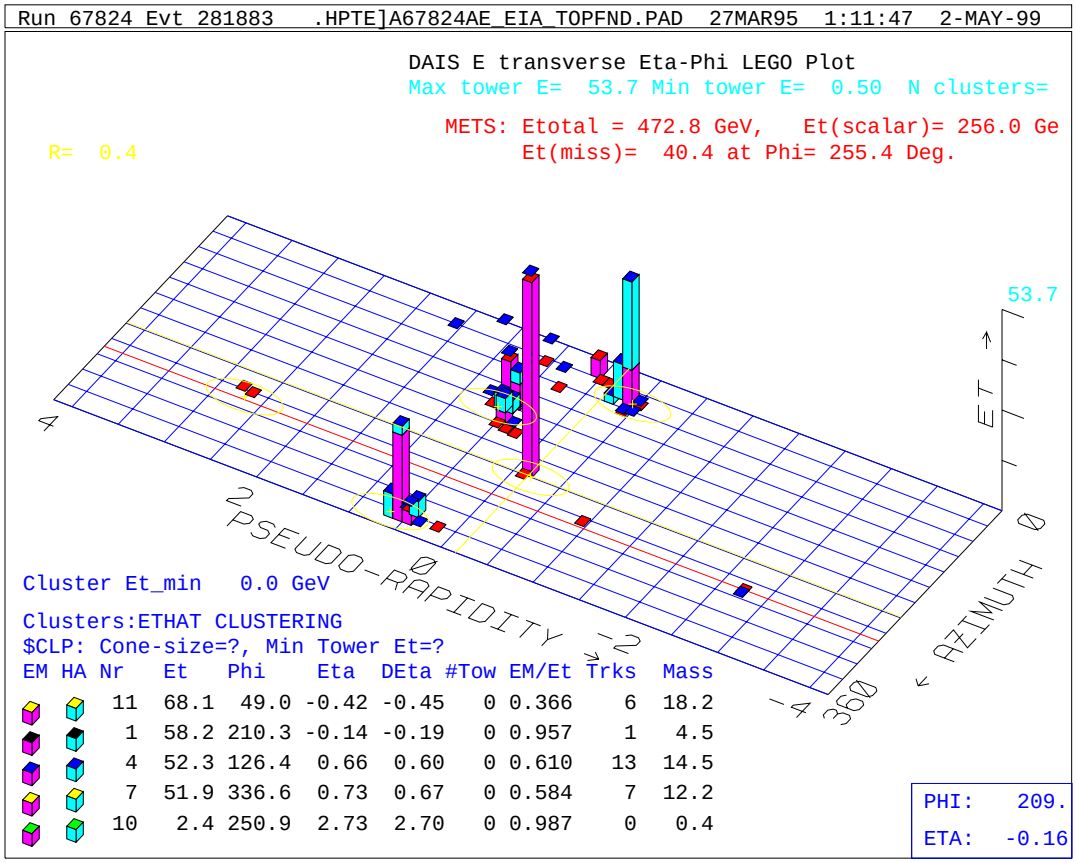


Figure 3.2: Event display of the calorimeter system from a top candidate event (Run 67824, Event 281883). The blocks in the figure indicate energy deposition in the calorimeter. The height of the blocks is proportional to the amount of energy deposited. The dark (light) blocks indicate energy deposition in the electromagnetic (hadronic) calorimeter. The primary electron is indicated by the tallest dark tower at $\eta_d = -0.14$ and $\phi = 210^o$. Also shown are jets, which are indicated by clusters of calorimeter towers contained in ellipses with size $\Delta R = 0.4$ (the ellipses are difficult to see, but visible, in the figure).

- The ratio of the hadronic energy to the electromagnetic energy in the cluster (HAD/EM) must be less than 5%.

- The ratio of the calorimeter energy to the track momentum, E/P, must be less than 1.5.

- The lateral shower profile in the calorimeter cluster, L_{shr}, must be consistent with that produced by a comparable test-beam electron. Quantitatively, $L_{shr} < 0.2$. See [37] for details about L_{shr}. Basically, L_{shr} is a measure of how much energy is deposited in towers neighboring the seed tower, given the amount of energy, the shower position, and angle of incidence of the electron in the seed tower.

- The CTC track of the electron candidate must match the CES shower position in both the r-ϕ view (Δx) and the z view (Δz), where Δx is the r-ϕ difference between the extrapolated track position and the shower position, while Δz is the corresponding difference in the z direction. The cuts are $\Delta x < 1.5$ cm and $\Delta z < 3.0$ cm.

- The CES shower profile must be consistent with that produced by a comparable test-beam electron. Quantitatively, $\chi^2_{strip} < 10$.

- The CTC track of the electron candidate must match the interaction vertex in the z direction to within 5.0 cm.

- The calorimeter cluster for the electron candidate must be *isolated*. Calorimeter isolation, I_{cal}, is defined as the total transverse energy deposited within a cone of radius $R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2} = 0.4$ centered about the electron cluster, but *excluding* the electron cluster's transverse energy, $E_T(ele)$. Quantitatively, $I_{cal}/E_T(ele) < 0.10$.

In order to benchmark the primary electron selection efficiency, these cuts were applied to a sample of $Z \rightarrow e^+e^-$ events. The efficiency is found to be 84%, excluding efficiency loss from photon conversion removal and from isolation cuts. See chapter 4 for a discussion on $t\bar{t}$ event selection efficiency and data sample purity.

3.2.1.2 Momentum Reconstruction

The momentum of an electron is obtained from the calorimeter energy and the event vertex (the CTC track of the electron is not used). The event vertex is used with position of the electron calorimeter cluster to determine the direction of the momentum, while the energy measurement gives its magnitude.

The event vertex is measured using a standard algorithm with the VTX. During the relatively high-luminosity conditions of Run 1b, there was an average of 2.5 interactions per beam crossing. Thus it was not uncommon for there to be several good interaction vertices in an event. In this analysis, the vertex that best matches the CTC track of the primary lepton was chosen.[3]

The energy of an electron is measured from the sum of the tower energies in the electron calorimeter cluster. The energy in a tower is, in turn, obtained from the geometric mean of the charge in the two phototubes, one on each side of the tower (in azimuth). This energy is then corrected for the following effects (see [97], [98], [99], [100], [37], [38] for details):

CEM Mapping Correction

The response of a CEM tower to an electron/photon of a given energy depends on where the electron/photon impacts the tower. This position-dependent response is corrected for using a "response map" obtained in the test-beam in 1984-1985 [86], [87], [88], [89], [90].

CEM Map Trimming

The response of CEM towers in the detector is somewhat different from that in the test-beam. These differences are accounted for by applying further mapping corrections. The corrections are obtained from the distribution of E/p, where E is the CEM tower energy, and p is the electron track momentum.

[3] This is in contrast to the top mass analysis [8], which uses the vertex with the greatest number of associated VTX hits. In the 163-event top mass sample, 13% of events use the wrong event vertex.[39] The authors of the top mass analysis have determined that the effect of using the wrong vertex has negligible effect on the top mass measurement, and have, therefore, decided to stick with the original (*i.e.* partly wrong) vertices.

CEM Tower-by-tower Correction

The average gain of each tower is slightly different from each other. This is corrected for by applying relative normalization.

Time-dependent Energy-response Correction

The energy response of the CEM decreases over time because of decreasing light attenuation length in the scintillator, decreasing photomultiplier tube (PMT) gain, and other effects whose origins are not well-understood. These effects are corrected for by normalizing with a factor that decreases linearly with run number. Different slopes are used for several different periods in Run I, each period being demarked by prolonged detector shutdown and/or detector access[4].

Absolute Energy Scale Correction

After all of the above corrections, the Z mass measured from $Z \rightarrow e^+e^-$ is found to be slightly different from the world-average value. A global correction factor is applied in order to account for this residual inconsistency.

The primary electron E_T resolution is obtained from studies of the $Z \rightarrow e^+e^-$ width. It is found to be $\delta E/E = 13.5\%/\sqrt{E_T} \oplus 1\%$.[38]

3.2.2 Muon

3.2.2.1 Identification

The signature of a primary muon is basically a high-P_T CTC track pointing to a track segment in CMU, CMP, or CMX (figure 3.3). Because of the muon detector geometry, a primary muon is necessarily in the central rapidity region $|\eta_d| < 1.0$.

Two of the important sources of fake muons are: (1) hadronic "punch-throughs" (*i.e.* particles that escape the outer radius of the hadronic calorimeter and enter the muon detector system); and (2) cosmic ray muons. The first source is removed by the following set of cuts:

[4]The behavior of the detector is observed to be somewhat different before and after prolonged shutdown and detector access. Thus the time-dependent corrections are also different, and must be determined separately before and after shutdown/access.

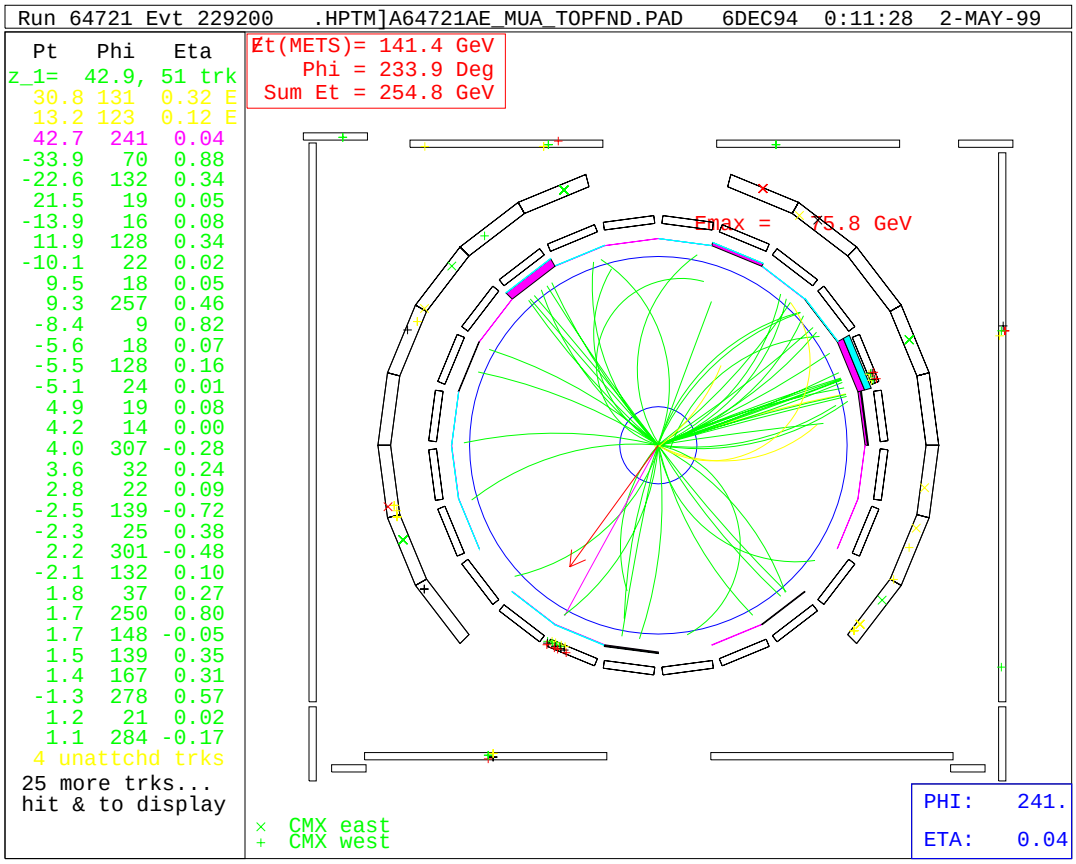


Figure 3.3: Event display of the CTC from a top candidate event (Run 64721, Event 229200). The large circular region indicates the CTC. Just outside of this is a view of the energy deposition in the calorimeter system (the dark (light) cells inidcate electromagnetic (hadronic) energy deposition). Outside of the calorimeters is the CMU (the '+'s indicate hits in the detector). Outside of this is the CMX detector (the '+'s and '×'s indicate hits). Finally, the ractangular region surrounding everything is the CMUP detector (with '+'s indicating hits). The primary muon is indicated by the stiff track pointing to the lower left of the figure and extrapolating to hits in the CMU and CMUP system.

- The energy deposition in the electromagnetic and hadronic calorimeter towers must be consistent with that for a minimum ionizing particle — *i.e.* a muon. More specifically, one extrapolates the CTC track of the muon candidate to an electromagnetic and hadronic calorimeter tower, and determine how much energy is contained in those towers. The cuts are $E < 2GeV$ and $E < 6GeV$ for the electromagnetic and hadronic towers, respectively.

- The r-ϕ distance $\Delta x = r \times \Delta\phi$ of the extrapolated CTC track to the track segment in the muon detector must be $\Delta x < 2$ cm for CMU and $\Delta x < 5$ cm for CMP and CMX.

- The muon candidate must satisfy the calorimeter isolation $I_{cal}/P_T(\mu) < 0.10$, where I_{cal} is the calorimeter energy within $\Delta R < 0.4$ of the muon track, subtracting the "typical" amount of energy deposited by a minimum ionizing particle.

The cuts used to remove cosmic ray muons are the following:

- The impact parameter (the closest approach in r-ϕ of the CTC track to the beam line) must not be too large. Specifically, it must be less than 3 mm.

- The CTC track of the muon candidate must be within 5 cm of a good interaction vertex.

As a benchmark for primary muon selection efficiency, these cuts were applied to $Z \rightarrow \mu^+\mu^-$ events. The efficiency was found to be 90.6%, excluding losses due to isolation cuts. See chapter 4 for a complete discussion of $t\bar{t}$ event selection efficiency and data sample purity.

3.2.2.2 Momentum Reconstruction

The momentum of a muon is determined from its CTC track. In a uniform axial (*i.e.* constant along the beam line) magnetic field, a charged particle track follows a helical path — it travels with constant speed along the beam line, and its motion projected in a plane transverse to the beam has constant curvature. This helix is parametrized by:

Curvature, C

This is the inverse-diameter of the circle segment described by the track in the r-ϕ plane.

Impact Parameter, D_0

Distance of closest approach to $r = 0$.

Azimuthal Angle, ϕ_0

The azimuthal angle at the point of closest approach to $r = 0$.

The Axial Position, z_0

The axial position at the point of closest approach to $r = 0$.

Track-beam Opening Angle, $\cot\theta$

θ is the opening angle between the CTC track and the beam line.

A CTC track is reconstructed by fitting hits in the CTC wires to a helical path, as parametrized by the above variables. In the fit, small variations ($\sim 1\%$) in the magnetic field are taken into account. The track is then re-fit by constraining it to have originated from the event vertex. The z position of the event vertex is determined by the VTX, while the r-ϕ position is measured by the SVX. This re-fitting improves the track momentum resolution by a factor of 2.[38] There are other smaller effects, such as ionization energy loss (dE/dx), bremsstrahlung, and false curvature (*i.e.* the misalignment of the CTC wires relative to each other, and of the CTC as a whole in relation to the beam and the SVX); these are negligible in high-P_T muons, and, therefore, are ignored.

The high-P_T CTC track resolution is determined from studies of the width of $Z \rightarrow \mu^+\mu^-$ to be $\delta P_T/P_T^2 = 8.10 \times 10^{-4}$.[38]

3.3 Quarks

3.3.1 Identification

Quarks are identified as jets through what is known as the "fixed-cone" jet clustering algorithm. Details of this algorithm can be found in [40]. Here, the algorithm is sketched.

In the first step of the jet clustering algorithm, a list of calorimeter towers (electromagnetic and hadronic) containing more than 1.0 GeV of transverse energy E_T is made. From this list, "pre-clusters" are made by grouping together contiguous towers with continually decreasing E_T, going out from the tower with maximum E_T. These pre-clusters are then used to initiate jet clustering. For each pre-cluster, an E_T-weighted centroid is calculated. Then all towers with $E_T > 0.1\ GeV$ within an η-ϕ cone of $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} = 0.4$ about the pre-cluster centroid are grouped into a cluster.[5] An E_T-weighted centroid of this set of towers is then calculated; this centroid is usually somewhat different from the pre-cluster centroid. Then towers in a cone about this new centroid are grouped together, and the centroid for this new cluster is recalculated. This process is repeated until the list of towers in a cluster remains unchanged.

The energy deposited in a jet cluster is primarily due to the quark or gluon that gave rise to it. The energy and momentum of a jet give a fairly reliable estimate of the energy and momentum of parent quark/gluon; the reconstruction of the energy and momentum of jets is discussed in the next section.

The difference between quark and gluon jets is quite miniscule, so it is difficult (if not impossible) to tell whether a given jet originates from a quark or a gluon. In $t\bar{t}$ candidate events, one simply assumes that the four leading jets (in E_T) are due to the four quarks in the lepton + jets decay channel. Exceptions to the indistinguishability of quark and gluon jets are found in jets due to the b-quark (and, to a lesser extent, the c-quark). Since the b-quark has a relatively long lifetime, a b-flavored hadron often travels a measurable distance before decaying. The decay of such hadron shows up as a displaced vertex of charged tracks, which can be reliably measured with the SVX. A b-flavored hadron can also be identified through the identification of the charged lepton (electron or muon) from the semileptonic decay of the b-quark. Such a lepton can be identified fairly reliably by the soft-lepton tagging (SLT) algorithm. The identification of b-quark jets is discussed in section 3.4.

[5] The choice of ΔR is somewhat arbitrary. Studies have shown that $\Delta R = 0.4$ to 1.0 are reasonable choices for reliably detecting jets and reconstructing their momenta.[41] For reconstructing $t\bar{t}$ events in the lepton + jets channel, studies have found that the choice $\Delta R = 0.4$ gives the best jet-counting efficiency.[16]

3.3.2 Momentum Reconstruction

The energy and momentum of a jet are obtained in three steps. First, one obtains the *raw* energy and momentum. Then, generic corrections are applied to the raw quantities. Finally, corrections specific to $t\bar{t}$ events are applied. This final class of corrections is discussed in chapter 6. In this section, the measurement of jet energy and momentum before and after the application of generic corrections is discussed.

The raw jet energy is simply the scalar sum of all the tower energies in the cluster. The jet momentum is the vector sum of the tower momenta, where tower momentum is defined such that its magnitude is the tower energy and its direction is defined by the line that connects the event vertex to the nominal shower-maximum position of the tower.

The generic jet energy corrections[6] correct for systematic shifts in the jet energy measurement due to: (1) imperfect behaviors of the calorimeters; and (2) limitations of the jet identification algorithm. Five categories of corrections are applied (see [42] and [44] for details about the generic corrections):

Relative Energy Scale

The relative energy scale correction accounts for the difference in the response to jets of various calorimeter subsystems. This correction is obtained by examining dijet data in which one of the jets is in the central region (where the jet energy resolution is good) and the other jet can be in any other region. The first jet is referred to as the "trigger jet", and the latter is called the "probe jet". In principal, the dijet system should have nearly zero net transverse momentum; in practice, the difference in relative response of different calorimeter systems often results in significant net transverse momentum. To correct for this difference, one assumes that all the jet energy mismeasurement is in the probe jet, and one obtains the scale factor for the probe jet energy needed to remove the transverse momentum of the dijet system. The relative energy scale correction, for a given η_d, is the average correction factor for a large number of jets in a bin around η_d (the correction is, to a good approximation, independent of jet E_T).

[6] The generic corrections are applied by the CDF FORTRAN jet library routine JTC96S.CDF.

Figure 3.4(a) shows the correction factor as a function of η_d. The largest corrections occur in the boundary regions between different calorimeter subsystems.

Absolute Energy Scale

The absolute energy correction accounts for nonlinearity in hadron calorimeter response to low-energy hadrons. Correction factors are determined by monte carlo studies in which the total energy of all monte carlo particles within a jet cone is compared to the the jet energy obtained from detector simulation. A correction factor as a function of raw jet E_T is obtained from such studies (see figure 3.4(b)).

Out-of-cone Correction

Some particles that ought to be counted as being part of a jet are lost because of the arbitrary size of the cone used to reconstruct jets. This loss is due to: (1) low-P_T charged particles curling outside of the cone; or (2) fragmentation effects causing occasional particles with large transverse momentum relative to the jet momentum. Monte carlo studies have been performed to estimate average energy loss as a function of raw jet E_T (figure 3.4(d)).

Underlying Event Subtraction

Underlying event energy is that part of the jet energy that is due to particles emanating from the debris of $p\overline{p}$ collision. The energy of these particles should be subtracted from the jet energy. The amount of energy to subtract is estimated from minimum-bias data (*i.e.* data in which the trigger requires only that an inelastic collision of $p\overline{p}$ has occurred). The basic idea is to assume that the amount of energy in a random cone in minimum bias data is similar to the underlying event energy. For Run 1a, 0.72 GeV was subtracted from the raw energy of each jet; in Run 1b, the corresponding figure is 0.65 GeV.

Secondary Interaction Subtraction

Typical instantaneous luminosities in Run 1b were significantly larger than that in Run 1a (the average instantaneous luminosity at the beginning of a store in Run 1b was $16 \times 10^{30} cm^{-2} s^{-1}$, compared to $5.4 \times 10^{20} cm^{-2} s^{-1}$ in Run 1a[105]). This resulted

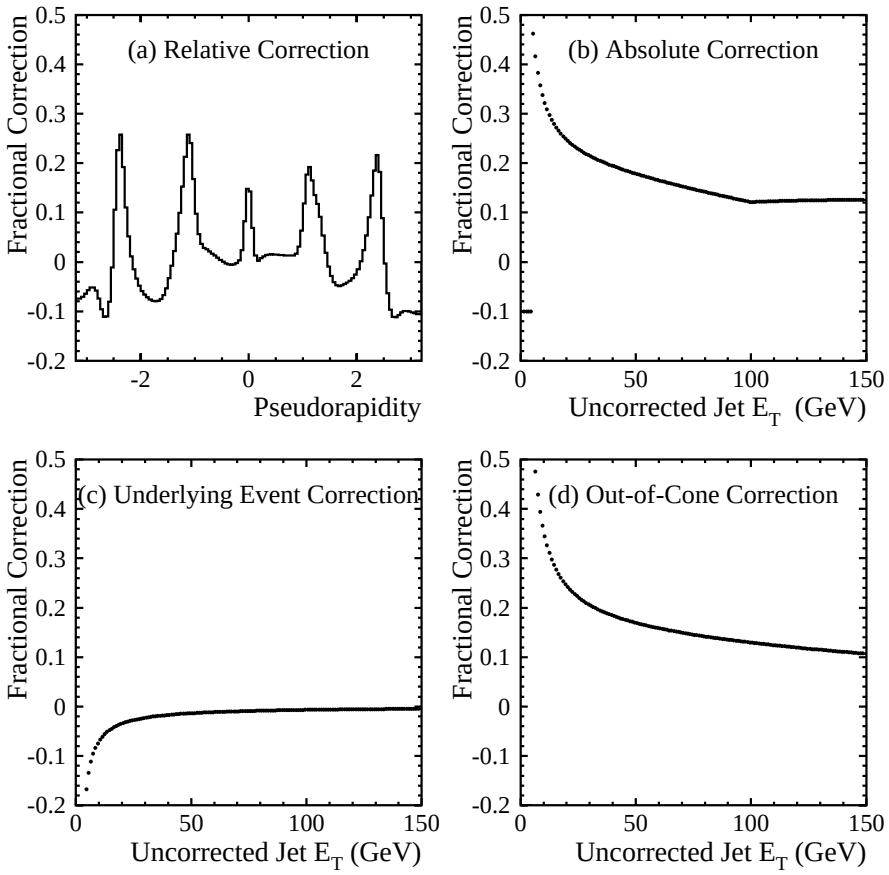


Figure 3.4: Generic jet corrections plotted as fractional change in raw jet E_T vs. jet η_d or raw jet E_T. This figure is from [44].

in larger number of interactions per beam crossing in Run 1b (2.5 in Run 1b vs. 0.85 in Run 1a). Thus, for Run 1b data, it is important to account for underlying events originating from extra inelastic $p\overline{p}$ collisions in the event. The raw jet energy of each jet was reduced by 0.297 GeV for each extra interaction in the event.

The nominal jet momentum resolution after the application of the generic corrections is given by the following:[7]

[7]See the CDF top physics library routine JET_ERROR.CDF and [101] for details about these errors.

$$\delta E_T = \kappa \cdot (0.1 * E_T + 1.0\ GeV) \tag{3.1}$$

The factor κ is about 1.0 to 1.1 for $|\eta_d| < 2.4$, and 1.6 otherwise. This nominal form, however, does not directly apply to jets in $t\bar{t}$ events. See chapter 6 for a discussion of this issue.

3.4 b-Quark

A b-quark jet is often distinguishable from a jet from a lighter quark for the following reasons:

Long b-quark Lifetime

The b-quark lifetime is rather long in spite of the fact that it is the second heaviest known quark. This long lifetime is due mainly to the fact that the b-quark mixes rather weakly with lower-generation quarks ($|V_{cb}| \approx 0.043$, $|V_{cd}| \approx 0$). Because of their relatively long lifetimes, b-flavored hadrons often travel $10^2 \sim 10^3\ \mu m$ before decaying. The decay vertex of such particles, when measured with the SVX, is displaced from the primary event vertex. Such displaced vertices are not very likely in jets from lighter quarks. Thus one can use a displaced vertex associated with a jet as a means of identifying b-quark jets. The identification of b-quark jets by this means is referred to as *secondary vertex tagging* (SVX tagging).

b-quark Semileptonic Decay

In $t\bar{t}$ events, the charged lepton $\ell = \mu, e$ from the decay $b \to \ell + \nu_\ell + X$ has two important properties: (1) it has relatively large P_T (see figure 3.5); and (2) it is inside a relatively energetic jet. These conditions are not very likely to be met in jets from lighter quarks. Thus one can use these conditions to identify b-quark jets. The identification of b-quark jets by this means is referred to as *soft lepton tagging*, or *SLT*.

In this section, the identification of b-quark jets by the SVX and SLT methods is discussed.

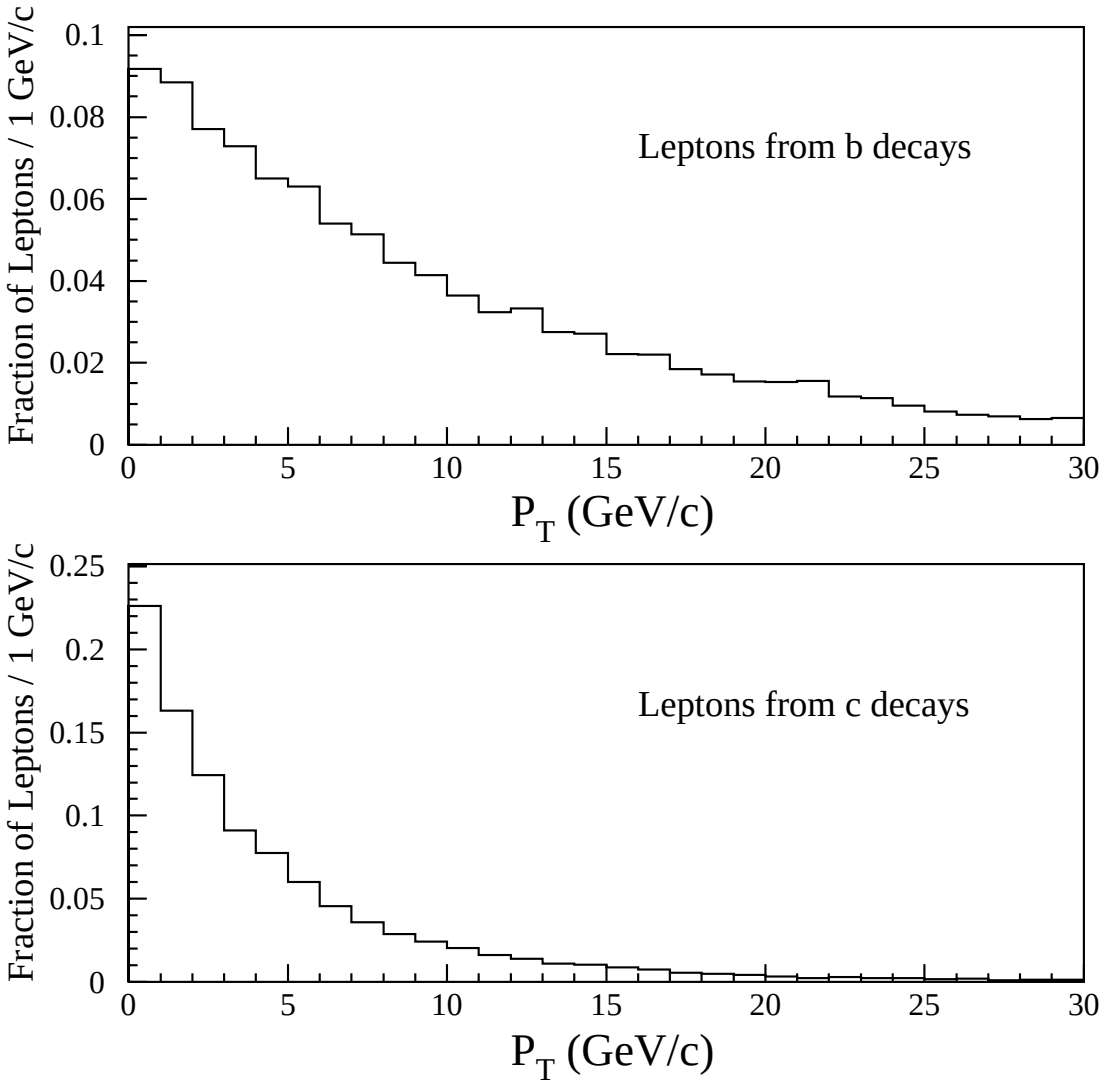


Figure 3.5: The lepton P_T spectrum from the semileptonic decay of b- and c-quarks that originate from the decay of $t\bar{t}$ in the lepton + jets channel. The distribution is obtained from the HERWIG monte carlo with $m_{top} = 160\ GeV$. The distributions are not too different from those at $m_{top} = 175\ GeV$. The figure is from [16].

3.4.1 SVX b-tagging

A displaced vertex associated with high-E_T jets can be found by various algorithms. Three types have been developed in CDF: (1) the "d-ϕ" algorithm; (2) the "jet-probability" algorithm; and (3) the "secondary vertex" algorithm. Reference [16] describes the basic principles involved in these algorithms. In the study of $t\bar{t}$ physics at CDF, the third type is most often used.

The basic operating principle of the third type of algorithm is as follows. One makes a list of jets with raw $E_T > 15\ GeV$. One also makes a list of "good" SVX-CTC tracks, where an SVX-CTC track is one where an extrapolated SVX track segment matches (in the r-ϕ plane) a CTC track well. The qualifier "good" indicates, among other things, that the track must be sufficiently energetic ($P_T > 2\ GeV$), and its SVX segment must be formed of a sufficient number of good SVX clusters (*e.g.* the clusters are not shared with other tracks, the charge profile is signal-like, *etc.*). Then, given these lists, one associates tracks to jets using the rule that if a track lies within 35^o of a jet's momentum, that track is presumed to be part of the jet. Then, from among tracks associated with each jet, one looks for tracks that form a displaced vertex. A jet which has tracks that form a displaced vertex is labeled as being very likely to be a b-quark jet.

The details of this algorithm differed somewhat between Run 1a and Run 1b. In Run 1a, displaced vertices were formed from two or more SVX-CTC tracks satisfying track-quality cuts. In Run 1b, displaced vertices were sought in two passes. In the first pass, one looks for displaced vertices formed from three or more SVX-CTC tracks satisfying loose track-quality cuts. Then, if no displaced vertex is found, one looks for displaced vertices formed from two tracks satisfying tighter track-quality cuts.

The Run 1b algorithm is much improved compared to that of Run 1a. The 1b algorithm tags b-quarks from $t\bar{t}$ events with an efficiency of about 40%; the corresponding figure for 1a is about 20%. Also, in Run 1b, the fractional contribution of mistags in b-tagged $W+ \geq 3$ jet events is expected to be $2.3/34 \approx 7\%$ [7], compared to $0.76/6 \approx 13\%$ [16] in Run 1a.

3.4.2 SLT b-tagging

The SLT algorithm looks for a "soft" lepton (electron or muon with $P_T > 2\ GeV$) associated with a jet with $E_T > 8\ GeV$. The lepton is required to be within a cone of $\Delta R < 0.4$ of the jet's momentum. Soft leptons are required to pass a large array of cuts intended to maximize b-tagging efficiency while minimizing background contamination; the details of the soft lepton selection criteria can be found in [16].

The SLT b-tagging efficiency in $t\bar{t}$ events is about 17%.[44] This is to be compared to 40% for the SVX tagger. Also, SLT is inferior to SVX in that it has greater background. In spite of these weaknesses, the SLT algorithm uses information that is only weakly correlated with that in the SVX algorithm, so it adds significant acceptance for tagging b-quark jets.

3.5 Neutrino

The presence of a high-P_T neutrino in an event is inferred from a large imbalance in the total transverse energy deposited in the calorimeters. This imbalance is quantified by the missing transverse energy, $\not\!\!E_T$, which is defined as the *negative* of the vector sum of transverse component of the "tower momentum" (see 3.3.2 for definition) of all calorimeter towers with $|\eta_d| < 3.6$. The range of η_d is restricted because the focusing magnets in the Tevatron obscure parts of the forward hadron calorimeter. Also, there are minimum threshold energies that towers in the various calorimeter systems must satisfy in order to be included in this sum. See [16] for details. The nominal $\not\!\!E_T$ resolution is given by $\delta\not\!\!E_T = 0.7 \cdot \sqrt{\sum E_T}$, where $\sum E_T$ is the *scalar* sum of the transverse tower energies measured in GeV.

In events with a high-P_T muon, $\not\!\!E_T$ must be corrected for the muon momentum. This is because the muon is a minimum-ionizing particle, so that it only deposits a fraction of its original energy as it traverses the calorimeters. To correct for the muon, its P_T as measured by the CTC is added to the original vector sum of calorimeter tower momenta. From this resultant vector, the expected amount of calorimeter energy deposited by the muon is subtracted. The negative of the final vector is the corrected $\not\!\!E_T$. Similarly, in events with a high-P_T electron, $\not\!\!E_T$ must be adjusted for the electron energy corrections described in

section 3.2.1. To do this, the calorimeter towers associated with the *uncorrected* electron is first clustered together. The contribution by these towers to $\not\!\!E_T$ is then subtracted from $\not\!\!E_T$. Finally, the corrected electron transverse momentum is added back in to obtain the corrected $\not\!\!E_T$.

The $\not\!\!E_T$ described in this section is used in this analysis as the first estimate of the neutrino's transverse momentum. The neutrino momentum obtained from the kinematic reconstruction of $t\bar{t}$ events is somewhat different from this. Chapter 6 discusses this issue in detail.

Chapter 4

Event Selection

4.1 Introduction

This chapter presents a discussion of the criteria used to obtain a sample of $t\bar{t}$ candidate events from the experimental data. Section 4.2 discusses why this analysis deals only with $t\bar{t}$ events in the lepton + jets decay channel. Section 4.3 discusses the cuts on physics objects used to obtain a sample of $t\bar{t}$ candidate events. Section 4.4 discusses the fact that the $t\bar{t}$ candidate events are subdivided into eight mutually exclusive sets, and explains why this is done. Finally, section 4.5 discusses the backgrounds in each of the eight subsamples of the data.

4.2 Choice of Decay Channel

According to the standard model, the top quark decay is a two-step, sequential, 3-body process:

1. $t \rightarrow b + W$

2. $W \rightarrow \ell + \nu_\ell$ or $q_d + q_u$

Here, ℓ is either an electron, muon, or a tau-lepton, ν_ℓ is the neutrino associated with ℓ, and q_d and q_u are the up- and down-type quarks, respectively, of the first two generations of quarks. Experimentally, it is useful to classify the decay of the $t\bar{t}$ system according to

how the intermediate Ws from t and $\bar{t}$ decay. This classification is summarized in table 1.2 of chapter 1. They are the: (1) dilepton; (2) lepton + jets; (3) all-hadronic; and (4) tau.

In this thesis, only the lepton + jets decay channel is examined. The decision to not include the other decay channels from this analysis is based on the following considerations:

- According to the standard model, the charged lepton from the decay of the top quark is the most sensitive probe of: (1) the top quark's spin polarization; and (2) the property of the t-b-W decay vertex. This issue is discussed in chapter 7.

- In order to *independently* probe the top quark spin polarization and the property of the t-b-W decay vertex, one must examine the top quark decay in the top quark rest frame. Experimentally, this means that one must reconstruct the top quark momentum in the lab frame. This issue is discussed in chapters 6 and 7.

These considerations impose the following constraints on the decay channels of $t\bar{t}$: (1) it should contain at least one primary charged lepton that is easily identifiable; and (2) it should be such that the momentum of t and $\bar{t}$ are reliably reconstructable. The all-hadronic decay channel fails both tests. The tau-decay channel fails the first test because the tau-lepton cannot be identified very reliably. The dilepton channel fails the second test: because two high-P_T neutrinos carry away undetected a significant fraction of the total energy in the event, the equations for the lab-frame t and $\bar{t}$ momenta are under-constrained, and, therefore, one cannot reliably reconstruct their momenta. This leaves only the lepton + jets channel, which easily passes the first test, and adequately passes the second.

4.3 Event Selection Cuts

The "nominal"[1] signature of the lepton + jets decay channel of $t\bar{t}$ in this analysis (and in the CDF lepton + jets top mass analysis [44]) is: (1) a high-P_T, isolated electron or muon; (2) four or more jets; and (3) large $\not\!\!E_T$. The electron/muon is presumed to originate from the semileptonic decay of t or $\bar{t}$, $\not\!\!E_T$ is presumed to be due primarily to the neutrino

[1] The qualification "nominal" is discussed later in this section. It indicates the fact that, although the event selection cut is designed to select only events from $t\bar{t}$ in the lepton + jets channel, the actual data sample passing the cuts is contaminated by other decay channels of $t\bar{t}$ faking lepton + jets.

from the semileptonic decay of t or $\bar{t}$, and the four leading-E_T jets are presumed to originate from the two b-quarks and two light quarks in the decay of t and $\bar{t}$.[2]

The event selection cuts designed to select $t\bar{t}$ in the lepton + jets decay channel are the following:

1. A candidate event must pass the high-P_T electron or muon trigger (lepton $P_T > 20\ GeV$ and $|\eta_d| < 1.0$).

2. A candidate event must have raw $\not\!\!E_T > 20\ GeV$ (see section 3.5).

3. The trigger (*i.e.* primary) lepton in a candidate event must satisfy the primary lepton identification requirements given in section 3.2.

4. A veto is placed on candidate events passing the event selection cuts for the dilepton channel of $t\bar{t}$ (see [6] for the dilepton cuts).

5. A veto is placed on candidate events having a Z-boson-like object. Such an object is defined as an e^+-e^- or μ^+-μ^- pair with invariant mass in the window 75-105 GeV. Also, candidate events containing an object consistent with a radiative Z-boson decay is removed. This object is defined as an e^+-e^- or μ^+-μ^- with an associated high-E_T photon, the combined invariant mass of which lies in the window 75-105 GeV.

6. A candidate event's primary vertex must be within 60 cm of the nominal interaction point ($z = 0$). In events with several primary vertex candidates, the vertex chosen is that which has the greatest number of VTX hits.[3]

7. A candidate event must have ≥ 3 jets with $E_T > 15\ GeV$ and $|\eta_d| < 2.0$.

8. A candidate event must have, in addition to the three jets described above, ≥ 1 jet with $E_T > 8\ GeV$ and $|\eta_d| < 2.4$.

[2]The presumption that the four leading E_T jets originate from the two b-quarks and two light quarks from the decay of t and $\bar{t}$ does not always correspond to reality: in a fraction of events, one or more jets from hard gluon radiation are among the four leading E_T jets.

[3]Incidentally, this choice of primary vertex is incorrect (*i.e.* disagrees with the primary lepton vertex) in about 13% of candidate events. This issue is discussed further later this section.

The first seven criteria are exactly the same as those used in the CDF lepton + jets $t\bar{t}$ cross section analysis.[7] The complete set (1-8) is almost identical to that used in the CDF top mass analysis. There are, however, a couple of differences:

- In this analysis, no cut is placed on the event reconstruction χ^2, which is discussed in chapter 6. This is because the cut does not improve the resolutions of the quantities measured in this analysis. See appendix A for details.

- In this analysis, the event vertex used to reconstruct the jet momenta is that of the primary lepton. The CDF top mass analysis, on the other hand, uses the vertex with the greatest number of VTX hits. The former choice is the correct event vertex. The latter choice disagrees with the former in 13% of the candidate events. A study by one of the authors of the CDF top mass analysis has shown that the effect of the wrong choice of vertex on the top mass measurement is small.[39] In this study, however, the correct vertex is used since the observables (the energy and angle of the charged lepton in the rest frame of the semileptonically decaying top quark) are probably more sensitive to the effect of choosing the wrong vertex than is the case with top quark mass.

The number of $t\bar{t}$ candidate events satisfying the cuts 1-8 is 159 (it is 163 in the CDF top mass analysis). The efficiency of this set of cuts is about 10%. The signal content of the 159 events is estimated to be 35%. See 4.5 for more on signal and background fractions.

One final point is in order before leaving this section. The set of cuts 1-8 described above is designed to select $t\bar{t}$ decaying in the lepton + jets channel. However, other decay channels of $t\bar{t}$ can fake the lepton + jets signal, so the signal portion of the final event sample is not purely lepton + jets. The break-down of the various decay channels in the lepton + jets candidate sample is shown in table 4.1. This contamination from the wrong $t\bar{t}$ decay channels is not treated as a background. Instead, during the $t\bar{t}$ event reconstruction process (chapter 6), such fake lepton + jets are treated as true lepton + jets. This causes the energy and angular distributions examined in this analysis to be smeared somewhat compared to those one would obtain from a pure sample lepton + jets; in effect, fake lepton + jets are treated as badly-measured true lepton + jets.

CHANNEL	FRACTION
lepton + jets	85%
tau	11%
dilepton	4%
all-hadronic	0%

Table 4.1: Estimated composition of the "lepton + jets" sample in the signal portion of the lepton + jets $t\bar{t}$ candidate events. The estimate is obtained with the HERWIG monte carlo generating $t\bar{t}$ with $m_{top} = 175\ GeV$. The generated $t\bar{t}$ events were allowed to decay to all possible decay channels.

4.4 Subdivision of the Data Sample

The event selection cuts shown in the last section result in an experimental data sample of 159 lepton + jets $t\bar{t}$ candidate events. Of these events, an estimated 35% are believed to originate from the production and decay of $t\bar{t}$, while the rest are believed to be fake $t\bar{t}$ events from various sources of background. The nature of the background sources and the estimated contributions they make to the $t\bar{t}$ sample are discussed in the next section. Before discussing background issues, however, a discussion on the fact that the data sample is divided into eight mutually exclusive subsamples is in order.

The 159 $t\bar{t}$ candidate events are subdivided according to two sets of criteria:

- The sample is first divided according to the number of tight jets — *i.e.* jets that pass the cuts $E_T > 15\ GeV$ and $|\eta_d| < 2.0$. Events with ≥ 4 tight jets belongs to the set labeled NJ4; events with exactly three such jets (with ≥ 1 additional "loose jet") belongs to the complementary set labeled NJ35. The label NJ4 is short-hand for "number of jets = 4"; similarly, NJ35 is short for "number of jets = 3.5" — *i.e.* not quite 4-jets.

- These two mutually exclusive samples are then further subdivided according to the type of b-tags. The four mutually exclusive b-tag categories are: (1) svx-only (XO); (2) slt-only (TO); (3) svx and slt (XT); and (4) no-tag (NT).

In total, there are eight, mutually exclusive subsamples labeled (NJ35, XO), (NJ4, XO), (NJ35, TO), *etc.* Table 4.2 shows the number of events in each subsample.

	NJ35	NJ4
XO	4	11
TO	5	9
XT	4	4
NT	77	45

Table 4.2: The number of events in each subsample.

The $t\bar{t}$ candidate sample is subdivided in this manner because this improves the measurement resolution. A complete discussion of how this improvement comes about is given in section 8.2.4. Here, the reason will be sketched. Imagine that one has a sample of $t\bar{t}$ events with N events. Imagine subdividing the samples into two, mutually exclusive subsets with N_1 and N_2 events. Also, let us assume that the background fraction of the two subsamples are β_1 and β_2. It is shown in section 8.2.4 that if $\beta_1 \neq \beta_2$, then the combined statistical error of the subdivided sample is smaller than that in which the whole data sample is treated as a single unit with background fraction being the statistics-weighted average of β_1 and β_2. In fact, the greater the difference between β_1 and β_2, the greater the degree of improvement in measurement resolution. This logic applies inductively to subdivision of the original sample into any number of mutually exclusive sets.

Given the above observations, the choice of categories is easily understood. Since background events tend to have softer and fewer jets than $t\bar{t}$ events at $m_{top} = 175\ GeV$, the subdivision of the data into the sets NJ4 and NJ35 results in subsamples of data with a rather large differential in background fractions. Similarly, the background fraction changes significantly from svx-slt events (smallest) to no-tag events (largest). Thus the background fractions in the eight subsamples of data vary considerably; this large range of variation is the key to improved measurement resolution.

4.5 Backgrounds to the $t\bar{t}$ Candidate Events

Earlier, it was stated that 159 $t\bar{t}$ candidate events pass the event selection criteria described in section 4.3. Of those events, it is estimated that only 35% originate from

	NJ35	NJ4
XO	$0.28^{+0.14}_{-0.09}$	$0.07^{+0.03}_{-0.02}$
TO	$0.65^{+0.13}_{-0.13}$	$0.25^{+0.11}_{-0.09}$
XT	$0.14^{+0.09}_{-0.05}$	$0.03^{+0.02}_{-0.01}$
NT	$0.90^{+0.05}_{-0.06}$	$0.57^{+0.12}_{-0.16}$

Table 4.3: Estimated background fraction in the subsamples. These numbers are the ones used in the top mass analysis[46]; they are slightly different from the ones used in this analysis. The errors are statistical; the systematic errors are small in comparison, so they are ignored.

the production and decay of $t\bar{t}$; the remaining 65% of the events originate from background processes that fake the lepton + jets $t\bar{t}$ signature. There are about a dozen physics processes that contribute to the background. Details can be found in [44], [7], and [16]. Essentially, they can be classified into three categories:

- W + jets (67%)
- QCD multijets (20%)
- Z + jets, $Z \rightarrow \tau\tau$, diboson, and single-top (13%)

The percentage in parenthesis is the estimated fractional contribution to the total background. QCD multijets include $b\bar{b}$ production in association with hard gluon jets, while the term diboson refers to the production and decay of WW, WZ, and ZZ in association with hard gluon jets.

An estimate of the background fraction in each of the eight subsets of data is necessary in order to compare the distribution of observables in the experimental data with those in monte carlo predictions of $t\bar{t}$ and background processes. This estimate has been made in [46] and [44], and the reader is referred there for details on the method used to obtain the results. Table 4.3 gives the background fractions in each subset.

The background fractions used in this analysis are slightly different from the ones given in table 4.3 because no χ^2 cut is applied in this analysis, while $\chi^2 < 10$ is applied in the CDF top mass analysis. To correct for this, the numbers in table 4.3 are extrapolated to $\chi^2 < \infty$ using the following formula:

	NJ35	NJ4
XO	$0.29^{+0.14}_{-0.09}$	$0.07^{+0.03}_{-0.02}$
TO	$0.66^{+0.13}_{-0.13}$	$0.25^{+0.11}_{-0.09}$
XT	$0.18^{+0.09}_{-0.05}$	$0.04^{+0.02}_{-0.01}$
NT	$0.90^{+0.05}_{-0.06}$	$0.58^{+0.12}_{-0.16}$

Table 4.4: Estimated background fraction in the subsamples for this analysis. These numbers are extrapolations from the numbers in table 4.3. The errors are assumed to be the same as in 4.3.

$$\beta(\infty) = \frac{\epsilon_s(10) \cdot \beta(10)}{\epsilon_b(10) + (\epsilon_s(10) - \epsilon_b(10)) \cdot \beta(10)} \tag{4.1}$$

The quantity $\beta(\infty)$ is the background fraction without a χ^2 cut (*i.e.* $\chi^2 < \infty$), while $\beta(10)$ is the background fraction when the cut $\chi^2 < 10$ is applied. The quantities $\epsilon_s(10)$ and $\epsilon_b(10)$ are the fraction of signal and background events surviving the cut $\chi^2 < 10$ (see figure A.1 in appendix A for plots of $\epsilon_s(\chi^2)$ and $\epsilon_b(\chi^2)$). Table 4.4 shows $\beta(\infty)$ for the eight subsets. The quantity $\beta(\infty)$ is almost unchanged from $\beta(10)$ in all subsets because $\epsilon_s(\chi^2)$ and $\epsilon_b(\chi^2)$ are not too different from each other (see appendix A). One can see from equation 4.1 that $\beta(\infty) \equiv \beta(\chi^2)$ when $\epsilon_s(\chi^2) \equiv \epsilon_b(\chi^2)$.

Chapter 5

Simulation Tools

5.1 Introduction

Simulation tools are essential for meaningfully comparing the top quark decay kinematics in the experimental data with those predicted by the standard model. In this analysis, two types of simulations tools are used: (1) monte carlo event generators for simulating the production and decay of the signal and background processes; and (2) a detector simulator that takes the output particle information from the monte carlos and simulates the CDF detector response. These simulation tools are described in this chapter. Also, several of the more important limitations concerning the use of these tools are noted.

5.2 Monte Carlo Generator for $t\bar{t}$

The production and decay of $t\bar{t}$ is simulated with the HERWIG V5.6 monte carlo[47], [48]. HERWIG uses leading-order QCD matrix elements to calculate the phase-space weighting for the hard process. The hadronic decay products from the hard process are hadronized using color-coherent parton shower evolution and cluster hadronization. The underlying event is generated using phenomenological models from experimental data. The decay of b-mesons is performed by the QQ monte carlo from the CLEO experiment [50], [51]. The QQ monte carlo is used instead of HERWIG because QQ's treatment of b-meson decay agrees with data better than that of HERWIG.

5.3 Monte Carlo Generator for Background Processes

The VECBOS [52] monte carlo is used to simulate all of the background processes (the issue of the whether or not it is appropriate to use VECBOS alone for modeling all of the dozen or so background processes is discussed in the next section). VECBOS is a parton-level monte carlo that uses tree-level matrix elements for W plus a fixed number of quarks and/or hard gluons. In this analysis, the matrix element for W + 3 quarks/gluons is used. The output partons from VECBOS are fed into the program HERPRT [53], which hadronizes the quarks and gluons (HERPRT uses the hadronization machinery in HERWIG). Because HERPRT causes some fraction of the generated W + 3 quarks/gluons events to radiate extra hard gluons, some of these events end up having four or more jets, which is one of the prerequisites for $t\bar{t}$ event selection.

5.4 Parton Distribution Function

The parton distribution function MRSD0′ is used in the generation of both the signal and background events.

5.5 Detector Simulation

In order to compare monte carlo events with events from experimental data, the output from the monte carlos (a list of the 4-momentum of several hundred particles) must be converted into simulated detector signals. This conversion is performed by QFL′, which is one of CDF's detector simulation code. The output of QFL′ is a collection of "data banks", which are basically formatted arrays containing simulated response of the tracking and calorimeter systems to the input particles. This is then analyzed in very much the same way the experimental data are in order to identify physics objects, such as electrons, muons, jets, and $\not\!\!E_T$. See [71] to [77] for more details on QFL′.

5.6 Caveats on the Monte Carlo Tools

There are several caveats concerning the monte carlo tools that should be pointed out. They are the following:

- The VECBOS monte carlo is used in this analysis to model all of the background processes, including those that do not originate from W + quarks/gluons. This choice is justified if the distribution of the observable of interest in true W + jets background is similar to those in non-W backgrounds. This choice seems to be adequate in the CDF top mass analysis [54]. In this analysis, however, this may not be adequate, as indicated by [55], in which the helicity fraction of the intermediate Ws in $t\bar{t}$ decay is measured using the primary lepton P_T spectrum. It is shown there that the lepton P_T spectrum in $b\bar{b}$ + jets events (which accounts for about 20% of the background) is considerably softer than that predicted by VECBOS.

- Multiple interactions are not dealt with in the simulation of both signal and background events. CDF studies on W + n-jets cross section [56], [57] show that these extra interactions can have rather large effects on: (1) the observed number of W + n-jets events; (2) the electron identification efficiency; and (3) the $\not\!\!E_T$ spectrum. Since [56] and [57] deal with W in the electron decay channel, it says nothing about muon events; similar problems, however, are expected in the muon decay channel too. It is not clear to what extent multiple interactions affect the shape of the observable distributions in both signal and background; nor is it clear how this would change the background fraction and the relative composition of the background (*i.e.* true-W *vs.* non-W).

- No trigger simulation is applied to the simulated signal and background events. This could affect the shape of the observable distributions in signal and background (it does not affect the background fraction and the relative background composition because trigger efficiencies are included in determining these [44], [46]). It is believed, however, that there will not be a dramatic change in the shape.

- The monte carlo statistics of the VECBOS sample is rather small. The generation of VECBOS events is a very laborious process; increasing monte carlo statistics is not a trivial matter. This analysis uses about 3000 VECBOS events. These events are divided into eight subsamples, as described in section 4.4. These subsets are populated unevenly; subsamples with large background fraction have lots of ($\sim 10^3$) VECBOS events, while those with small fraction have few ($\sim 10^2$) events. The background shape in sparsely populated subsets is expected to have large uncertainty. This uncertainty, however, is offset by the fact that the background fraction is small, so its effect on the measurement is also small.

In this analysis, these issues are not dealt with. The modeling of both the signal and background events is, therefore, not as accurate as it could be. The author has decided not to deal with these issues in this analysis because it would add a degree of detail that is not warranted given the predicted Run I measurement resolution (chapter 8). In other words, there are not enough events in Run I to meaningfully compare the experimental data with standard model predictions; given this, there is no point trying to refine the measurement by improving the signal and background model. Rather, this analysis focuses on the technical aspects involved in the use of a $t\bar{t}$ reconstruction algorithm to examine the kinematics of the semileptonic decay of the top quark. It is the author's hope that the technical details discussed in this thesis will be used as a starting point in similar analyses conducted in Run II. These Run II analyses can then focus on refining the monte carlo models.

Chapter 6

The $t\bar{t}$ Reconstruction Algorithm

6.1 Introduction

One of the chief aims of this thesis is to examine the kinematics of the semileptonically decaying top quark, t_ℓ, in the top quark rest frame. To do this, the t_ℓ momentum in the lab frame must be known. Obtaining this from the raw data sample, however, is not straightforward. The t_ℓ momentum is the sum of the decay product momenta: $t_\ell \rightarrow \ell + \nu_\ell + b_\ell$ ($\ell = e^\pm$ or $\mu^\pm$, ν_ℓ = neutrino partner of ℓ, and $b_\ell = b$-quark). Although the momentum of ℓ is usually well measured, that of ν_ℓ is incompletely known (the transverse component is given only poorly by the missing transverse momentum $\vec{\not{E}}_T$, and the longitudinal component is not known), and the momentum of b_ℓ is unknown because one does not know which of the four or more jets in the event originates from b_ℓ. Thus the momentum of t_ℓ is incompletely known. It can, however, be estimated fairly well using the $t\bar{t}$ reconstruction algorithm. The implementation and performance of this algorithm is described in this chapter.

6.2 The Algorithm

The $t\bar{t}$ reconstruction algorithm described here is that used by the CDF collaboration in the measurement of the top quark mass in the lepton + jets channel [44]. The idea underlying the algorithm is this:

1. Assume that the physics objects ($e^\pm$ or $\mu^\pm$, $\not{E}_T$, jets) originate from $p\bar{p} \rightarrow t\bar{t}$ in the lepton + jets channel.

2. Note the following mass constraints on the event subsystems: (1) the mass of the semileptonically decaying top quark t_ℓ and the hadronically decaying top quark t_h are approximately equal; (2) the invariant mass of the charged lepton and neutrino is approximately equal to the W mass; (3) the invariant mass of the light quark pair that originates from the hadronic decay of W is approximately equal to the W mass.

3. Make use of the assumption in step 1 and the mass constraint relations in step 2 to obtain estimates of the momenta of the decay products of t and $\bar{t}$.

In order to explain how the algorithm works, it will be useful to trace how the decay product momenta change from the parton-level to the final estimate output by the algorithm.

6.2.1 From the Parton to the Raw Data

At the parton-level, the $t\bar{t}$ decay in the lepton + jets channel is denoted as follows:

$$t_\ell \rightarrow b_\ell + \ell + \nu_\ell$$
$$t_h \rightarrow b_h + W_d + W_u$$

The symbols are defined as follows:

t_ℓ	Semileptonically decaying top quark
b_ℓ	b-quark from t_ℓ decay
ℓ	$e^\pm$ or $\mu^\pm$
ν_ℓ	Neutrino
t_h	Hadronically decaying top quark
b_h	b-quark from t_h decay
W_d	Down-type quark from hadronic W decay
W_u	Up-type quark from hadronic W decay

These decay products are found in the detector as *raw physics objects* (see chapter 3). The association between the parton-level objects and the raw physics objects is given below:

Parton-Level		Raw Physics Objects
ℓ	$\rightarrow$	ℓ
ν_ℓ	$\rightarrow$	$\not\!\!E_T$
b_ℓ		$j1$
b_h		$j2$
W_d	$\rightarrow$	$j3$
W_u		$j4$
		$\vdots$
		jn

The charged lepton ℓ is found in the detector as is — *i.e.* electrons and muons are identified reliably by the detector. The neutrino escapes detection; its presence is inferred from the imbalance in the transverse energy deposited in the calorimeters. The quarks give rise to jets. The symbols $j1$, $j2$, *etc.* denote the jets, numbered in descending order of raw transverse energy. Since there are four quarks at the parton-level, one may expect exactly four jets. However, because of gluon radiation, fluctuation in the fragmentation of the quarks, multiple interactions, and various other factors, more than four jets can sometimes be found in the raw data sample. Also, the four quarks can give rise to events with less than four jets because the jet from one of the quarks fails the jet acceptance cut; such events are rejected in this analysis.

6.2.2 Jet-quark Combinatorics

When the quarks b_ℓ, b_h, W_d, and W_u hadronize to form jets, their identity become obscured. That is, if there are no b-tagged jets in the event, then one cannot tell which jet originates from which quark. If there are b-tagged jets, then one can with confidence rule out the light quarks W_d and W_u as likely candidates for these jets; however, one cannot tell whether a given tagged jet originates from b_ℓ or b_h. One of the most important tasks performed by the $t\bar{t}$ reconstruction algorithm is the matching of the jets to the quarks.

The first step in implementing this matching is to consider all possible combination of jets to quarks. In order to simplify the algorithm, it is assumed that the four quarks give rise to the four leading jets $(j1, j2, j3, j4)$.[1] The jets $(j1, j2, j3, j4)$ are to be assigned

[1] Not infrequently, gluon jets can be among the four leading jets. In such cases, jets originating from the $t\bar{t}$ decay could be among the fifth or lower jets. Whether or not to include these extra jets as candidates is a matter of balancing cost and benefit — the cost of increased combinatorics *versus* the benefit of finding

the quark labels b_ℓ, b_h, W, W. Note that the labels for the light quark jets are identical — *i.e.* the distinction between the up-type quark W_u and down-type quark W_d is not made. This is because there is no reliable way to tell whether a jet originates from a light up-type or down-type quark. The combinatorics problem is to find all possible ways of assigning the labels b_ℓ, b_h, and W to the four jets $(j1, j2, j3, j4)$. The total number of possible combinations depends on the number of b-tagged jets:

0 Tags

All jets can be assigned all quark labels. There are 4 ways to assign b_ℓ to the four jets, 3 ways to assign b_h to the remaining three jets, and one way to assign the two identical labels W to the remaining two jets. This gives a total of 12 combinations.

1 Tag

The three untagged jets can be assigned any quark label; the tagged jet can be assigned only b_ℓ or b_h. If b_ℓ is assigned to the tagged jet, there are three ways of assigning b_h to the untagged jets, and one way to assign W to the remaining jets, giving a total of 3 combinations. The same argument holds for b_h assigned to the tagged jet. The total number of combinations is, therefore, 6.

2 Tags

b_ℓ can be assigned to one of the tagged jets, and b_h to the other, and vice versa. There are, therefore, only 2 combinations.

The possibility of more than two tagged jets is not considered because such events are not observed in the data.

Let us see where this discussion is leading. It will be shown later that, for each jet-quark combination, there are two solutions for the longitudinal component of the neutrino's momentum. Thus, there are 24, 12, and 4 possible candidate solutions for events with 0, 1, and 2 b-tags, respectively. Each candidate solution is assigned the quantity χ^2, which is an indicator of how likely a given configuration is the correct one. The χ^2 of each configuration

jets from $t\bar{t}$ that would otherwise be lost. An internal study in the CDF collaboration [58] has shown that the costs outweigh the benefits.

is minimized, and the configuration with the smallest minimized χ^2 is chosen as the output of the $t\bar{t}$ reconstruction algorithm.

6.2.3 Standard Raw Data Corrections

The raw data are used to trigger on events during data acquisition. Since triggering is a time-critical operation, only minimal processing takes place in the conversion of detector signals to raw data. After the data are written to tape, further processing of the raw data is necessary in order to more accurately reconstruct the 4-momenta of the particles that give rise to events. The standard set of corrections applied to the physics objects in lepton + jets events is given in chapter 3.

6.2.4 Special Jet Correction

The standard jet correction was developed and optimized primarily for the study of events with $2 \sim 4$ light quark- and gluon-jets [40], [42]. There is, therefore, no reason to believe that they provide adequate corrections for $t\bar{t}$ events in the lepton + jets decay channel, which usually have $3 \sim 6$ jets, two of which are b-quark jets. A study done by CDF [43] shows, indeed, that the standard correction is insufficient. The following factors are believed to be (in varying degrees) responsible:

- The standard correction does not account for the energy carried away by the neutrino in the semileptonic b-quark decay $b \rightarrow e\ \nu\ X$. In the decay $b \rightarrow \mu\ \nu\ X$, the standard correction fails to account for the energy of the neutrino and the muon (the muon deposits only a fraction of its energy in the calorimeter).

- The standard out-of-cone correction and underlying event correction are obtained for QCD events that have, on average, smaller number of jets and weak jet P_T spectrum compared to what is found in lepton + jets $t\bar{t}$ events. One would expect these corrections to depend on the number and hardness of jets.

- The monte carlo-dependent part of the standard correction is based on studies using ISAJET [49]; $t\bar{t}$ events, on the other hand, are simulated using HERWIG [47], [48].

These two event generators use very different schemes for simulating the hadronization process.

- The standard absolute jet energy scale correction was obtained using a monte carlo sample with flat parton-level P_T spectrum[42]. Lepton + jets $t\bar{t}$ events, however, have parton-level P_T spectrum with significant structure. The authors of the reference [43] claim that amount of absolute energy scale correction depends on the parton-level P_T distribution.

- The standard relative jet energy scale correction is obtained by balancing dijet events[42]. Lepton + jets $t\bar{t}$ events, however, typically have $3 \sim 6$ jets.

The relative importance of each factor in causing the standard correction to be insufficient is not clearly known. In spite of this, the insufficiency is a clearly established fact — the standard-corrected jet E_T in simulated $t\bar{t}$ events are systematically off from the true quark E_T[43]. The standard corrections can be improved by adding this systematic shift to the standard-corrected jet E_T. This new jet E_T is (on average) a better estimate of the true E_T. This new correction that is applied on top of the standard one is referred to in CDF as the "AA" correction[2].

The amount of systematic shift depends a lot on the nature of the jet. Thus the AA correction distinguishes between the following four types of jets:

- Generic b-quark jets. This includes b-quarks decaying hadronically and b-quarks decaying in the electron/muon channel, but whose electron/muon is not identified by the soft lepton tagging algorithm.

- $b \rightarrow e + \nu + X$ identified by the soft electron tagging algorithm.

- $b \rightarrow \mu + \nu + X$ identified by the soft muon tagging algorithm.

- Light quark jets.

[2] "AA" is in honor of the authors of the study: *A*llessandra Caner and *A*vi Yagil. They are both members of the CDF collaboration.

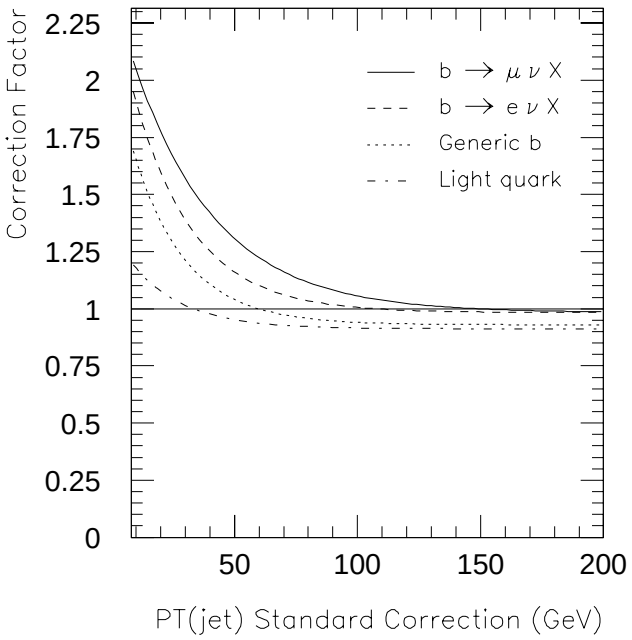


Figure 6.1: AA corrections as a function of E_T(jet) after standard correction.

Figure 6.1 shows the correction factors for each jet category as a function of jet E_T after standard correction. These corrections are applied in the following manner.

- If a jet is assigned to b_ℓ or b_h and it does not have a soft lepton tag, the generic b-quark correction factor is applied.
- If a jet is assigned to b_ℓ or b_h and it has a soft electron tag, the correction factor for $b \to e\ \nu\ X$ is applied. If it has a soft muon tag, the factor for $b \to \mu\ \nu\ X$ is applied.
- If a jet is assigned to W, then the correction factor for light quarks is applied.

It should be emphasized that the generic b-quark correction can be applied to jets without a b-tag; the correction applied to a jet depends on the quark label assigned to it, not on the presence or absence of a b-tag. Also, in events with one or no b-tag, the correction applied to an untagged-tagged jet is not unique because it could be assigned to a b-quark or to a light quark.

6.2.5 *The Neutrino Momentum*

The neutrino momentum is obtained in two steps: (1) *calculate* the transverse components; (2) *choose* the longitudinal component. The distinction between *calculate* and *choose* is significant, as will be discussed presently. Step (1) is the same throughout the $t\bar{t}$ reconstruction algorithm. Step (2), on the other hand, involves two separate procedures, one for the initial estimate of the neutrino momentum, and the other used during the χ^2 minimization process.

Let us first discuss the definition of the neutrino transverse momentum, $\vec{P}_T(\nu)$. It is defined to balance all of the other objects in the event:

$$\begin{aligned} \vec{P}_T(\nu) \;=\; - \;\; & (\vec{P}_T(\ell) \\ & +\vec{P}_T(\text{4 leading jets}) \\ & +\vec{P}_T(\text{extra jets}) \\ & +\vec{P}_T(\text{very forward jets}) \\ & +\vec{P}_T(\text{unclustered energy})) \end{aligned} \tag{6.1}$$

The terms corresponding to the "extra jets" and "very forward jets" account for the jets against which the $t\bar{t}$ system recoils. The "unclustered energy" term is described in detail in the next section; roughly speaking, it accounts for the $t\bar{t}$ transverse recoil momentum from all factors other than jets.

Now let us describe how the longitudinal neutrino momentum $P_z(\nu)$ is chosen. First, let us consider the *initial estimate* — *i.e.* the values assigned to $P_z(\nu)$ at the beginning of the χ^2 minimization process. The longitudinal momentum $P_z(\nu)$ is *chosen* to be such that the invariant mass of the ℓ-ν system is equal to the W mass: $(p_\ell + p_\nu)^2 = {M_W}^2$. Solving this equation for $P_z(\nu)$, one obtains:

$$P_z(\nu) = A \pm \frac{\sqrt{D^2}}{\sin^2\theta_\ell} \tag{6.2}$$

$$A = R\frac{\cos\theta_\ell}{\sin^2\theta_\ell} \tag{6.3}$$

$$D^2 = R^2 - \sin^2\theta_\ell \left|\vec{P}_T(\nu)\right|^2 \tag{6.4}$$

$$R = \frac{{M_W}^2}{2E_\ell} + \cos\Delta\phi_{\ell\text{-}\nu}\sin\theta_\ell \left|\vec{P}_T(\nu)\right| \tag{6.5}$$

The angle θ_ℓ is the polar angle of the charged lepton, E_ℓ is the charged lepton energy, and $\Delta\phi_{\ell\text{-}\nu}$ is the azimuthal angular separation between ℓ and ν_ℓ (*i.e.* the angle between $\vec{P}_T(\ell)$ and $\vec{P}_T(\nu)$). All of these quantities are defined in the lab frame. It is seen from equation 6.2 that, in general, $P_z(\nu)$ has two solutions. Also, in general, different jet-quark combinations give rise to somewhat different values of $P_z(\nu)$ because the AA correction applied to each jet depends on the jet-quark assignment. In other words, the vector sum of the jet transverse momenta depends somewhat on the type of AA correction applied to each jet; differences in the jet transverse momenta propagates to differences in $P_z(\nu)$.

Not infrequently (in about 36% of events), the quantity D^2 in equation 6.4 is negative, in which case $P_z(\nu)$ is complex. These complex solutions are primarily the result of either $\left|\vec{P}_T(\nu)\right|$ or $\Delta\phi_{\ell\text{-}\nu}$ being overestimated. Under these circumstances, the two $P_z(\nu)$ solutions are taken as follows:

$$P_z(\nu) \;= A \pm 20\,GeV \tag{6.6}$$

This prescription is only a technical work-around to get the χ^2 minimization process started. Exactly what real numbers are assigned in place of the complex solutions is unimportant so long as they are chosen such that when χ^2 has two minima, the minimization process finds them both. Monte carlo studies have shown that the choice of $P_z(\nu)$ given in equation 6.6 satisfies these requirements. See appendix B for more details.

Now let us consider how $P_z(\nu)$ is chosen during the χ^2 minimization process. In this process, $P_z(\nu)$ is one of the parameters involved in minimizing χ^2. The quantity χ^2 consists of two parts:

$$\chi^2 = \chi^2_{mass} + \chi^2_{kinematics} \tag{6.7}$$

χ^2_{mass} is the term that favors configurations that have event topology consistent with the $t\bar{t}$ hypothesis. $\chi^2_{kinematics}$, on the other hand, is the "penalty term" — *i.e.* the energies of the physics objects are allowed to stretch or shrink in order to decrease χ^2_{mass}, but at the cost

of increased $\chi^2_{kinematics}$. See section 6.2.7 for details. The important point to note in this context is the details of χ^2_{mass}:

$$\chi^2_{mass} = \chi^2(t_\ell) + \chi^2(t_h) + \chi^2(W_\ell) + \chi^2(W_h) \tag{6.8}$$

The first two terms favor configurations where the mass of t_ℓ and t_h are close together. The third term favor configurations where the ℓ-ν_ℓ invariant mass is close to M_W. The final term favors configuration where the two-jet system presumed to originate from the hadronically decaying W has invariant mass close to M_W. The choice of the neutrino longitudinal momentum $P_z(\nu)$ influences the first three terms of equation 6.8. The choice is essential in determining the size of $\chi^2(W_\ell)$, while it is moderately important in setting the size of $\chi^2(t_\ell) + \chi^2(t_h)$.[3]

One of the most essential tasks performed by the $t\bar{t}$ reconstruction algorithm is the minimization of χ^2. At the start of the minimization process, χ^2_{mass} is large — typically on the order of 10^2 — while $\chi^2_{kinematics}$ is exactly zero. χ^2_{mass} is decreased by: (1) stretching or shrinking the physics object energies; and (2) choosing $P_z(\nu)$ appropriately. In order to keep χ^2_{mass} as small as possible, $P_z(\nu)$ must almost always be chosen such that the invariant mass of the ℓ-ν_ℓ system is very close to M_W, since this keeps $\chi^2(W_\ell)$ very close to zero.

6.2.6 Unclustered Energy

The unclustered energy $\vec{P}_T(uce)$ is a two-component vector that estimates the total transverse momentum deposited in the calorimeter from all sources except leptons and jets. The beam line component of the unclustered energy is defined to be zero.[4] See appendix C for a discussion of how $\vec{P}_T(uce)$ is estimated from the physics object momenta.

Before leaving this section, let us consider what role the unclustered energy plays in the $t\bar{t}$ reconstruction algorithm. In appendix C, it is stated that $\vec{P}_T(uce)$ is composed of two parts: (1) the *physical* part; and (2) the *resolution* part. The physical part

[3] It may seem counterintuitive that ν_ℓ, which is a t_ℓ decay product, should influence the size of $\chi^2(t_h)$. $\chi^2(t_\ell)$ and $\chi^2(t_h)$, however, are related to each other through a common parameter, and, therefore, $P_z(\nu)$ influences size of both terms. See section 6.2.7 for details.

[4] No attempt is made to estimate the longitudinal component of the system comprising the unclustered energy because there is no reliable way to do so.

$\vec{P}_T(physical)$ is due to actual particles depositing energy in the calorimeters. The resolution part $\vec{P}_T(resolution)$, on the other hand, is due to the finite resolution of the physics object transverse momenta. In order to illustrate the role of the unclustered energy, let us consider two mutually exclusive scenarios:

The Physical Part is More Important than the Resolution Part

In this case, $\vec{P}_T(uce)$ is a more-or-less accurate estimate of the true physical unclustered energy. If this term is left out of equation 6.1 for $\vec{P}_T(\nu)$, then the estimated value of $\vec{P}_T(\nu)$ would, on average, be less accurate than if $\vec{P}_T(uce)$ were included. Thus, in this situation, $\vec{P}_T(uce)$ plays the role of directly improving the estimate of the neutrino transverse momentum.

The Resolution Part is Comparable to, or More Important than, the Physical Part

In this case, $\vec{P}_T(uce)$ is a lousy estimate of the true physical unclustered energy. Thus the addition of the term $\vec{P}_T(uce)$ to equation 6.1 would not significantly improve the initial estimate of $\vec{P}_T(\nu)$. However, because $\vec{P}_T(uce)$ has a large measurement error, this term can be altered significantly during the χ^2 minimization process (see section 6.2.7). This, in turn, allows greater freedom in choosing the transverse and longitudinal components of the neutrino's momentum. This freedom is a reflection of the uncertainty in the neutrino's momentum, and is important in order not to overconstrain it.

Of these two scenarios, the latter accurately describes the situation in this analysis. Thus, in this analysis, the unclustered energy serves as an error term in estimating the neutrino momentum.

6.2.7 χ^2

At the starting point of the $t\bar{t}$ reconstruction algorithm, there are 24, 12, and 4 candidate solutions in events with 0, 1, and 2 b-tagged jets. From all of these candidates, a unique output is to be chosen. Preferably, the chosen output is the configuration in which the decay product momenta most closely resemble the parton-level momenta. The

simplest way to choose a unique output is to select, event-by-event, a random configuration. Then, in a fraction of events, the "best" solution (*i.e.* the solution whose decay product momenta most closely resemble those at the parton-level) is guaranteed to be chosen by pure luck. One should, however, be able to do better by making use of the event topology that is characteristic of $t\bar{t}$ events in the lepton + jets channel — *i.e.* that (1) the mass of t_h and t_ℓ must be approximately equal; (2) that the invariant mass of the ℓ-ν_ℓ system is approximately equal to the W mass M_W; and (3) the invariant mass of the light quark pair presumed to originate from the hadronically decaying W is approximately equal to M_W. By quantifying these mass conditions, one should be able to use them to select the correct configuration more often than by random selection.

In the $t\bar{t}$ reconstruction algorithm, the mass condition is quantified by a number referred to, suggestively, as χ^2.[5] The quantity χ^2 consists of many terms, which can be placed in two categories: the *mass terms* and the *kinematic terms*:

$$\chi^2 = \chi^2_{mass} + \chi^2_{kinematic} \tag{6.9}$$

The mass terms are defined such that χ^2_{mass} is small when the mass conditions are well-satisfied, and large when they are not. The kinematic terms are defined so that $\chi^2_{kinematic}$ is zero when the magnitude of the physics object momenta have their initial value, and increases as they are varied. The amount of increase in $\chi^2_{kinematics}$ is determined by the expected errors of the momenta. A detailed quantitative definition of χ^2 is given in appendix D.

The strategy used by the $t\bar{t}$ reconstruction algorithm to choose a unique output configuration is the following. At the starting point of the algorithm, $\chi^2_{kinematic}$ is zero because the energy scale of the physics objects have not been changed. The term χ^2_{mass}, however, is defined so that, typically, the starting value is on the order of 10^2. By stretching or shrinking the energy scale of the objects in the event, χ^2_{mass} can usually be made close to zero; the cost of doing this is the increase in $\chi^2_{kinematic}$, which is typically on the order

[5] This quantity "χ^2" is not χ^2 in the strict, statistical sense. Its definition involves ideas that are analogous to those employed in χ^2. However, the statistical properties of the "χ^2" defined in the text is not as simple as that of χ^2 in the usual sense of the word.

of $10^0 \sim 10^1$ by the time the total χ^2 is minimized. The configuration with the smallest minimized χ^2 is chosen as the unique output.

Intuitively, one can expect this choice of output to correspond to the correct configuration more often than when the output is chosen randomly. This is because, at the parton-level, the correct configuration always satisfies the mass conditions exactly, whereas the incorrect configurations almost always do not satisfy them. After the parton-level momenta are smeared by the measurement process, the correct configuration does not always satisfy the mass conditions well. Also, incorrect configurations that do not satisfy the mass conditions at the parton-level may end up satisfying them well after the parton-level momenta are smeared. Even so, it seems reasonable to expect that the correct configuration will more often match the mass conditions than incorrect configurations will. In order to prove that this is actually the case, one would have to demonstrate that the algorithm chooses the correct configuration more often than in random selection. That this is so, and the extent to which it is, is shown in the next section, which discusses the algorithm's performance.

6.3 The Algorithm's Performance

One way to measure the degree of success of the $t\bar{t}$ reconstruction algorithm is by examining how much more often it outputs the "best configuration" compared to the case where the output is chosen at random. In this situation, "best configuration" refers to the following:

- The configuration in which the greatest number of quarks are assigned to the correct jets.
- If χ^2 has two minima, the configuration which corresponds to the minimum whose $P_z(\nu)$ is closest to the true value.

Although these criteria for "correctness" of output solution may be appropriate in the top mass measurement, in which the properties of both the t_ℓ and t_h decay products play essential roles, they are not very useful in this study, where only the t_ℓ decay products are of

central importance. The problem with the criteria, from the point of view of this analysis, is the following: too many factors that are not essential in determining measurement resolution of the t_ℓ decay product kinematics are involved. In other words, the criteria given above are unnecessarily complicated, and they obscure the aspects of the $t\bar{t}$ reconstruction algorithm that are important in determining the measurement resolutions.

The first step in finding a more suitable measure of the $t\bar{t}$ reconstruction algorithm's performance is to focus on the aspects of the algorithm that deal with t_ℓ. They are the following:

- Charged lepton energy scale
- Neutrino momentum
- b_ℓ energy scale
- Whether or not b_ℓ is matched to the correct jet

In chapter 9, it is shown that the last item is by far the most important source of measurement degradation. Therefore it will be used as the sole measure of the performance of the $t\bar{t}$ reconstruction algorithm. Specifically, the fraction of events with b_ℓ matched to the correct jet, $f_{b\ell}$, will be used.

Figure 6.2 shows $f_{b\ell}$ in events with 0, 1, and 2 b-tagged jets. The points show $f_{b\ell}$ from the algorithm; the hatched histogram shows $f_{b\ell}$ when the output is chosen at random; the open histogram shows the maximum possible $f_{b\ell}$.[6] Table 6.1 shows the numerical values of $f_{b\ell}$ corresponding to those shown in figure 6.2. Clearly, the algorithm does much better than randomly selecting the output configuration.

The following are some technical comments on figure 6.2:

- The maximum possible $f_{b\ell}$ is less than 100% because in a fraction of events, the jet from b_ℓ either fails the jet acceptance cut or is not among the four leading jets.

[6] The maximum $f_{b\ell}$ is somewhat less than 100% because in some events the jet from b_ℓ is not among the four leading jets; in such events, the fitter cannot possibly assign the correct jet to b_ℓ.

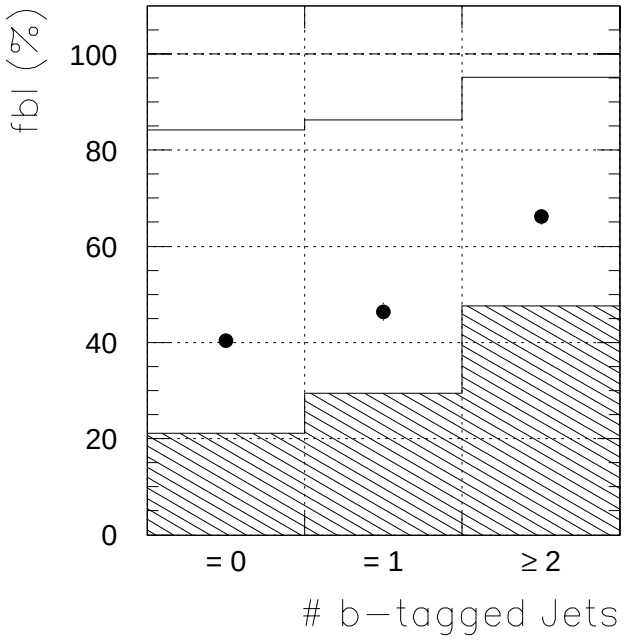


Figure 6.2: $f_{b\ell}$ vs. # b-tagged jets. Points = expected from algorithm; hatched histogram = random; open histogram = maximum. See text for the reason why the maximum $f_{b\ell} < 100\%$.

	# B-TAGGED JETS		
	$= 0$	$= 1$	≥ 2
ALGORITHM	40.4	46.4	66.2
MAXIMUM	84.2	86.3	95.1
RANDOM	21.1	29.4	47.6

Table 6.1: The predicted value of $f_{b\ell}$ (in %) from the $t\bar{t}$ reconstruction algorithm. Also shown are the maximum possible $f_{b\ell}$, and $f_{b\ell}$ when the output is chosen randomly.

- In events with 2 or more b-tagged jets, one might imagine that $f_{b\ell}$, by definition, must have a maximum value of 100%. It is, however, less than 100% because of the following possibility: the jet from b_ℓ is missed for the reasons stated above; a charm or light quark (from the hadronic W decay or from a gluon) can be tagged, although with much smaller efficiency compared to b jets; this "tagged" jet is treated as a b-quark jet by the algorithm.

Chapter 7

Observables

7.1 Introduction

The objective of this thesis is to examine, in the semileptonic decay of the top quark, the energy and angular distribution of the charged lepton in the top rest frame. This statement of the objective raises the following questions:

- Why examine the decay product kinematics in just the semileptonic decay channel of the top quark, and not in the hadronic channel?
- Why examine just the charged lepton energy and angular distributions, and not those of the neutrino or the b-quark?
- Why examine the top quark decay product kinematics in the top rest frame rather than in the lab frame?

These questions, essentially, are about the choice of observables. This chapter is devoted to answering these questions, and to defining observables that allow one to compare the experimental data with the standard model prediction of the top quark decay.

Before these questions can be answered and observables can be defined, however, a theoretical framework for describing the top quark decay is necessary. To do this, the first step is to find the number of parameters (*i.e.* energy and angles) that are necessary and sufficient to completely describe the top quark decay in the top rest frame. Once a set of parameters is chosen, it is necessary to examine the standard model prediction of the

distribution of these parameters. With these theoretical framework in place, one can define observables that can be used to compare the experimental data with the standard model prediction.

The first several sections of this chapter are devoted to building up this theoretical framework. After this, the observables are defined, and the distribution of these observables as predicted by monte carlo models are shown. At the end of the chapter, the questions raised above concerning the choice of observables are answered.

7.2 Parameters Describing the Top Quark Decay in the Top Rest Frame

The top quark decay is a sequential 3-body process, where first $t \rightarrow b + W$, and then $W \rightarrow$ a lepton or quark pair:

$$t \rightarrow b + W \hookrightarrow \begin{cases} \ell + \nu_\ell \\ \text{or} \\ W_d + W_u \end{cases}$$

The symbol ℓ denotes e, μ, or τ; ν_ℓ is the neutrino partner of ℓ; W_d and W_u are down- and up-type quarks, respectively, from the W decay. The goal of this section is to determine how many parameters (energies and angles) are necessary and sufficient to describe this process in the rest frame of t.

The upper limit for the number of parameters is nine: there are three particles in the decay, and the momentum of each particle is described by two angles and one energy. However, the energy and angles of the particles are correlated because:

1. b and W originate from the decay of t, and ℓ and ν_ℓ (or W_d and W_u) originate from the decay of W.

2. The top and W masses are fixed (their widths can be ignored for the purpose of this study).

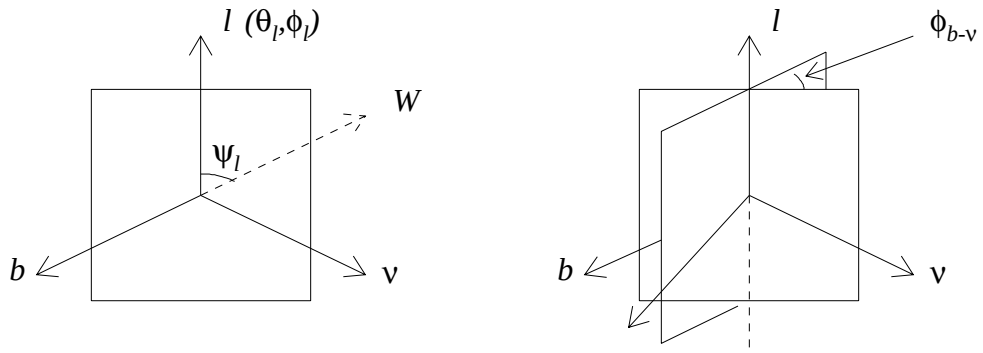


Figure 7.1: The four angles θ_ℓ, ϕ_ℓ, $\phi_{b\text{-}\nu}$, and ψ_ℓ that completely describe the top quark decay. The angles θ_ℓ, ϕ_ℓ, and $\phi_{b\text{-}\nu}$ are needed to span all possible orientation of the decay plane, while ψ_ℓ fixes the magnitude and relative orientation of the momenta in the plane. The angle ψ_ℓ can be replaced with the opening angle between ℓ and ν_ℓ, since there is a one-to-one relationship between the two angles.

3. They are to be described in the top rest frame.

Therefore the number of parameters is less than nine. In the following, it will be shown that four parameters are necessary and sufficient. For the sake of concreteness, the decay $t \rightarrow \ell + \nu_\ell + b$ will be used. To translate the result to hadronic decay, ℓ and ν_ℓ are replaced with W_d and W_u, respectively.

The fact that the top quark decay is to be described in the top rest frame implies that the sum of the momentum of ℓ, ν_ℓ, and b is zero. Thus the three momentum vectors $\vec{p}_\ell$, $\vec{p}_\nu$, and $\vec{p}_b$ are coplanar. Three angles are necessary to span all possible orientation of the plane containing these vectors. For reasons that will become clear in the next section, the angles will be chosen as θ_ℓ, ϕ_ℓ, and $\phi_{b\text{-}\nu}$. Figure 7.1 illustrates these angles. The angles θ_ℓ and ϕ_ℓ are the polar and azimuthal angle of $\vec{p}_\ell$ defined relative to some coordinate system (the exact specification of which is not important in the context of this discussion), while $\phi_{b\text{-}\nu}$ describes the orientation of the plane when it is rotated about the axis containing $\vec{p}_\ell$.[1]

At this point, the following quantities have not yet been fixed:

[1] This angle is called $\phi_{b\text{-}\nu}$ for the following reason. Imagine the vectors $\vec{p}_\ell$, $\vec{p}_\nu$, and $\vec{p}_b$ forming a fork, with $\vec{p}_\ell$ as the handle and $\vec{p}_\nu$ and $\vec{p}_b$ as the prongs. Then $\phi_{b\text{-}\nu}$ is the angle that describes the rotation of the b-ν prongs around the axis containing $\vec{p}_\ell$.

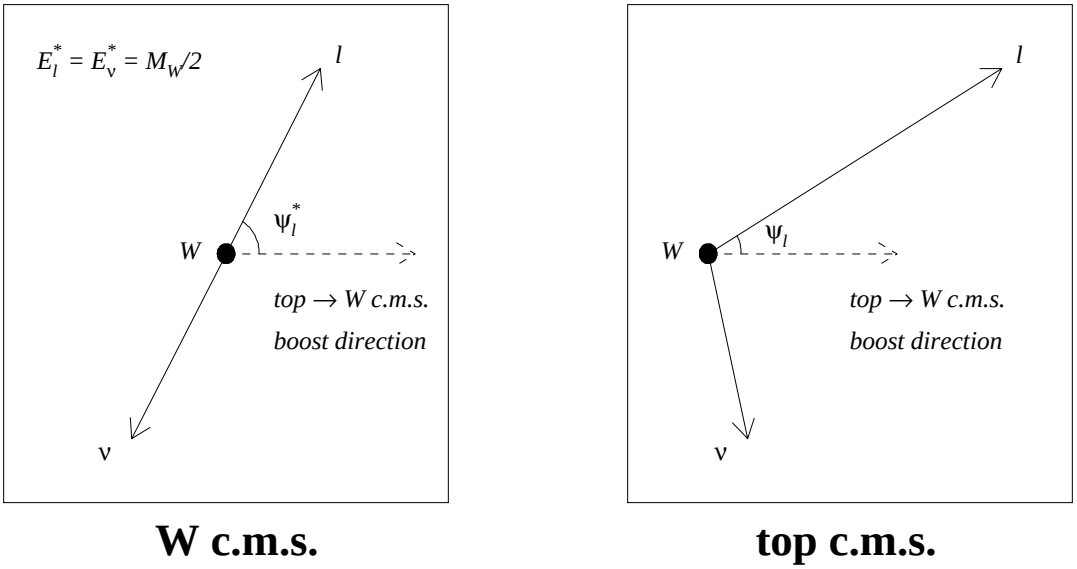


Figure 7.2: W decay kinematics in the W and top rest frames.

- The orientation of $\vec{p}_{\nu_\ell}$ and $\vec{p}_b$ relative to $\vec{p}_\ell$
- The magnitude of $\vec{p}_\ell$, $\vec{p}_{\nu_\ell}$, and $\vec{p}_b$

As it turns out, by specifying the orientation of $\vec{p}_{\nu_\ell}$ relative to $\vec{p}_\ell$, everything else is fixed. To see that this is so, it is useful to examine the decay kinematics of ℓ and ν_ℓ in the W rest frame (see figure 7.2). The energy of ℓ and ν_ℓ in the W rest frame, E_ℓ^* and E_ν^*, are simply equal to $M_W/2$ (lepton masses are neglected). Their momenta are equal and opposite. The angle ψ_ℓ^* is that between the charged lepton momentum and the boost direction from the top to the W rest frame.

The energy of ℓ and ν_ℓ in the top rest frame, E_ℓ and E_ν, are related to E_ℓ^* and E_ν^* by a Lorentz transformation. After some rearrangements, one obtains the following:

$$E_\ell = \frac{1}{2}(E_W + P_W \cos\psi_\ell^*) \tag{7.1}$$

$$E_\nu = \frac{1}{2}(E_W - P_W \cos\psi_\ell^*) \tag{7.2}$$

The quantities E_W and P_W are the W energy and momentum in the top rest frame. They are constant: $E_W = ({m_t}^2 + {M_W}^2 - m_b^2)/(2m_t)$, $P_W = \sqrt{E_W^2 - M_W^2}$. Because they are constant, E_ℓ is related to $\cos\psi_\ell^*$ by a simple linear transformation. In the equation for E_ν, the fact that the neutrino momentum is equal and opposite that of ℓ in the W rest frame is used. The angle ψ_ℓ between the momentum of ℓ and that of W in the top rest frame, and the corresponding angle ψ_ℓ^* in the W rest frame, are also related to one another through a Lorentz transformation. After rearrangements, it looks like the following:

$$\cos\psi_\ell = \frac{\cos\psi_\ell^* + \beta_W}{1 + \beta_W \cos\psi_\ell^*} \tag{7.3}$$

The quantity β_W is the relativistic speed of W in the top rest frame; it is a constant equal to P_W/E_W. Similarly, the angle between the momentum of ν_ℓ and that of W in the top rest frame is related to $\cos\psi_\ell^*$ according to the following:

$$\cos\psi_\nu = \frac{-\cos\psi_\ell^* + \beta_W}{1 - \beta_W \cos\psi_\ell^*} \tag{7.4}$$

The relations in equations 7.1 through 7.4 show that: (1) the magnitude of $\vec{p}_\ell$ and $\vec{p}_{\nu_\ell}$; and (2) the angle between ℓ and ν_ℓ; are fixed by the single parameter $\cos\psi_\ell^*$. Since the angle between $\vec{p}_\ell$ and $\vec{p}_{\nu_\ell}$ is just $\psi_\ell + \psi_\nu$, one can also say that the magnitude of $\vec{p}_\ell$ and $\vec{p}_{\nu_\ell}$ and the angle $\cos\psi_\ell^*$ is fixed by specifying the angle between $\vec{p}_\ell$ and $\vec{p}_{\nu_\ell}$. Finally, once the direction and magnitude of $\vec{p}_\ell$ and $\vec{p}_{\nu_\ell}$ are fixed, the direction and magnitude of $\vec{p}_b$ is fixed by momentum conservation.

Let us summarize the results of this section. Four parameters are necessary and sufficient to completely specify the decay of the top quark in the top rest frame. Three are needed to obtain all possible orientation of the plane containing $\vec{p}_\ell$, $\vec{p}_{\nu_\ell}$, and $\vec{p}_b$; one is needed to fix the orientation of the three vectors relative to one another, and to fix their magnitude. The three parameters for specifying the orientation of the plane can be chosen as $\cos\theta_\ell$, ϕ_ℓ, and $\phi_{b\text{-}\nu}$. The one remaining parameter can be chosen as the opening angle between $\vec{p}_\ell$ and $\vec{p}_{\nu_\ell}$; other possible choices are $\cos\psi_\ell$, $\cos\psi_\nu$, $\cos\psi_\ell^*$, $\cos\psi_\nu^*$, E_ν, and E_ℓ — all of these parameters are related to one another by a one-to-one correspondence. A specific choice of parameters will be made on the basis of the form of the standard model prediction.

Before leaving this section, it should be noted that the standard model top quark decay is similar in many ways to the 3-body $(V-A)\times(V-A)$ decay of other fermions, such as the muon and tau lepton, and the c, and b quarks.[2] One critical difference, however, is the fact that the W from the top quark decay is on mass shell, whereas that from the other fermions is off mass shell. One consequence of this difference is the fact that the 3-body $(V-A)\times(V-A)$ decay of all other fermions are described by *five* parameters, the extra parameter being the virtual W mass. Appendix E discusses the similarities and differences between the $V-A$ decay of the top quark and all other fermions.

7.3 Standard Model Prediction of the Distribution of the Four Parameters

The tree-level standard model prediction of the distribution of the four parameters is obtained by summing the matrix elements for all possible Feynman diagrams for $p\overline{p} \rightarrow t\bar{t}$ for a given production and decay channel, and then squaring that sum. Which matrix element to use, however, depends on an assumption regarding the top quark lifetime and the typical timescale involved in the hadronization of quarks. Therefore, before any prediction regarding the four parameters can be made, this issue must be settled.

7.3.1 The Full Matrix Element Calculation Versus the Independent Decay Procedure

Let us take τ_{top} to be the top quark lifetime, and τ_{had} the typical timescale required for the hadronization of a quark. The top quark lifetime is $\tau_{top} = 1/\Gamma_{top}$, where:

$$\Gamma_{top} = \frac{G_F}{8\sqrt{2}\pi} |V_{tb}|^2 \, {m_t}^3 (1 - \frac{{M_W}^2}{{m_t}^2})^2 (1 + 2\frac{{M_W}^2}{{m_t}^2}) \tag{7.5}$$

For $m_t = 175\ GeV/c^2$ and $|V_{tb}| = 1$, $\Gamma_{top} = 1.6\ GeV$, and $\tau_{top} = 4.2 \times 10^{-25}$ sec. Radiative corrections are expected to decrease Γ_{top} by about 10%[62]. Thus Γ_{top} is expected to be about $1.4\ GeV$. The hadronization timescale, in contrast, is not a precisely defined quantity.

[2] The s- and d-quarks are not included in this list of fermions because they are too light for the spectator approximation to be valid.

It is a timescale on the same order of magnitude as the time required for light to travel the diameter of a proton [29]:

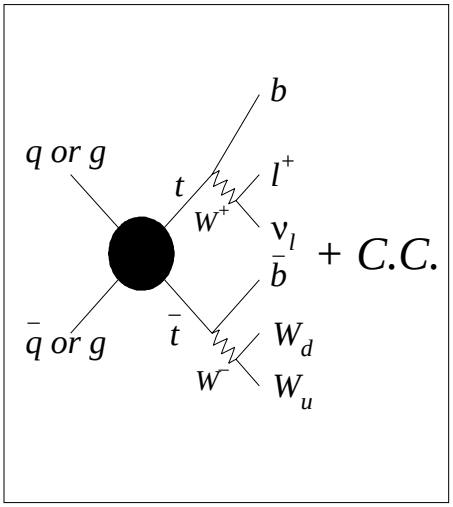
$$\begin{aligned} \tau_{had} &\approx (1fm)/c \\ &\approx (0.2 \cdot GeV)^{-1} \\ &\approx 3 \times 10^{-24} s \end{aligned}$$

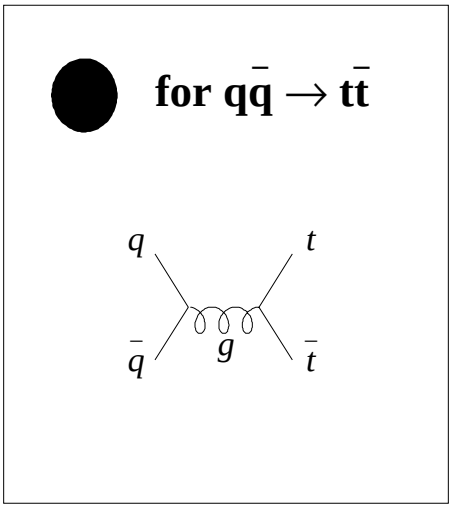
Therefore $\tau_{top} \approx \frac{1}{7} \times \tau_{had}$. This suggests that, usually, the top quark probably decays well before it hadronizes.

The length of τ_{top} relative to τ_{had} is critical since hadronization randomizes (1) the relative orientation of the t and $\bar{t}$ spin and (2) any phase coherence between matrix elements corresponding to various spin states of t and $\bar{t}$. Therefore if $\tau_{top} \ll \tau_{had}$, then the t and $\bar{t}$ spin dependence must be taken account of in order to correctly model the $p\bar{p} \rightarrow t\bar{t}$ production and decay process. On the other hand, if $\tau_{top} \gg \tau_{had}$, this spin dependence should be ignored. If $\tau_{top} \approx \tau_{had}$, then perhaps information regarding t and $\bar{t}$ spin may be only partially lost, and some special hybrid treatment may be necessary to deal with this situation.

Equation 7.6 suggests that the top quark decays well before it hadronizes. If this is the case, then the correct way to model the tree-level production and decay of t and $\bar{t}$ in $q\bar{q}$, $gg \rightarrow t\bar{t}$ is to: (1) evaluate the matrix elements for all spin configurations for all diagrams in figure 7.3; (2) sum them; and (3) square the sum. If, on the other hand, the top quark decays well after it hadronizes (in spite of equation 7.6), then a procedure, which will here be called the *independent decay procedure*, is appropriate for generating the production and decay of t and $\bar{t}$. The independent decay procedure is described below:

1. In the $t\bar{t}$ rest frame, the top quark momentum is generated according to the tree-level differential cross section $d\sigma^{q\bar{q}}/d\hat{t}$ or $d\sigma^{gg}/d\hat{t}$, depending on whether the $t\bar{t}$ is produced by a $q\bar{q}$ or gg pair ($\hat{t}$ is one of the Mandelstam variables). The differential cross section is obtained using the Feynman diagrams in the bottom of figure 7.3. See appendix G for details on the cross section formulas. The $\bar{t}$ momentum is equal and opposite that





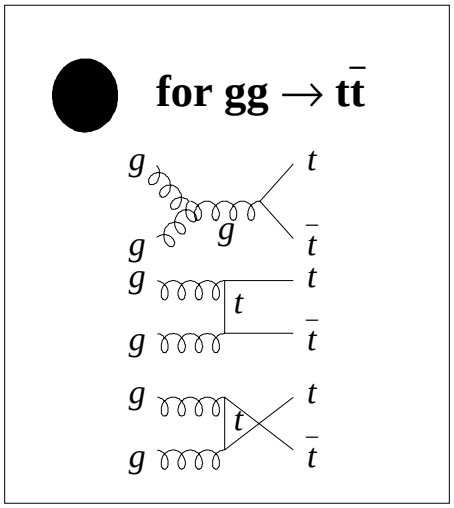


Figure 7.3: The Feynman diagrams for $q\overline{q}$ or $gg \to t\overline{t} \to \ell + \nu_\ell + b_\ell + b_h + W_d + W_u$. The circular hatched region in the top figure is to be replaced with the diagrams in the bottom figures.

of t. The lab frame momentum of t and $\bar{t}$ are obtained from the generated momentum by boosting along the beam line by negative the momentum of the incoming $t\bar{t}$ system.

2. In the top quark rest frame, the top quark is decayed using the Feynman diagram obtained by chopping off the portion starting from the top quark in the top of figure 7.3. The $\bar{t}$ decay in the $\bar{t}$ rest frame is performed in the same way. The t and $\bar{t}$ spin are assumed to be randomly oriented, and the t and $\bar{t}$ decay are performed independently.

3. The lab frame momenta of the top quark decay products are obtained by boosting the momenta in step 2 back to the lab frame using the lab frame top quark momentum generated in step 1. The same procedure is used to obtain the lab frame momenta of the decay products of $\bar{t}$.

As far as the author is aware at the time of this writing, all standard monte carlos available for modeling $p\bar{p} \rightarrow t\bar{t}$ use the independent decay procedure.[3] As of this writing, all studies of the top quark (that do not specifically deal with $t\bar{t}$ spin correlation effects) carried out by the CDF and D0 collaboration use standard monte carlos, and, therefore, they all assume the independent decay procedure. This study, too, will assume the independent decay procedure. This choice may seem irrational given that $\tau_{top} \ll \tau_{had}$. Yet, this choice is adopted for the following reasons:

- The decay product kinematics obtained from the independent decay procedure is very similar to that obtained from the full matrix element calculation (see [29] and appendix F). In fact, the choice of procedure does not make any detectable difference in the final results of this study.

- Whether or not hadronization takes place before or after the top quark decays is an issue that should be settled experimentally. Therefore, until the expectation that the top quark decays well before it hadronizes is demonstrated experimentally, it seems natural to choose the default hypothesis adopted so far by the CDF and D0 collaborations.

[3]Monte carlos that incorporate the full tree-level spin-dependent calculations have been developed by various people, including one by the author[61]. However, as of this writing, the author is not aware of any that have been accepted as a standard tool that is publicly available in the HEP community.

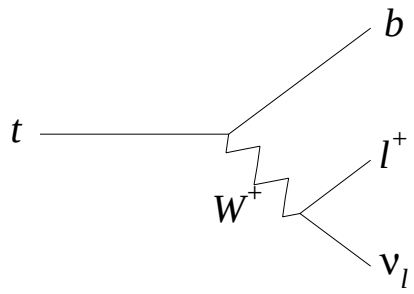


Figure 7.4: The Feynman diagram for the semileptonic decay of the top quark.

See appendix F for further discussion of this issue.

7.3.2 The Distribution of the Four Parameters Assuming the Independent Decay Model

Having settled on the assumption of independent decay, one can obtain the distribution of the four parameters using the Feynman diagram in figure 7.4 and the assumptions that: (1) the t and $\bar{t}$ spin are randomly oriented (*i.e.* they are unpolarized); and (2) the t decay is not influenced by, and does not influence, the $\bar{t}$ decay. In what follows, the first assumption will be relaxed to allow the possibility of top quark spin polarization — *i.e.* the possibility for the t and $\bar{t}$ spin to point preferentially along some fixed direction. Allowing the possibility of spin polarization enables one to quantify the degree to which distributions in the experimental data agree with the standard model prediction.

The details of the matrix element calculations are shown in appendix H. The result is the following:

$$F(\cos\theta_\ell, \cos\psi_\ell^*) = f(\cos\theta_\ell) \cdot g(\cos\psi_\ell^*) \tag{7.6}$$

In this context, $\cos\theta_\ell$ is the angle that $\vec{p}_\ell$ makes with the top quark spin polarization vector,

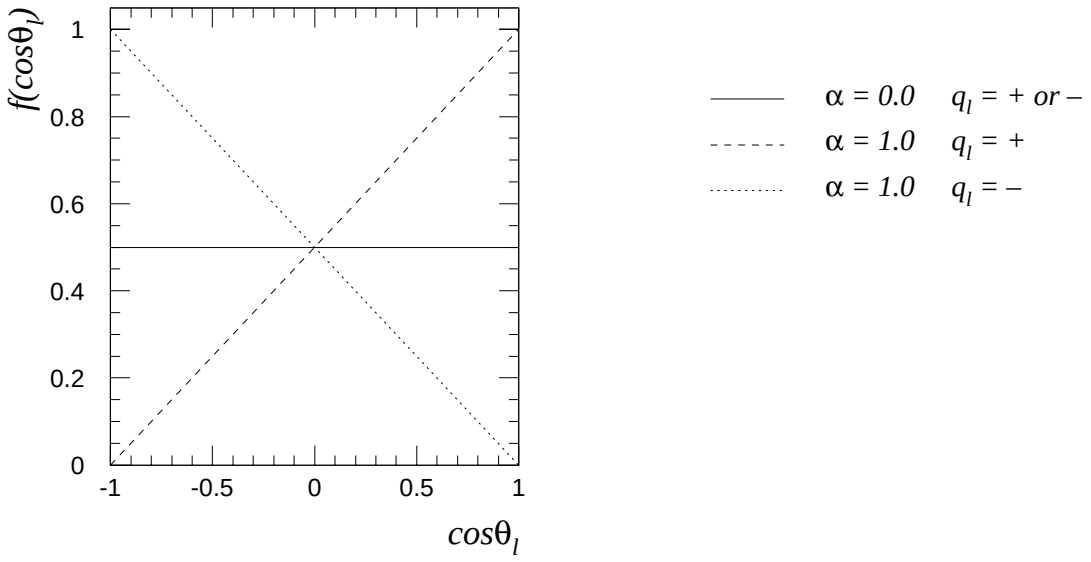


Figure 7.5: The distribution of $\cos\theta_\ell$ for unpolarized ($\alpha = 0$) and 100% polarized ($\alpha = 1.0$) top quark.

and $\cos\psi^*_\ell$ was described earlier in section 7.2. Let us examine the functions $f(\cos\theta_\ell)$ and $g(\cos\psi^*_\ell)$ in detail.

7.3.2.1 The Distribution of $\cos\theta_\ell$

The formula for $f(\cos\theta_\ell)$ is:

$$f(\cos\theta_\ell) = \frac{1}{2}(1 + q_\ell\,\alpha \cdot \cos\theta_\ell) \tag{7.7}$$

The symbol q_ℓ is the sign of the charge of ℓ, and α is the degree of spin polarization ($0 \leq \alpha \leq 1$). According to the standard model and the independent decay assumption, $\alpha = 0$ — *i.e.* there is no dependence on $\cos\theta_\ell$, and, therefore, $\vec{p}_\ell$ is randomly oriented. If $\alpha \neq 0$, then the distribution of $\vec{p}_\ell$ is asymmetric: ℓ^+ tends to decay toward the spin polarization vector, and ℓ^- tends to decay away from it. Figure 7.5 shows a plot of $f(\cos\theta_\ell)$. In chapter 10 where the experimental data are compared with the standard model prediction, the parameter α is used to measure the degree of consistency of the data with the standard model.

7.3.2.2 The Distribution of $\cos\psi^*_\ell$

The formula for $g(\cos\psi^*_\ell)$ is:

$$
\begin{aligned}
g(\cos\psi_\ell^*) &= (h_-,\, h_0,\, h_+)\cdot(g_b,\, g_n,\, g_f) \\
&= h_-\cdot g_b(\cos\psi_\ell^*) + h_0\cdot g_n(\cos\psi_\ell^*) + h_+\cdot g_f(\cos\psi_\ell^*) \qquad (7.8)
\end{aligned}
$$

The first line in the above equation is a short-hand notation indicating that $g(\cos\psi_\ell^*)$ is obtained by taking the dot product of a triplet of scalars $(h_-,\, h_0,\, h_+)$ with a triplet of functions $(g_b,\, g_n,\, g_f)$[4]. The three components are for the three helicity states of W: left-handed, longitudinal, and right-handed. For the decay $t \to b + W^+$, the first, second, and third component of the triplet correspond to the left-handed, longitudinal, and right-handed helicity state; for the charge conjugate process $\overline{t} \to \overline{b} + W^-$, the components correspond to the right-handed, longitudinal, and left-handed helicity state. The assignment of the helicity states to each component of the triplet is summarized below:

$t \to b + W^+$	(left, long, right)
$\overline{t} \to \overline{b} + W^-$	(right, long, left)

The components of $(h_-,\, h_0,\, h_+)$ are scalar quantities that depend on m_t, M_W, and m_b, and they are called the *helicity fractions*. Since $m_b \ll m_t$ and M_W, the b-quark mass can, to a good approximation, be set to zero — doing so introduces to h_i an error on the order of $m_b^2/M_W^2 \approx 0.004$. In this approximation, the standard model prediction for the helicity fractions are:

$$h_- = \frac{1}{1 + x^2/2} \qquad (7.9)$$

$$h_0 = \frac{x^2/2}{1 + x^2/2} \qquad (7.10)$$

$$h_+ = 0 \qquad (7.11)$$

[4]These 3-component objects are obviously not vectors. They are written in vector-like notation as a suggestive, short-hand notation

$$x = \frac{m_t}{M_W} \tag{7.12}$$

For $m_t = 175\ GeV$, $h_- = 0.30$, $h_0 = 0.70$, and $h_+ = 0$. The fact that $h_+ \approx 0$ — *i.e.* t ($\overline{t}$) almost decouples from the right-handed (left-handed) helicity state — is a consequence of the $V - A$ coupling in the t-W-b vertex. The fact that h_+ is not exactly zero is due to the small helicity-flip amplitude resulting from the non-zero mass of the b-quark.

In the decay $t \to b + W^+$, h_-, h_0, and h_+ can be thought of as coupling strengths of t to the left-handed, longitudinal, and right-handed helicity states of W. In the charge-conjugate case $\overline{t} \to \overline{b} + W^-$, h_-, h_0, and h_+ are coupling strengths of $\overline{t}$ to the right-handed, longitudinal, and left-handed states. The helicity fractions, by definition, are normalized, so the following relation holds: $h_- + h_0 + h_+ = 1$. In addition, h_i are non-negative. These properties of h_i suggest a probabilistic interpretation — *i.e.* that h_i gives the probability (branching fraction) that a top quark decay produces W in the helicity state i. This, however, is incorrect because the transverse helicity states of W interfere with the longitudinal state, and, therefore, a top quark decay does not produce W in a definite helicity state. Appendix J discusses this issue in detail.

The functions $g_b(\cos\psi_\ell^*)$, $g_n(\cos\psi_\ell^*)$, and $g_f(\cos\psi_\ell^*)$ have the following form:

$$g_b(\cos\psi_\ell^*) = \frac{3}{8}(1 - \cos\psi_\ell^*)^2 \tag{7.13}$$

$$g_n(\cos\psi_\ell^*) = \frac{3}{4}(1 - \cos^2\psi_\ell^*) \tag{7.14}$$

$$g_f(\cos\psi_\ell^*) = \frac{3}{8}(1 + \cos\psi_\ell^*)^2 \tag{7.15}$$

These functions give the $\cos\psi_\ell^*$ distribution for the three helicity states of W. The subscripts b, n, and f stand for *backward*, *normal*, and *forward*, and they indicate the direction in which g_i peaks (see figure 7.6). In the decay $t \to b + W^+$, the left-handed, longitudinal, and right-handed helicity state of W give rise to $\cos\psi_\ell^*$ distribution given by g_b, g_n, and g_f, respectively. In the charge-conjugate process $\overline{t} \to \overline{b} + W^-$, the left-handed, longitudinal, and right-handed helicity state of W have the $\cos\psi_\ell^*$ distributions g_f, g_n, and g_b, respectively.

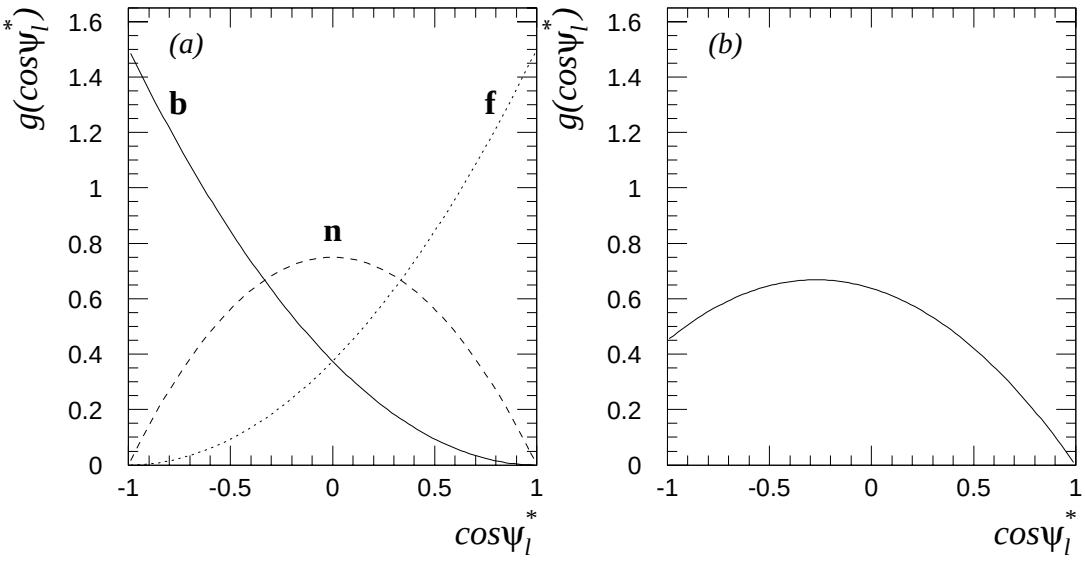


Figure 7.6: (a) The functions $g_i(\cos\psi_\ell^*)$, $i = b, n, f$. (b) Standard model distribution of $\cos\psi_\ell^*$ assuming $m_t = 175\ GeV$. The curve is obtained by the combination 0.30*(backward) + 0.70*(normal).

	h_i			$g_i(\cos\psi_\ell^*)$		
	LEFT	LONG	RIGHT	LEFT	LONG	RIGHT
$t \to b + W^+$	r_T	$1 - r_T$	0	g_b	g_n	g_f
$\overline{t} \to \overline{b} + W^-$	0	$1 - r_T$	r_T	g_f	g_n	g_b

Table 7.1: The helicity fractions h_i and the $\cos\psi_\ell^*$ distributions $g_i(\cos\psi_\ell^*)$ for the three helicity states of W. The role of left and right are reversed for t and $\overline{t}$. The quantity r_T is equal to $1/(1 + x^2/2)$, where $x = m_t/M_W$.

A summary of the results on the distribution of $\cos\psi_\ell^*$ is given in table 7.1. An interesting feature about these results is the fact that the roles of left and right are reversed for t and $\overline{t}$. This role-reversal occurs at the t-b-W vertex (described by h_i) *and* at the W-ℓ-ν_ℓ vertex (described by g_i). Because of this double-reversal, the final distribution of $\cos\psi_\ell^*$ is identical for t and $\overline{t}$. The physical basis for this behavior is discussed in detail in section E.7 of appendix E.

7.3.2.3 Summarizing the Discussion on $f(\cos\theta_\ell)$ and $g(\cos\psi^*_\ell)$

Let us summarize the results obtained so far. According to the standard model, the four parameters that describe the top quark decay in the top rest frame are distributed according to $F(\cos\theta_\ell, \cos\psi^*_\ell)$. If the top quark spin is unpolarized, then the $\cos\theta_\ell$ dependence is absent. Therefore, the standard model together with the independent decay assumption implies the following about the distribution of the four parameters:

> Of the four parameters describing the top quark decay in the top rest frame, the three required to orient the plane containing $\vec{p}_\ell$, $\vec{p}_{\nu_\ell}$, and $\vec{p}_b$ are randomly distributed. The one parameter that fixes the magnitude and direction of $\vec{p}_\ell$, $\vec{p}_{\nu_\ell}$, and $\vec{p}_b$ in the decay plane is distributed according to equation 7.8.

7.3.2.4 A Change of Variable: $\cos\psi^*_\ell \rightarrow E_\ell$

Before leaving this section, a change of variable will be introduced for aesthetic reasons. Equation 7.1 shows that $\cos\psi^*_\ell$ is related to E_ℓ, the energy of ℓ in the top rest frame, by a linear transformation. Therefore the function $F(\cos\theta_\ell, \cos\psi^*_\ell)$ could equally well have been written as $F(\cos\theta_\ell, E_\ell)$; no information is gained or lost by making this change in variable. By making this change, the standard model prediction regarding the top quark decay in the top rest frame can be expressed solely in terms of the direction and magnitude of $\vec{p}_\ell$:

$$F(\cos\theta_\ell, E_\ell) = f(\cos\theta_\ell) \cdot g(E_\ell) \tag{7.16}$$

The function $f(\cos\theta_\ell)$ is the same as before, while $g(E_\ell)$ is:

$$\begin{aligned} g(E_\ell) &= (h_-,\, h_0,\, h_+) \cdot (g_s,\, g_m,\, g_h) \\ &= h_- \cdot g_s(E_\ell) + h_0 \cdot g_m(E_\ell) + h_+ \cdot g_h(E_\ell) \end{aligned} \tag{7.17}$$

The helicity fractions h_-, h_+, and h_0 have the same values as before, while the functions $g_s(E_\ell)$, $g_m(E_\ell)$, and $g_h(E_\ell)$ are obtained from $g_b(\cos\psi^*_\ell)$, $g_n(\cos\psi^*_\ell)$, and $g_f(\cos\psi^*_\ell)$, respectively, by the change of variable $\cos\psi^*_\ell \rightarrow E_\ell$. Figure 7.7 shows plots of the E_ℓ distributions. The E_ℓ distributions are related to their $\cos\psi^*_\ell$ counterparts by a translation and a change

of scale. The subscripts s, m, and h stand for *soft*, *medium*, and *hard*, and refer to the hardness of the E_ℓ distributions. The hardness of the E_ℓ distributions are due to the nature of Lorentz transformations (see equation 7.1). The functional form of $g_i(E_\ell)$ is given below:

$$g_s(E_\ell) = \frac{3}{P_W}\left(\frac{E_\ell^{max} - E_\ell}{E_\ell^{max} - E_\ell^{min}}\right)^2 \tag{7.18}$$

$$g_m(E_\ell) = \frac{6}{P_W} \times \frac{(E_\ell^{max} - E_\ell)(E_\ell - E_\ell^{min})}{(E_\ell^{max} - E_\ell^{min})^2} \tag{7.19}$$

$$g_h(E_\ell) = \frac{3}{P_W}\left(\frac{E_\ell - E_\ell^{min}}{E_\ell^{max} - E_\ell^{min}}\right)^2 \tag{7.20}$$

$$E_\ell^{max} = \frac{1}{2}m_t \tag{7.21}$$

$$E_\ell^{min} = \frac{1}{2}\left(\frac{{M_W}^2}{m_t}\right) \tag{7.22}$$

$$P_W = E_\ell^{max} - E_\ell^{min} \tag{7.23}$$

The quantities E_ℓ^{max} and E_ℓ^{min} are the upper and lower bounds of E_ℓ — outside of these bounds, $g_i(E_\ell) \equiv 0$. The quantity P_W is the W momentum in the top rest frame (this is a constant fixed by m_t, M_W, and m_b).

7.4 The Observables

The results of sections 7.2 and 7.3 lay the theoretical foundation for the study of the top quark decay kinematics in the top rest frame. In this section, observables that will allow one to compare distributions in the experimental data with the theoretical prediction will be defined. In the first part of this section, the observables will be introduced. In the second part, the distribution of the observables from monte carlo models will be shown. The effect of: (1) event selection cuts; and (2) smearing introduced by event reconstruction; on the observable distributions will be examined.

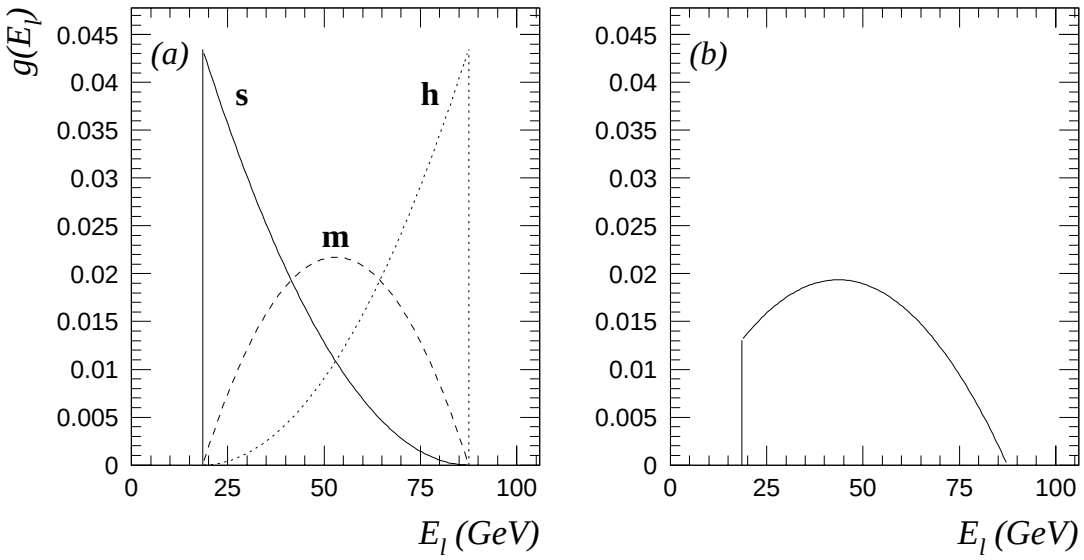


Figure 7.7: (a) $g_i(E_\ell)$, $i = s, m, h$. (b) Standard model distribution of E_ℓ assuming $m_t = 175\ GeV$. The distributions have the same shape as those shown in figure 7.6; they are related to each other by a translation and a scale change.

7.4.1 The Definition of the Observables

It was shown in section 7.3 that the four parameters describing the top quark decay in the top rest frame are distributed according to the $F(\cos\theta_\ell, E_\ell)$ given in equation 7.16. This equation treats the general situation where the top quark is allowed to have non-zero spin polarization. In the independent decay model of the top quark, however, the top quark is assumed to have zero spin polarization, and, therefore, the $\cos\theta_\ell$ dependence drops out of $F(\cos\theta_\ell, E_\ell)$. In other words, of the four parameters that describe top quark decay, three are randomly distributed, while one parameter — E_ℓ — is distributed according to $g(E_\ell)$ in equation 7.17. The three randomly distributed parameters are those required to orient the plane containing $\vec{p}_\ell$, $\vec{p}_{\nu_\ell}$, and $\vec{p}_b$ (the decay plane); the one parameter distributed in a non-random way is that which fixes the direction and magnitude of $\vec{p}_\ell$, $\vec{p}_{\nu_\ell}$, and $\vec{p}_b$ in the decay plane. The following is a summary of these observations:

- The direction of $\vec{p}_\ell$ is randomly distributed. In other words, $\cos\theta_\ell$ and ϕ_ℓ are randomly distributed, where θ_ℓ and ϕ_ℓ are the polar and azimuthal angle of $\vec{p}_\ell$ defined in some coordinate system.

- The magnitude of $\vec{p}_\ell$ is distributed according to $g(E_\ell)$ in equation 7.17.

- The angle $\phi_{b\text{-}\nu}$ is randomly distributed (see figure 7.1).

Of these four observables, $\cos\theta_\ell$, ϕ_ℓ, and E_ℓ will be studied in this thesis. The choice not to study $\phi_{b\text{-}\nu}$ is due to time constraints.

In what follows, prescriptions for obtaining the observables from the experimental data are given.

7.4.1.1 Obtaining E_ℓ from the Experimental Data

The primary lepton energy in the top rest frame, E_ℓ, is obtained by boosting the lab frame charged lepton 4-momentum to the top rest frame. In order to do this, one must first obtain the lab frame 4-momentum of the semileptonically decaying top quark, t_ℓ. This is obtained using the $t\bar{t}$ reconstruction algorithm, which is described in chapter 6.

7.4.1.2 Obtaining the Top Rest Frame Angular Distribution of $\vec{p}_\ell$ from the Experimental Data

Obtaining the top rest frame angular distribution of $\vec{p}_\ell$ from the experimental data is more involved than was the case for E_ℓ. The first step is the same as for E_ℓ: one uses the lab frame momentum of t_ℓ to boost the momentum of ℓ to the t_ℓ rest frame. In the t_ℓ rest frame, however, a coordinate system must be defined. In this thesis, it is defined as follows:

1. Using the lab frame momentum of the semileptonically and hadronically decaying top quark, t_ℓ and t_h, form the momentum of the $t\bar{t}$ system. This momentum is almost completely along the beam line.

2. Using the $t\bar{t}$ momentum, boost decay product momenta to the $t\bar{t}$ rest frame.

3. In the $t\bar{t}$ rest frame, define three unit vectors $\hat{u}_1$, $\hat{u}_2$, and $\hat{u}_3$, where $\hat{u}_1$ is along the t_ℓ momentum vector, $\hat{u}_2$ is along the cross product of $\hat{u}_1$ and the beam line, and $\hat{u}_3 = \hat{u}_1 \times \hat{u}_2$. See figure 7.8.

4. Using the coordinate system defined by $\hat{u}_1$, $\hat{u}_2$, and $\hat{u}_3$, boost decay product momenta to the t_ℓ rest frame.

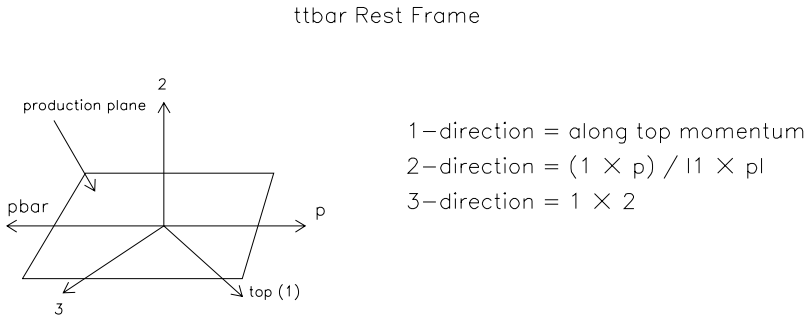


Figure 7.8: The unit vectors that define a coordinate system in the top rest frame. The vectors are defined in the $t\bar{t}$ rest frame, and are used to boost the decay product momenta to the top rest frame. In the top rest frame, the same unit vectors are used to define the coordinate system.

5. In the t_ℓ rest frame, use the same set of unit vectors to define a coordinate system.

The coordinate system is defined in this manner because if anomalous spin polarization were present, the 1- and 2-direction are good candidate directions for the polarization vector. For instance, polarization along or against the 1-direction may imply chiral anomaly in the production mechanism of $t\bar{t}$, and polarization along or against the 2-direction may imply anomalous loop contribution to the top quark decay amplitude. Reference [30] and [31] discuss these possibilities. The 3-direction is not a candidate direction for any anomalous polarization, but is included for completeness.

Once the coordinate system is defined, one can define the angle θ_ℓ and ϕ_ℓ as the polar and azimuthal angle of $\vec{p}_\ell$ in the top rest frame, taking the 1-direction as the polar direction and the 2-3 plane as the azimuthal plane. These angles, however, will not be used as the observables. Instead, the direction cosines of $\vec{p}_\ell$ will be chosen. They are defined as follows:

$$\cos U_i = \frac{\vec{p}_\ell \cdot \hat{u}_i}{|\vec{p}_\ell|}, \qquad i = 1, 2, 3 \tag{7.24}$$

This equation defines three observables where only two exist, so $\cos U_i$ are not independent of one another. However, this choice is convenient because, if anomalous polarization were present, the signature will show up as an asymmetry in the distribution of $\cos U_i$.

7.4.2 The Distribution of the Observables

In the monte carlo simulation of $p\bar{p} \rightarrow t\bar{t}$, one can view the simulation process as going through the following stages:

$$\text{Analytic} \rightarrow \text{Parton-level} \rightarrow \text{Reconstructed}$$

The stages are defined as follows:

Analytic

In this stage, no cuts are applied to the generated events. The observable distributions in this stage are shown in figures 7.5 and 7.7 — they are either linear or parabolic.

Parton-level

The events in this stage are a subset of those in the analytic stage. The subset is defined as follows. First, one takes all events in the analytic stage and perform a full simulation — *i.e.* take account of gluon radiation, convert outgoing quarks and gluons to jets, perform detector simulation, *etc.* Second, event selection cuts are applied to the physics objects in the fully simulated events. The events that pass this cut belong to this stage. The 4-vectors of the physics objects in this stage, however, are not those of the fully reconstructed physics objects, but of the analytic-level. The observable distributions in this stage have similar shapes as those in the analytic level, but they are modified by the event selection cuts.

Reconstructed

The events in this stage are exactly the same as those in the parton-level. However, the 4-vectors of the physics objects in this stage are those obtained after full simulation. These 4-vectors are, in general, different from those in the parton-level because of uncertainties introduced by the $t\bar{t}$ reconstruction process. This change in 4-vectors introduces *smearing* in the observable distributions.

The change in shape of the observables in going from the analytic → parton-level stage will be referred to as being due to the *acceptance effect*. Similarly, the change in shape of the observable distributions in going from the parton-level → reconstructed stage will

be referred to as being due to the *smearing effect*. To convert observable distributions in the analytic stage to those in the parton-level stage, one takes the analytic distributions and multiply them by *acceptance curves*. To convert the distributions in the parton-level to those in the reconstructed stage, one takes the parton-level distributions and applies the *smearing matrices*. These conversion methods are discussed below. The validity of these methods hinge on an important property of the observable distributions in the standard model. What this property is, and why it is necessary for the methods to be valid, are discussed in appendix K.

7.4.2.1 The Acceptance Effect

The overall acceptance of $p\bar{p} \to t\bar{t}$ in the lepton + jets decay channel is approximately 10%. In this context, "acceptance" is the ratio of the number of monte carlo-generated events passing the event selection cuts to the total number of events generated (*i.e.* the number of events in the parton-level stage divided by the number in the analytic stage). Figure 7.9 shows the acceptance as a function of the true observable value.

The shape of each distribution is determined mostly by the primary charged lepton identification cuts (see chapter 3). The following is a detailed discussion of the shape of each distribution:

$\cos U_1$

Two noteworthy features of this distribution are: (1) the dip near $\cos U_1 = -1$; and (2) the asymmetry about $\cos U_1 = 0$. The dip is primarily due to the P_T cut. When ℓ has $\cos U_1 \approx -1$, its momentum in the top rest frame points against the boost direction from the $t\bar{t}$ rest frame to the top rest frame. Since the momentum of ℓ opposes the boost, the energy of ℓ (and, hence, P_T of ℓ) in the $t\bar{t}$ rest frame is smaller than it would be for larger values of $\cos U_1$. The P_T of ℓ in the $t\bar{t}$ rest frame is usually about the same as it is in the lab frame. Thus $\cos U_1 \approx -1$ implies small P_T in the lab frame, and hence these events are likely to fail the P_T cut.

The asymmetry in the distribution is due to the fact that the '1' direction is along the boost direction from the $t\bar{t}$ rest frame to the top rest frame. As illustrated above, P_T

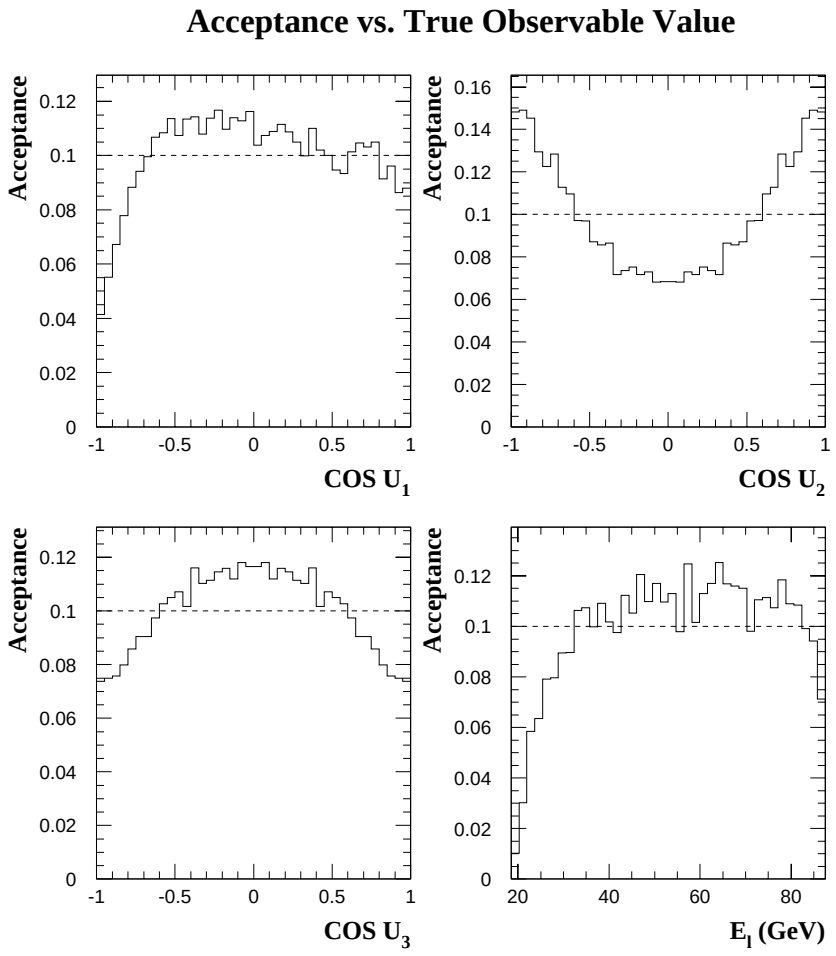


Figure 7.9: Acceptance versus true observable value for the observables $\cos U_1$, $\cos U_2$, $\cos U_3$, and E_ℓ. The dashed line in each graph shows the average acceptance. The $\cos U_2$ and $\cos U_3$ distributions have been symmetrized.

of ℓ in the lab frame tends to be larger when ℓ travels along the '1' direction than when it travels against it. Thus the acceptance curve is forward-backward asymmetric.

$\cos U_2$

The noteworthy features of this distribution are: (1) the dip around $\cos U_2 = 0$; and (2) the symmetry about $\cos U_2 = 0$. The dip is caused primarily by the η cut. By definition, the 2-direction is perpendicular to the beam line. Therefore the direction along or against $\hat{u}_2$ — *i.e.* $|\cos U_2| \approx 1$ — tends to point into the detector, where the acceptance is greatest. Thus the peaks of the distribution occur at $\cos U_2 = \pm 1$. The symmetry of the distribution reflects the azimuthal symmetry of both the detector and the $t\bar{t}$ production process.

$\cos U_3$

The noteworthy features of this distribution are: (1) the peak around $\cos U_3 = 0$; and (2) the symmetry about $\cos U_3 = 0$. The first feature is due to the fact that the 3-direction tends to point toward the beam line. Thus when $\cos U_3 \approx \pm 1$, $\hat{u}_3$ tends to point out of the detector, so the acceptance curve in those regions has minimum value. The symmetry of the distribution is due to the symmetry along the beam line both of the detector and the $t\bar{t}$ production process.

E_ℓ

The single noteworthy feature of this distribution is the dip at small E_ℓ. This dip is primarily due to the P_T cut. Events with little energy in the top rest frame tends to have little energy in the lab frame; such events are more likely to fail the P_T cut than events with larger E_ℓ.

The parton-level distribution of the observables is obtained by modulating the analytic distributions in figures 7.5 and 7.7 by the acceptance curves. This procedure is illustrated in figure 7.10 for the distribution of $\cos U_1$ for a top quark that is 100% polarized *against* the 1-direction. The contents of the histogram for the analytic distribution is multiplied bin-by-bin by the efficiency curve to obtain the parton-level distribution.

Figures 7.11 and 7.12 show the parton-level distribution of the observables $\cos U_i$, $i = 1, 2, 3$ for unpolarized and 100% polarized top quark spin. For $\cos U_1$, the situation

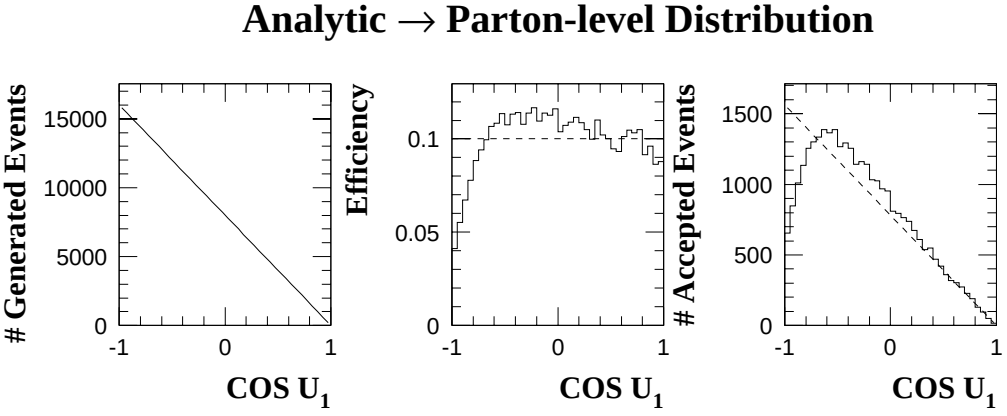


Figure 7.10: Obtaining the parton-level distribution from the analytic distribution and the efficiency curve. The parton-level distribution (right) is obtained by multiplying bin-by-bin the histogram for the analytic distribution (left) by the efficiency (center). The dashed line in the figure on the right is the parton-level distribution if the efficiency were flat. The dashed line in the middle figure is the average acceptance.

where the polarization vector is along and against the 1-direction are both shown because the acceptance curve is asymmetric about $\cos U_1 = 0$. For $\cos U_2$ and $\cos U_3$, the acceptance curves are symmetric, so only polarization along the respective coordinates is shown. Figure 7.13 shows, for $t \to b + \ell^+ + \nu_\ell$, the parton-level distribution of E_ℓ for W^+ 100% in the left-handed, longitudinal, and right-handed helicity state, respectively. Also shown is the standard model prediction.

Before leaving this section, a comment on the total number of accepted events is in order. First, let us consider the parton-level E_ℓ distribution. To make the argument concrete, let us specify the charge state of the top quark as follows: $t \to b + \ell^+ + \nu_\ell$. Note in figure 7.9 the acceptance curve for E_ℓ: it has small acceptance at small E_ℓ, and as E_ℓ gets larger, the acceptance quickly increases and levels off. The analytic distribution for the left-handed helicity state peaks where the acceptance dips to very small values, whereas those for the longitudinal and right-handed helicity states peak where the acceptance is level (see figure 7.7). These observations explain the following fact:

> Before event selection cuts, the standard model distribution of E_ℓ is obtained by combining the left-handed, longitudinal, and right-handed distributions in the proportion 30% : 70% : 0% (taking $m_t = 175\ GeV$). After event selection cuts, the proportion becomes 25% : 75% : 0%.

The exact amount of the shift in the value of the helicity fractions is determined by the ratio

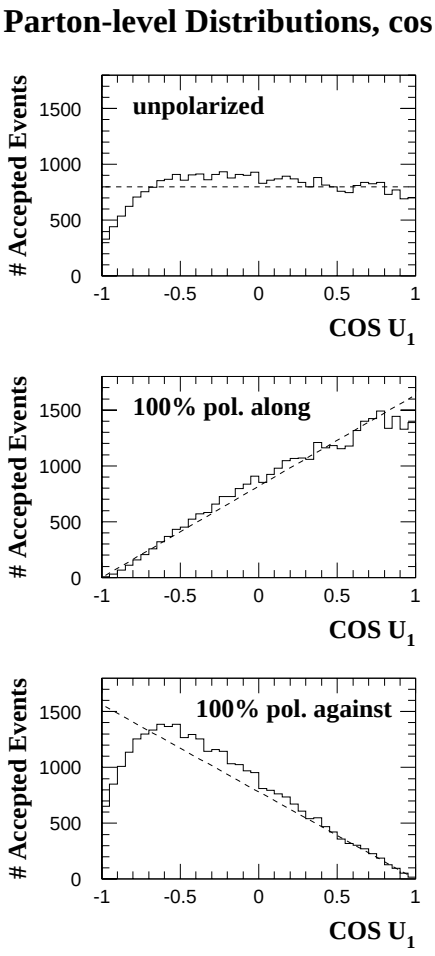


Figure 7.11: Parton-level distribution of $\cos U_1$ for unpolarized (top), 100% polarized along $\hat{u}_1$ (center), and 100% polarized against $\hat{u}_1$ (bottom). The dashed line in each plot shows what the distributions would look like if the acceptance curve were flat.

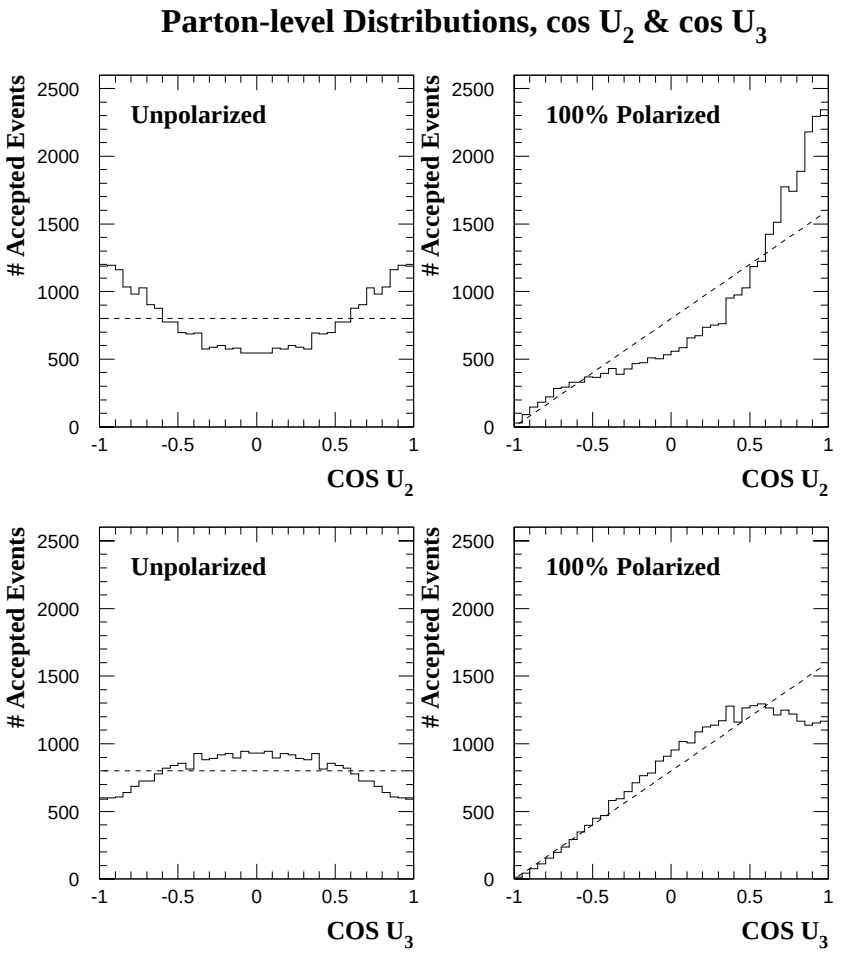


Figure 7.12: Parton-level distribution of $\cos U_2$ and $\cos U_3$ for unpolarized (left column), and 100% polarized along $\hat{u}_2$ or $\hat{u}_3$ (right column). The dashed line in each plot shows what the distributions would look like if the acceptance curve were flat.

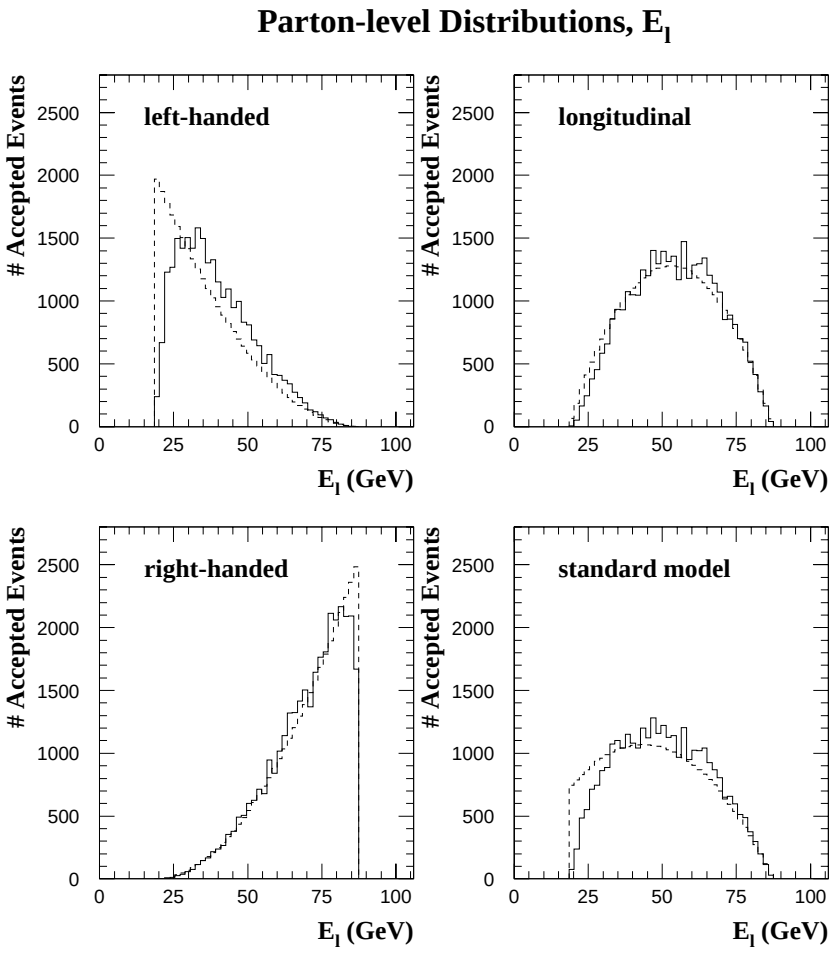


Figure 7.13: Parton-level distribution, in $t \to b + \ell^+ + \nu_\ell$, of E_ℓ for 100% left-handed (upper left), 100% longitudinal (upper right), 100% right-handed (lower left), and standard model prediction (lower right). The dashed line in each plot shows what the distributions would look like if the acceptance curve were flat.

of the acceptances of the left-handed and longitudinal W. This ratio is approximately 0.80. This change in composition of the W helicity states is an effect that must be accounted for when the helicity fractions are measured in the experimental data.

For the angular observables $\cos U_i$, the issue at hand is whether 100% polarized events are more or less likely to be accepted than unpolarized events. It can be shown that there is no difference in acceptance for polarized and unpolarized events if and only if the acceptance curve is symmetric about $\cos U_i = 0$. Thus the 2- and 3-directions are bias-free. On the other hand, the acceptance curve for the 1-direction is asymmetric, so a bias exists. Events originating from top quark 100% spin polarized *along* the 1-direction are 2.4% *more* likely to be accepted compared to events originating from unpolarized top quark. For 100% spin polarization *against* the 1-direction, the percentage is -2.4%.

7.4.2.2 The Smearing Effect

The parton-level distribution of the observables shown in the last section correspond to what one would observe if the $t\bar{t}$ event reconstruction process introduced no error in the measurement of the decay product momenta. In reality, significant amount of error is introduced for a variety of reasons. It will be shown in chapter 9 that the most important source of error is the matching of the wrong jet to the b-quark. This, in turn, is mostly due to the imprecise measurement of the quark momenta, as deduced from the jet energy.

The smeared observable distribution can be obtained from the parton-level distributions by using the *smearing* or *input-output* matrix. An illustration of this method is shown below:

$$\underbrace{\begin{pmatrix} \\ \\ \\ \end{pmatrix}}_{\text{output}} = \begin{pmatrix} \text{Input-Output} \\ \text{Matrix} \\ n \times n \end{pmatrix} \underbrace{\begin{pmatrix} \\ \\ \\ \end{pmatrix}}_{\text{parton}} \tag{7.25}$$

In this equation, the column vectors represent histograms with n bins. The input-output matrix is $n \times n$, and can be thought of as a 2-dimensional histogram. The columns of the matrix represent narrow bands of true (input) observable values, while the rows represent

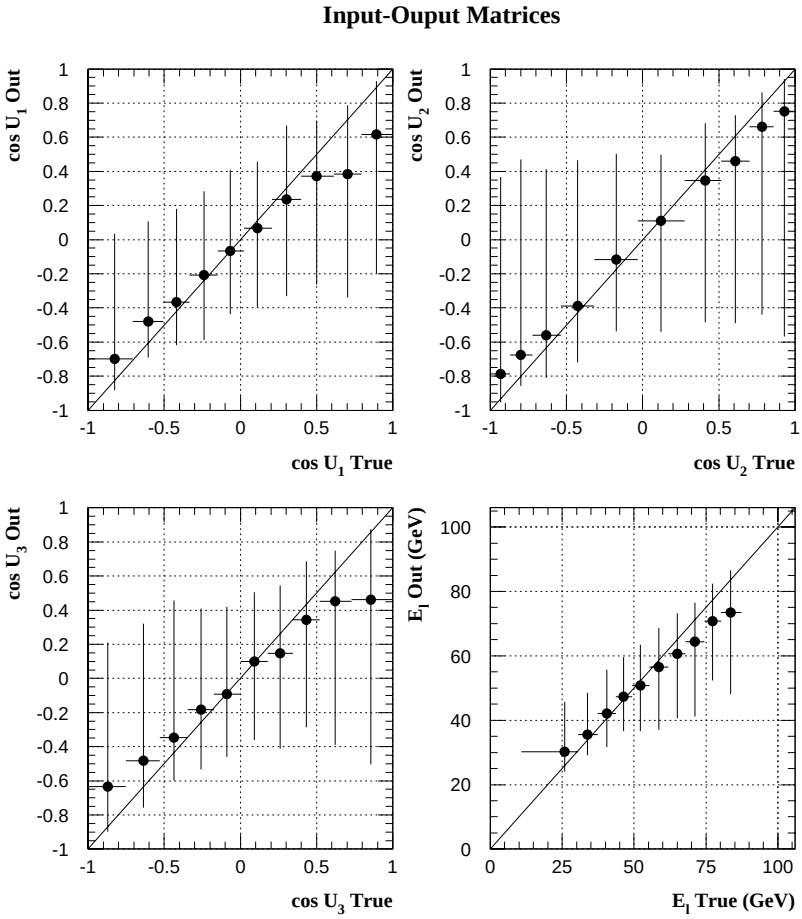


Figure 7.14: Schematic representation of the input-output matrices. The points show the median output value for each band of input (true) observable values. The vertical error bars represent the spread of events containing 34% of the population above and below the median value. The horizontal error bar is the bin size. The bin sizes were chosen so that each bin has the same population.

narrow bands of the smeared (output) observable values. Each column can be thought of as a histogram of the output values for a given narrow range of input values. Figure 7.14 shows a schematic representation of the input-output matrix for all of the observables.

Figures 7.15 through 7.17 show the smeared distributions for all of the observables. The smearing makes the polarized and unpolarized $\cos U_i$ distributions less distinguishable from each other than they are at the parton-level. Similarly, the left-handed, longitudinal, and right handed distributions become less distinguishable from each other after smearing. This loss in distinction between the distributions results in degraded measurement resolution. This degradation can be quantified using the method described in chapter 8.

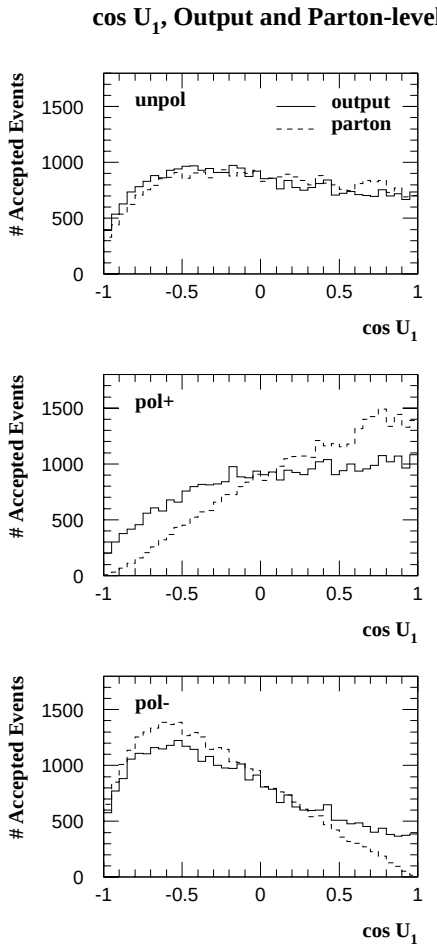


Figure 7.15: The distribution of $\cos U_1$ before and after smearing.

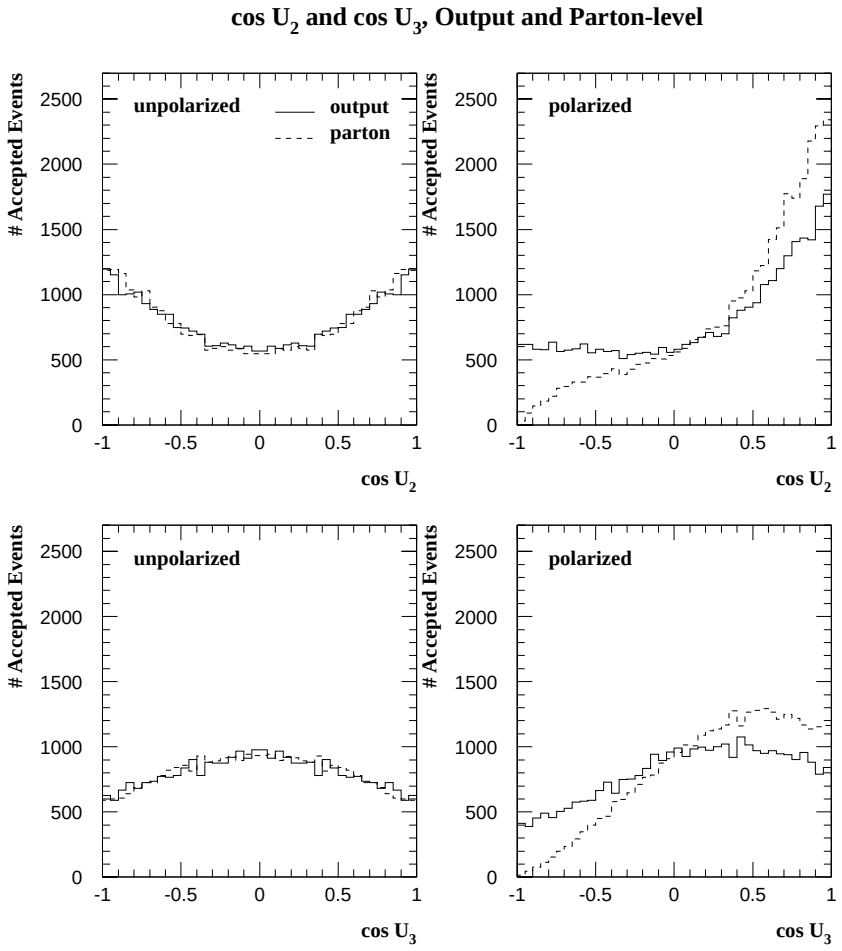


Figure 7.16: The distribution of $\cos U_2$ and $\cos U_3$ before and after smearing.

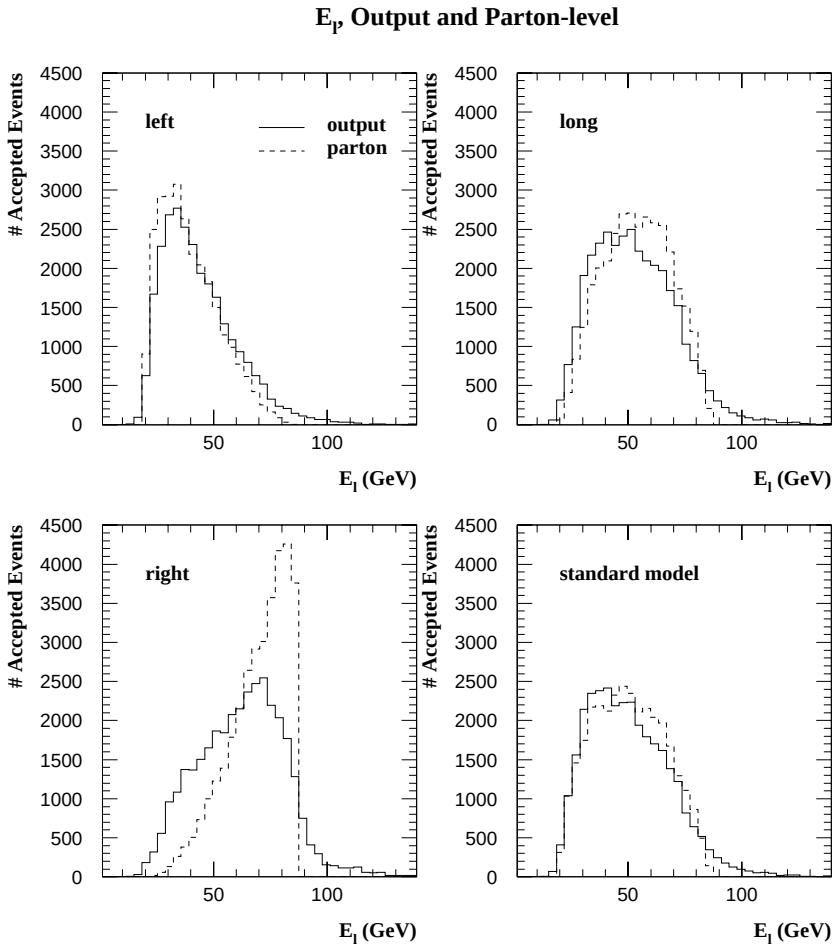


Figure 7.17: The distribution of E_ℓ before and after smearing. The distributions assume the following charge state: $t \rightarrow b + \ell^+ + \nu_\ell$.

7.5 Answers to Questions Raised Earlier

At the beginning of this chapter, three questions regarding the choice of observables were raised:

- Why examine the decay product kinematics in just the semileptonic decay channel of the top quark, and not in the hadronic channel?
- Why examine just the charged lepton energy and angular distributions, and not those of the neutrino or the b-quark?
- Why examine the top quark decay product kinematics in the top rest frame rather than in the lab frame?

Equipped with the information presented in the last several section, these questions can now be answered.

7.5.1 Why Examine Only $t \to \ell + \nu_\ell + b$?

It was mentioned in chapter 4 that it is useful, from an experimental point of view, to classify the possible decay channels of $t\bar{t}$ events into the following categories: the dilepton, lepton + jets, and the all-hadronic. It was argued there that the study of the decay kinematics of the top quark in the top rest frame can be carried out reliably only in the lepton + jets channel. It is for this reason that, in this thesis, one examines events in which one of t or $\bar{t}$ decays semileptonically, and the other decays hadronically. The semileptonically decaying top quark is referred to as t_ℓ, and the hadronically decaying one is called t_h.

Since there are two top quarks in each event, it may seem strange that, in this thesis, only the decay kinematics of t_ℓ is examined. Could one not double the available statistics by examining the hadronic decay channel?

The answer to this question is: no. In order to explain this answer, let us recall that four parameters are necessary and sufficient to describe the top quark decay in the top rest frame. In the semileptonic decay of the top quark, these parameters are: $\cos\theta_\ell$, ϕ_ℓ, $\phi_{b\text{-}\nu}$, and E_ℓ. The following equation gives the distribution of these parameters:

$$F(\cos\theta_\ell, E_\ell) = f(\cos\theta_\ell)g(E_\ell) \tag{7.26}$$

The function f, and only the function f, provides information about the top quark spin. Similarly, the function g, and only the function g, provides information about the helicity structure of the t-W-b interaction vertex. In other words, the spin property of the top quark and the helicity structure of the t-W-b interaction vertex decouple from one another.

Now let us see how one would go about parameterizing the decay of t_h. The decay of t_h is denoted thus: $t_h \rightarrow b_h + W_d + W_u$. The decay kinematics of b_h is the same as that of b_ℓ, that of W_d is the same as that of ℓ, and that of W_u is the same as that of ν_ℓ. Thus, if one chooses to describe the decay of t_h the same way as in the decay of t_ℓ, the four parameters become the following: $\cos\theta_d$, ϕ_d, $\phi_{b\text{-}u}$, and E_d. The angles θ_d and ϕ_d specify the direction of the momentum of the W_d, $\phi_{b\text{-}u}$ is the analog of $\phi_{b\text{-}\nu}$ where ν_ℓ is changed to W_u, and E_d is the energy of the W_d.

This parameterization is fine if the down-type quark W_d can be identified as reliably as its analog ℓ in the decay of t_ℓ. In reality, this is not the case because W_d is detected as a jet. Since W_d is a light quark, one cannot distinguish between a W_d-jet from a W_u-jet, a b jet without b-tag, or a gluon jet. Worse yet is the fact that even when the correct jet is found, its momentum is usually quite different from the true W_d momentum. Because of these difficulties, the distribution of any observables based on the W_d momentum is badly smeared. The use of such observables is, therefore, highly unsuitable in the study of the top quark decay kinematics.

In order to salvage some information, one may try using b_h to probe the decay properties of t_h. The quark b_h is, in a sense, superior to W_d in probing the decay of t_h because b-quarks can be tagged, and, therefore, the combinatoric background for assigning the correct jet to b_h is smaller. The problem with this method, however, is the fact that, at the parton-level, b_h is a poor probe of the decay properties of the top quark.

To demonstrate this point, let us parameterize the momentum of b_h by the following: θ_b, ϕ_b, and E_b. The energy E_b of b_h gives no information about the top quark spin and the t-W-b vertex because it is constant: $E_b = ({m_t}^2 - {M_W}^2 + {m_b}^2)/(2m_t)$. Thus only θ_b and ϕ_b remain as candidate observables. Let us assume that t_h is spin polarized, the degree of

polarization being α. Let us take θ_b as the angle between the momentum of b_h and the spin polarization vector. Then $\cos\theta_b$ is distributed according to the following formula:

$$f(\cos\theta_b) = \frac{1}{2}\left[1 + K\alpha\left(1 - 2r_T\right)\cdot\cos\theta_b\right] \tag{7.27}$$

The factor $K = -1$ if t_h is a particle, and $+1$ if it is an anti-particle, and $r_T = 1/(1+x^2/2)$, where $x = m_t/M_W$. This formula indicates the following:

- If $\alpha = 0$, then the $\cos\theta_b$ distribution is flat; if $\alpha \neq 0$, it is asymmetric. In this respect, b_h behaves like ℓ.

- The degree of asymmetry of the $\cos\theta_b$ distribution depends on both α and r_T — *i.e.* equation 7.27 contains information on both the top quark spin and the t-W-b vertex. This is unlike the case of ℓ, where the information on these two aspects of the top quark decay are contained in two independent functions. This is problematic for two reasons. First, a measurement of the asymmetry of $\cos\theta_b$ yields a mixture of information about the top quark decay properties. In other words, a measurement provides incomplete information about both the top quark spin and the t-W-b vertex. Second, because equation 7.27 contains a mixture of information about top decay properties, the acceptance curve and smearing matrix for the observable $\cos\theta_b$ depend on the degree of asymmetry in the $\cos\theta_b$ distribution. As discussed in appendix K, this implies that a simple 2-component log-likelihood technique cannot be used to extract the asymmetry parameter from the experimental data.

- For $m_t = 175\ GeV$, and assuming that t_h is 100% spin polarized, the $\cos\theta_b$ distribution is given by the following:

$$f(\cos\theta_b) = \frac{1}{2}\left[1 \pm 0.4\cdot\cos\theta_b\right] \tag{7.28}$$

  This shows that the degree of asymmetry in the $\cos\theta_b$ distribution is only 0.4 times that of the $\cos\theta_\ell$ distribution from a 100% spin polarized t_ℓ. This implies that, even without any smearing, the statistical error on the asymmetry parameter for $\cos\theta_b$ is $1/0.4 = 2.5$ times larger than that for ℓ. In order for the statistical error on the $\cos\theta_b$ distribution to be equal to that of $\cos\theta_\ell$, one would need $2.5^2 = 6.25$ times as

many events. Considering the fact that $\cos\theta_b$ has much worse smearing than $\cos\theta_\ell$, it is clear that the statistical error on the asymmetry parameter of $\cos\theta_b$ after event reconstruction is prohibitively large.

The above observations show that very little information can be salvaged by using b_h instead of W_d as a probe of the decay of t_h. It is for this reason that, in this thesis, the decay kinematics of only t_ℓ is analyzed. The kinematics of t_h is used only in the $t\bar{t}$ reconstruction process and in the boosting of physics object momenta from the lab frame to the $t\bar{t}$ rest frame. After these tasks are done, the decay products of t_h are ignored.

7.5.2 Why Examine Only the Charged Lepton Energy and Angular Distributions?

One of the more important results of this chapter is the fact that, in the semileptonic decay of the top quark, the charged lepton energy and angular distribution can be used to probe the top quark spin structure and the nature of the t-W-b interaction vertex. The decay of t_ℓ, however, gives rise to two other objects: b_ℓ and ν_ℓ. A natural question to ask is why these objects are ignored in the analysis presented in this thesis.

One way to answer this question is to note the fact that four parameters are necessary and sufficient to describe the top quark decay in the top rest frame. Three of the parameters can be chosen as the direction and magnitude of the charged lepton momentum vector ($\cos\theta_\ell$, ϕ_ℓ, E_ℓ). This leaves one parameter $\phi_{b\text{-}\nu}$, which is an angle that describes the orientation of the decay plane when the direction of ℓ is fixed. This shows that, once ℓ is chosen as the probe of top quark decay, there is hardly any room left for the other objects — the only *independent* information b_ℓ and ν_ℓ provide is in the definition of $\phi_{b\text{-}\nu}$. Asides from this, the information provided by the direction and magnitude of the b_ℓ and ν_ℓ momentum is redundant.

The answer given above leads to another question: why should one choose the charged lepton as the probe of the top quark decay? Why not used b_ℓ or ν_ℓ? The answer to this question is this: the charged lepton is an optimal object to use to probe the property of the top quark spin and the t-W-b interaction vertex. It is optimal because the distribution of $\cos\theta_\ell$ and E_ℓ, $F(\cos\theta_\ell, E_\ell)$, is *separable* — *i.e.* the function F can be written as a product

of a function of $\cos\theta_\ell$ and a function of E_ℓ: $F(\cos\theta_\ell, E_\ell) = f(\cos\theta_\ell)\, g(E_\ell)$. Furthermore, information regarding the top quark spin is, and only is, in the function f. Similarly, information regarding the t-W-b interaction vertex is, and only is, in the function g. This separation allows one to make independent measurements of the property of the top quark spin and the t-W-b vertex. Moreover, as discussed in appendix K, without this separation, measurements of the property of the top quark spin and the t-W-b vertex become very complicated. This separation does not occur in the energy and angular distribution of b_ℓ and ν_ℓ. Therefore b_ℓ and ν_ℓ are inferior objects to use in studying the top quark decay.

7.5.3 Why Examine the Top Quark Decay Kinematics in the Top Rest Frame?

The observables that are examined in this thesis ($\cos U_i$ and E_ℓ) are all measured in the rest frame of the semileptonically decaying top quark. In order to boost the 4-vector of ℓ from the lab frame to the t_ℓ rest frame, one needs to know the lab frame 4-vector of t_ℓ. Determining this from the experimental data is a complicated process requiring an elaborate algorithm described in chapter 6. Considering the great effort needed to obtain observables in the t_ℓ rest frame, one may wonder if the effort is worthwhile. Could not the properties of the top quark spin and the t-W-b decay vertex be studied using lab frame observables? If such observables exist, why examine the top quark decay in the top rest frame rather than in the lab frame?

First let us consider the necessary and sufficient condition for the existence of lab frame observables that can be used to probe the property of the top quark spin and the t-W-b vertex:

> Let x_{lab} be a lab frame observable, and x_0 be an observable in the top rest frame. Suppose x_0 probes one of the properties of top quark decay (spin or t-W-b vertex). The observable x_{lab} can be used to study this property if and only if x_{lab} and x_0 are correlated.

Lab frame observables that satisfy the above condition exist for the observables $\cos U_1$, $\cos U_3$, and E_ℓ. For $\cos U_1$ and E_ℓ, this observable can be taken as $P_T(\ell)$; for $\cos U_3$, it can be taken as η_ℓ. See figure 7.18.[5] Lab frame observables correlated with $\cos U_2$ do exist,

[5]Other choices for the lab frame observables obviously exist, *e.g.* the lab frame energy of ℓ instead of

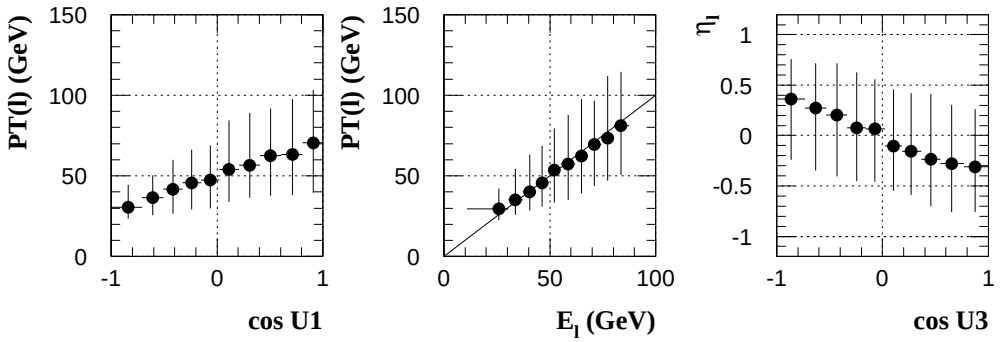


Figure 7.18: The correlation between the parton-level observable distributions in the top rest frame (horizontal axis) and the reconstructed observable distributions in the lab frame (vertical axis). The points are the median of the reconstructed distribution for each parton-level bin. The vertical bars show the spread of 34% of the reconstructed distribution above and below the median. The horizontal bars show the bin size (the bin sizes were chosen so that each bin has equal population).

but the correlation in each case is very weak as to render them practically unusable in measurement.

Figure 7.18 demonstrates the existence of lab frame counterpart of $\cos U_1$, $\cos U_3$, and E_ℓ that can be used to probe the properties of the top quark spin or t-W-b vertex. The next question that naturally arises is this: which observable is "better", lab frame or top rest frame? This question, it turns out, has no simple answer; in some ways the lab frame observables are better, and in others, the top rest frame observables are better. The advantages of using the top rest frame observables are the following:

Simultaneous Measurement of Polarization in the '1'-Direction and t-W-b Vertex

In the lab frame, top spin polarization in the '1'-direction and the t-W-b vertex property are both probed by $P_T(\ell)$. Polarization along (against) the '1'-direction shows up as a harder-than-expected (softer-than-expected) $P_T(\ell)$ distribution (figure 7.19(a)). Similarly, the hardness of the $P_T(\ell)$ distribution is determined by the size of r_T, the left-handed helicity fraction: the larger r_T is, the softer the $P_T(\ell)$ distribution (fig-

$P_T(\ell)$, and $P_z(\ell)$ instead of η_ℓ. No careful study has been done as to which lab frame observable has the greatest amount of correlation with a given top rest frame observable. The author believes, however, that one cannot do a whole lot better than those given in the text.

ure 7.19(b)). Because of this, the hardness/softness of the $P_T(\ell)$ distribution in a given data sample can be due to polarization in the '1'-direction or to the size of r_T. Performing a simultaneous measurement of polarization and r_T from a single $P_T(\ell)$ distribution, however, would require one to examine not only the hardness of the distribution, but its *degree of hardness*. This sort of measurement would require an enormous top quark data sample that is not expected in the foreseeable future. Thus, in practice, a lab frame measurement of '1'-polarization requires one to *assume* that the r_T has expected standard model values. Similarly, in making a lab frame r_T measurement, one must assume that the top quark is unpolarized in the '1'-direction. In contrast to this, separate observables for '1'-polarization and r_T measurements exist in the top rest frame ($\cos U_1$ and E_ℓ). One can, therefore, measure both properties simultaneously.

Possibility of Improving the Measurement Resolution

Let us imagine an hypothetical situation where the t_ℓ decay product momenta can be measured precisely. In this ideal condition, the observable distributions are sharply defined (figures 7.5 and 7.7). In reality, these distributions are smeared out. For the top rest frame observables, the smearing is due to limitations of the $t\bar{t}$ reconstruction process; smearing of the lab frame observables is due to the spread in the top quark momentum distribution. The effect of smearing on measurement resolution can be quantified by the K-factor, which is discussed in detail in section 8.2.2. The size of this factor depends on how distinguishable the component functions are in a 2-component fit: the more distinguishable they are, the smaller the K-factor. Since the measurement error scales as K, the greater the distinguishability between two component functions, the better the measurement resolution. Smearing causes the component functions to be less distinguishable from each other, and, therefore, worsens the measurement resolution. Table 7.2 compares the K-factor of various observables in the top rest frame and in the lab frame. Remarkably, the K-factors in both frames have very similar values. This similarity, however, is coincidental. In the future, with improved $t\bar{t}$ reconstruction algorithms, the K-factors for top rest frame

observables should get smaller. Those of the lab frame observables, however, cannot be changed because their sizes are determined by the underlying physics.

Possibility of Measuring Top Spin Polarization in the '2'-Direction

A lab-frame measurement of polarization in the '2'-direction is impractical because no observable with any significant degree of correlation with $\cos U_2$ exist. In the top rest frame, however, this measurement is trivial.

Arguments in favor of lab frame observables are the following:

Simple Methodology

The shape of lab frame observable distributions depends on just two factors: (1) the energy and angular distribution of ℓ in the t_ℓ rest frame; and (2) the lab frame t_ℓ momentum distribution. In contrast, the shape of top rest frame observables depends on many details of the $t\bar{t}$ event structure. Therefore, lab frame observables rely much less on the details of monte carlo modeling than do the top rest frame observables. Low reliance on monte carlo translates to smaller systematic uncertainty.

Greater Data Sample Size

The data sample for an analysis using top rest frame observables is restricted to *reconstructable* $t\bar{t}$ candidate events — *i.e.* events with e or μ, large $\not\!\!E_T$, and four or more jets. In contrast, analyses using lab frame observables do not require full event reconstruction, so events satisfying looser cuts can be examined. In addition, candidate events in the dilepton channel can be examined. This translates to increased sample size, which, in turn, translates to smaller measurement error. It is estimated in appendix L that the inclusion of these extra events decreases the measurement error by 16%; this is equivalent to a 40% gain in statistics for the top rest frame-based analysis.

It is seen that there are benefits and liabilities in both methods. The choice in this thesis — to examine only the top rest frame observables — was made because of time constraints. Ideally, both methods should be used, and the result of each should be compared to check for consistency.

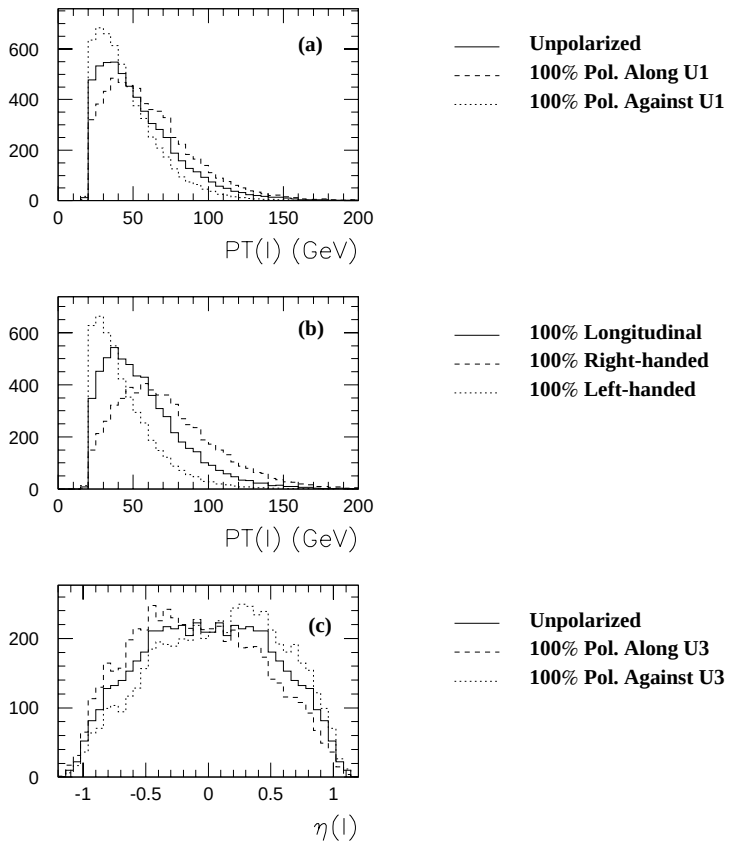


Figure 7.19: The distribution of the lab frame observables for probing: (a) Polarization along/against the '1'-direction; (b) the W helicity fractions; (c) Polarization along/against the '3' direction.

	K-factor	
Top Quark Property Being Measured	Lab	Top c.m.s.
Polarization *Along* the '1'-direction	3.76	3.78
Polarization *Against* the '1'-direction	3.51	3.48
Polarization Along/Against the '3'-direction	5.10	4.91
Transverse W Helicity Fraction	1.85	1.72

Table 7.2: The K-factors for various measurements of properties of the top quark. The column under "Lab" are the K-factors for lab frame observables, while that under "Top c.m.s." are the K-factors for top rest frame observables.

Chapter 8

Measurement Method and Resolution

8.1 Introduction

In chapter 7, the observables $\cos U_i$ $(i = 1, 2, 3)$ and E_ℓ were introduced. It was shown there that the shape of the $\cos U_i$ distribution depends on the degree of polarization along or against the "i"-direction. Similarly, the shape of the E_ℓ distribution depends on the W helicity fractions h_-, h_0, and h_+. In other words, the degree of polarization and the helicity fractions parameterize the possible shape of the observable distributions. Using this parameter dependence of the shape of the distributions, one can extract from the experimental data an estimate of the true parameter value. The extracted value can then be compared with the standard model parameter value to determine the degree of consistency between the experimental data and the standard model.

This chapter deals with the method used to extract from the experimental data the estimated true parameter value. The first section of this chapter is a discussion of the method — the 2-component minimum log-likelihood method. In this section, the method is defined, and an expression for the estimated statistical error of the parameter is derived. Following this is an application of the method, in which the estimated statistical error in the experimental data is calculated.

8.2 The Measurement Method

8.2.1 Introduction

The method used in this thesis to extract from the experimental data the parameters describing the top quark spin and helicity properties is the *2-component minimum log-likelihood method.* The full expression of the log-likelihood function is rather complicated. Therefore, in order to facilitate the description of the method, a *layered* approach is used. In other words, the description starts with simplified assumptions, and these assumptions are modified at the next level of description to take account of details that were left out of the previous level.

8.2.2 The Method, Part I

Suppose that there are N_{ev} events in the experimental data. As a first step, let us make the following simplifying assumptions:

- There are no background contributions to the N_{ev} events.
- The data are analyzed as a single, monolithic sample — *i.e.* no attempt is made to classify the data according to certain attributes, *e.g.* the number and type of b-tagged jets, the number of jets passing tight cuts, *etc.*

In later sections, these assumptions are relaxed, and the effect of relaxing them is examined.

Let x be the observable being studied, α a parameter that determines the shape of the distribution of x, and $f_0(x)$ and $f_1(x)$ the component distributions (*i.e.* the distribution of x when $\alpha = 0$ and 1, respectively). Both f_0 and f_1 are normalized to 1. Given these, the probability density that a measurement results in a value x, given that the parameter value is α, is given by the following:

$$f_\alpha(x) = \alpha \cdot f_1(x) + (1 - \alpha) \cdot f_0(x) \tag{8.1}$$

Table 8.1 gives the correspondence between the general symbols in the above equation with quantities in specific a measurement.

	MEASUREMENT	
	SPIN POLARIZATION	W HELICITY FRACTION
x	$\cos U_i$, $i = 1, 2, 3$	E_ℓ
α	Degree of polarization.	The "soft" transverse helicity fraction.
$f_0(x)$	The distribution of $\cos U_i$ from an *unpolarized* top quark.	The *medium* E_ℓ distribution.
$f_1(x)$	The distribution of $\cos U_i$ from a 100% *polarized* top quark.	The *soft* E_ℓ distribution.

Table 8.1: The meaning of the symbols in equation 8.1 for the top spin polarization measurement and the W helicity measurement.

An important point to note about the above equation is this: the shape of the x distribution when the parameter value is α is obtained by taking a "linear interpolation" between the shape of the distribution when $\alpha = 1$ and that when $\alpha = 0$. This is not generally true of all observables. It is, however, true of the observables $\cos U_i$ and E_ℓ that are examined in this thesis. That these distributions have the simple property expressed in equation 8.1 is due to the "separation" of the top quark spin property from the property of the t-b-W interaction vertex. This separation is discussed in section 7.3.2, and it is one of the important predictions about the top quark decay made by the standard model. See appendix K for more on this topic.

A special note is in order about the W helicity fraction measurement. It was shown in section 7.3.2 that, according to the standard model, E_ℓ is distributed according to the following relation:

$$g(E_\ell) = r_T \cdot g_s(E_\ell) + r'_T \cdot g_h(E_\ell) + (1 - r_T - r'_T) \cdot g_m(E_\ell) \tag{8.2}$$

The functions g_s, g_m, and g_h are the "soft", "medium", and "hard" E_ℓ distribution, respectively (see figure 7.7), while r_T and r'_T are the "soft" and "hard" transverse helicity fraction. According to the standard model, assuming $m_{top} = 175\ GeV$, $r_T = 0.30$ while $r'_T = 0.00$. In this thesis, the W helicity measurement will take r'_T to be fixed at zero. When this choice is adopted, the shape of the E_ℓ distribution is determined by a single parameter. This decision was made because a 1-parameter measurement has considerably smaller statistical error compared to a 2-parameter measurement. Since the amount of experimental data available from Tevatron Run I is quite limited, it is necessary to make such an assumption in order to obtain results that begin to be meaningful. See appendix N for a discussion on generalizing the W helicity measurement, where both r_T and r'_T are allowed to vary.

Let us denote by $\{x_i\}$ $(i = 1 \sim N_{ev})$ the set of measured value of the observable x in an experiment with N_{ev} events. The probability density that this set of measurement originate from a parent distribution with parameter value α is given by the product of the probability density in equation 8.1:

$$\prod_{i=1}^{N_{ev}} f_\alpha(x_i) = \prod_{i=1}^{N_{ev}} [\alpha \cdot f_1(x_i) + (1-\alpha) \cdot f_0(x_i)] \tag{8.3}$$

The parameter value α that maximizes this probability density is referred to as *the most probable (likely) value* of the parameter. The most probable value of α also *minimizes* the following quantity:

$$\mathcal{L}(\alpha) = \sum_{i=1}^{N_{ev}} -\log [\alpha \cdot f_1(x_i) + (1-\alpha) \cdot f_0(x_i)] \tag{8.4}$$

This equation is simply the negative of the logarithm of equation 8.3. The quantity $\mathcal{L}(\alpha)$ is referred to as the *negative log-likelihood function*, or simply the *log-likelihood function*. From a practical viewpoint, this quantity is more easily minimized than the product in equation 8.3 is maximized. For this reason, the log-likelihood function will henceforth be the primary object of interest in this discussion.

In order to understand the statistical properties of $\mathcal{L}(\alpha)$, it is useful to introduce the concept of *pseudo-experiment*. The following is a description of this concept:

- Assume that the true, theoretical value of the parameter is α_0. Then the theoretically expected distribution of the observable x is the following:

$$f_{\alpha_0}(x) = \alpha_0 \cdot f_1(x) + (1 - \alpha_0) \cdot f_0(x) \tag{8.5}$$

- Using monte carlo techniques, generate N_{ev} values of the observable x distributed according to $f_{\alpha_0}(x)$.
- Repeat the last step N_{exp} times. Each repetition is referred to as an "experiment".
- In each experiment, obtain the value of α that minimizes the log-likelihood function.

The result of a pseudo-experiment is a set of N_{exp} values of α that minimize the log-likelihood function in each experiment. This set of values is referred to as $\{\alpha_i\}$. In ideal situations, $\{\alpha_i\}$ is distributed as a gaussian centered at α_0. The measurements in this thesis correspond, to a good approximation, to this ideal situation. See appendix M for situations where the ideal distribution is not valid. The width of the gaussian, σ, is approximated by the following relation:

$$\frac{1}{\sigma^2} = \left.\frac{\partial^2 \mathcal{L}}{\partial \alpha^2}\right|_{\alpha=\alpha_0} \tag{8.6}$$

A useful approximation of σ can be obtained by making a continuous approximation of the log-likelihood function. To obtain the continuous version of equation 8.4, assume that N_{ev} is very large. Let x_{min} and x_{max} be the minimum and maximum value of the observable x.[1] Imagine dividing the range $[x_{min}, x_{max}]$ into N_{bin} equal-sized bins, so that all bins have width $\Delta x = (x_{max} - x_{min})/N_{bin}$. Then, if one assumes that the parent distribution has parameter value $\alpha = \alpha_0$, bin number i has approximately $N_{ev} \cdot f_{\alpha_0}(x_i) \cdot \Delta x$ events in it, where x_i is the average value of x in bin number i. Given this, the log-likelihood function $\mathcal{L}(\alpha)$ can be approximated as follows:

$$\mathcal{L}(\alpha) = -N_{ev} \sum_{i=1}^{N_{bin}} \Delta x \cdot f_{\alpha_0}(x_i) \cdot \log\left[\alpha \cdot f_1(x_i) + (1 - \alpha) \cdot f_0(x_i)\right] \tag{8.7}$$

[1] For the observable E_ℓ, which is unbounded above, one takes x_{max} to be a reasonably chosen cut-off value — *i.e.* a value beyond which the probability density is negligibly small.

Note that, unlike in equation 8.4, the summation is not over all events, but over all bins. Also, since the sum is over bins, the normalization factor N_{ev} is necessary. In the limit $\Delta x \rightarrow 0$, this expression becomes the following integral:

$$\mathcal{L}(\alpha) = -N_{ev} \int dx\, f_{\alpha_0}(x) \cdot \log\left[\alpha \cdot f_1(x) + (1-\alpha) \cdot f_0(x)\right] \tag{8.8}$$

The second derivative of this expression, evaluated at $\alpha = \alpha_0$, gives $1/\sigma^2$. The first derivative is the following:

$$\frac{\partial \mathcal{L}}{\partial \alpha} = -N_{ev} \int dx\, f_{\alpha_0}(x) \frac{f_1(x) - f_0(x)}{f_\alpha(x)} \tag{8.9}$$

When this expression is evaluated at $\alpha = \alpha_0$, $f_{\alpha_0}(x)$ in the numerator cancels out the $f_\alpha(x)$ in the denominator, giving the following:

$$\begin{aligned} \left.\frac{\partial \mathcal{L}}{\partial \alpha}\right|_{\alpha=\alpha_0} &= -N_{ev} \int dx\, \left[f_1(x) - f_0(x)\right] \\ &= -N_{ev} \cdot (1-1) \\ &= 0 \end{aligned} \tag{8.10}$$

This shows that the continuous approximation of $\mathcal{L}(\alpha)$ has an extremum at $\alpha = \alpha_0$. This is consistent with the expectation that, in the limit $N_{ev} \rightarrow \infty$, the value of α that minimizes $\mathcal{L}$ should be the true value α_0.

The second derivative of $\mathcal{L}$ is the following:

$$\frac{\partial^2 \mathcal{L}}{\partial \alpha^2} = N_{ev} \int dx\, f_{\alpha_0}(x) \frac{\left[f_1(x) - f_0(x)\right]^2}{f_\alpha(x)^2} \tag{8.11}$$

This expression, evaluated at $\alpha = \alpha_0$, is the following:

$$\left.\frac{\partial^2 \mathcal{L}}{\partial \alpha^2}\right|_{\alpha=\alpha_0} = N_{ev} \int dx\, \frac{\left[f_1(x) - f_0(x)\right]^2}{f_{\alpha_0}(x)} \tag{8.12}$$

Using equation 8.6, it is seen that this expression is equal to $1/\sigma^2$. In order to make the expression more compact, let us re-express the integral in equation 8.12 as follows:

$$\frac{1}{K^2} = \int dx\, \frac{\left[f_1(x) - f_0(x)\right]^2}{f_{\alpha_0}(x)} \tag{8.13}$$

Then one obtains the following for σ:

$$\sigma = \frac{K}{\sqrt{N_{ev}}} \tag{8.14}$$

An important point to note about this equation is the fact that σ factorizes into a part that depends on the size of the experimental data $(1/\sqrt{N_{ev}})$ and a part that is independent of it (K). The factor K is characteristic of an observable, and it is determined by the shape of the component functions f_0 and f_1 and the true parameter value α_0. It can be thought of as the purely geometric factor in determining the size of the parameter error. A particularly interesting point about the K factor is the fact that the size of K is determined, in large part, by the degree to which the component functions are distinguishable. This is clearly evident in equation 8.13, where the numerator in the integrand is the square of the difference between the component functions. The more distinguishable the component functions, the larger the difference between them, and hence the larger $1/K^2$ is, or the smaller K is. This property of K is exploited in chapter 9 to determine what factor in the $t\bar{t}$ reconstruction is most responsible for degrading the measurement resolution of the observables.

Before leaving this section, it should be noted that the expression for σ in equation 8.14 is only valid in the limit $N_{ev} \to \infty$. For finite N_{ev}, the K factor actually has a weak dependence on N_{ev}. K is largest for small N_{ev}, and decreases monotonically to the limiting value as N_{ev} increases. The difference between K at finite N_{ev} and K at $N_{ev} = \infty$ is not large. See appendix O for a discussion of the N_{ev} dependence of the K-factor.

8.2.3 The Method, Part II

In this section, let us relax the assumption about the background: instead of fixing the background to zero, let us allow an arbitrary value for the it. Specifically, let us say that a fraction β of the N_{ev} events is estimated to originate from background processes. Then, if $f_b(x)$ is the distribution of the observable x in the background events, the probability density for observing the value x, given the parameter value α, is the following:

$$f_\alpha(x) = \beta \cdot f_b(x) + (1-\beta) \cdot [\alpha \cdot f_1(x) + (1-\alpha) \cdot f_0(x)] \tag{8.15}$$

The log-likelihood function in this case is:

$$\mathcal{L}(\alpha) = -\sum_{i=1}^{N_{ev}} \log\left\{\beta \cdot f_b(x_i) + (1-\beta)\cdot[\alpha \cdot f_1(x_i) + (1-\alpha)\cdot f_0(x_i)]\right\} \tag{8.16}$$

In this expression, β is fixed — *i.e.* β is obtained from a separate measurement. Also, in this thesis, the uncertainty on β is ignored. This decision was made in order to focus on the most important issues surrounding the measurement; the measurement of β is not one of them.

The statistical properties of the log-likelihood function in this case is almost identical to those described in the last section. One important change, however, is in the size of the parameter error σ. Since the background fraction is β, the number of expected signal events is $S = (1-\beta)N_{ev}$. Then, naïvely, one might expect the statistical error to be the following:

$$\begin{aligned} \sigma_{naive} &= \frac{K}{\sqrt{S}} \\ &= \left(\frac{1}{\sqrt{1-\beta}}\right)\frac{K}{\sqrt{N_{ev}}} \end{aligned} \tag{8.17}$$

The K-factor here has nearly the same value as in the last section. This expression, however, is incorrect. The true expression for σ is obtained by evaluating the second derivative of $\mathcal{L}$ at $\alpha = \alpha_0$, where α_0 is the true parameter value:

$$\frac{1}{\sigma^2} = \left.\frac{\partial^2 \mathcal{L}}{\partial \alpha^2}\right|_{\alpha=\alpha_0} \tag{8.18}$$

If the continuum approximation of $\mathcal{L}(\alpha)$ is employed, then this expression evaluates to the following:

$$\frac{1}{\sigma^2} = N_{ev} \cdot (1-\beta)^2 \cdot \int dx \, \frac{[f_1(x) - f_0(x)]^2}{f_{\alpha_0}(x)} \tag{8.19}$$

If K is defined so that the integral in the above equation equals $1/K^2$, then the formula for σ is given by the following expression:

$$\sigma = \left(\frac{1}{1-\beta}\right)\frac{K}{\sqrt{N_{ev}}} \tag{8.20}$$

It is seen that the true error is larger than the naïve error by a factor of $1/\sqrt{1-\beta}$. The origin of this extra factor of $1/\sqrt{1-\beta}$ can be understood by examining the form of the continuum approximation of $\mathcal{L}$:

$$\mathcal{L}(\alpha) = -N_{ev} \int dx\, f_{\alpha_0}(x) \cdot \log\left[\beta \cdot f_b(x) + (1-\beta) \cdot \{\alpha \cdot f_1(x) + (1-\alpha) \cdot f_0(x)\}\right] \quad (8.21)$$

The α-dependent part of this expression is multiplied by the factor $1-\beta$. Thus, every derivative of $\mathcal{L}$ with respect to α introduces an overall factor of $1-\beta$. Since $1/\sigma^2$ is equal to the second derivative of $\mathcal{L}$, one sees that σ must be proportional to $1/\sqrt{(1-\beta)^2} = 1/(1-\beta)$. The fact that the true error is larger than the naïve error by a factor of $1/\sqrt{1-\beta}$ implies that the presence of background events degrades the measurement resolution in more ways than simply reducing the number of signal events.

The expression for σ in equation 8.20 can be re-expressed as follows:

$$\sigma = K\left(\frac{\sqrt{S+B}}{S}\right) \quad (8.22)$$

The quantity B is the expected number of background events. The fact that σ has this form is significant in two ways. First, it is known that the quantity H_T — the scalar sum of E_T of all $t\bar{t}$ decay products — is useful in distinguishing signal events from background events. By applying the cut $H_T > U$, one can improve the purity of the event sample, at the cost of losing some signal events. The quantities S and B are functions of U: $S = S(U)$ and $B = B(U)$. Given equation 8.22, it is clear that the value of the cut U that optimizes the statistical error is that which maximizes the ratio $S/\sqrt{S+B}$. This property of σ is used to decrease the statistical error in section 8.3.3.

The second way in which the form of equation 8.22 is significant is this: under certain conditions, the statistical error can be made smaller by subdividing the event sample $\{x_i\}$ into subsamples. This is described in detail in the next section.

8.2.4 The Method, Part III

So far, the experimental data $\{x_i\}$ has been treated as a unit, with a single background estimate β, single component functions $f_0(x)$ and $f_1(x)$, and a single background

function $f_b(x)$. However, because the statistical error has the form given in equation 8.22, it is, under certain conditions, advantageous to subdivide the event sample into subsamples.

In order to see that this is so, let us first, as an example, subdivide the event sample into two parts: $\{x_i\}_1$ and $\{x_i\}_2$. For the sake of concreteness, let us take subsample 1 to be the subset of events with at least one b-tagged jet, while subsample 2 is the complement of subsample 1, *i.e.* events with no b-tagged jets. The log-likelihood expression in this case is:

$$\mathcal{L}(\alpha) = \mathcal{L}_1(\alpha) + \mathcal{L}_2(\alpha) \tag{8.23}$$

$$\mathcal{L}_\lambda(\alpha) = -\sum_{i=1}^{N_\lambda} \log\left\{\beta_\lambda \cdot f_b^\lambda(x_i) + (1-\beta_\lambda)\cdot\left[\alpha \cdot f_1^\lambda(x_i) + (1-\alpha)\cdot f_0^\lambda(x_i)\right]\right\} \tag{8.24}$$

The index λ in equation 8.24 refers to subsamples 1 and 2. N_λ is the number of events and β_λ is the background fraction, both in subsample λ. Similarly, the functions $f_0^\lambda(x)$, $f_1^\lambda(x)$, and $f_b^\lambda(x)$ are the various distributions of x in subsample λ.

The statistical error on the parameter α for the subsample λ taken alone is:

$$\sigma_\lambda^2 = \frac{K_\lambda^2}{(1-\beta_\lambda)^2\, N_\lambda} \tag{8.25}$$

$$= \frac{K_\lambda^2}{(1-\beta_\lambda)^2\, r_\lambda \cdot N_{ev}} \tag{8.26}$$

The quantity $r_\lambda = N_\lambda / N_{ev}$ is the fraction of events belonging to subsample λ. When the subsamples are combined, the total statistical error is:

$$\frac{1}{\sigma^2} = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} \tag{8.27}$$

This rule for combining the subsample errors is obtained by noting: (1) the fact that the $1/\sigma^2$ is equal to the second derivative of $\mathcal{L}$; and (2) the log-likelihood function is additive (equation 8.23). Substituting equation 8.25 into equation 8.27, and after some rearrangements, one obtains the following:

$$\left(\frac{1}{N_{ev}}\right)\frac{1}{\sigma^2} = r\left[\frac{(1-\beta_1)^2}{K_1^2}\right] + (1-r)\left[\frac{(1-\beta_2)^2}{K_2^2}\right] \tag{8.28}$$

Typically, the K-factor has similar values across all subsamples. Thus one can use the following approximation: $K_1 \approx K_2 \approx K$, where K is an "average" value of K_1 and K_2. Using this approximation, equation 8.28 can be rewritten as follows:

$$\left(\frac{K^2}{N_{ev}}\right)\frac{1}{\sigma^2} = r\cdot(1-\beta_1)^2 + (1-r)\cdot(1-\beta_2)^2 \tag{8.29}$$

Let us compare this to the original situation where the experimental data is grouped into a single sample. The background fraction in this situation is the average of β_1 and β_2:

$$\beta = r\cdot\beta_1 + (1-r)\cdot\beta_2 \tag{8.30}$$

The expression corresponding to equation 8.29 is:

$$\left(\frac{K^2}{N_{ev}}\right)\frac{1}{\tilde{\sigma}^2} = (1-\beta)^2 \tag{8.31}$$

It can be shown from equations 8.29 and 8.31 that $\sigma \leq \tilde{\sigma}$ — *i.e.* that the statistical error is improved, or, at worst, stays the same, when the event sample is divided into subsamples. To see that this is so, let us examine the ratio of the two errors:

$$\frac{\tilde{\sigma}^2}{\sigma^2} = \frac{r\cdot(1-\beta_1)^2 + (1-r)\cdot(1-\beta_2)^2}{(1-\beta)^2} \tag{8.32}$$

$$= \left(\frac{1}{r}\right)\frac{1+R\varepsilon^2}{(1+R\varepsilon)^2} \tag{8.33}$$

$$R = \frac{1-r}{r} \tag{8.34}$$

$$\varepsilon = \frac{1-\beta_2}{1-\beta_1} \tag{8.35}$$

An analysis of equation 8.33 show the following:

- $\tilde{\sigma}^2/\sigma^2 \geq 1$ for all values of r, β_1, and β_2.
- The smallest possible value of the ratio is 1, and this occurs only when $\varepsilon = 1$. In other words, the statistical error always improves unless the subsamples have the same background fraction, in which case the error stays the same.

- As ε is decreased from 1 to 0, the ratio increases monotonically from 1 to $1/r$. Similarly, as ε is increased from 1 to ∞, the ratio increases monotonically from 1 to $1/(1-r)$. The points $\varepsilon = 0$ and $\varepsilon = \infty$ correspond to the situations where one of the subsamples contains 100% background events.

It is seen that the statistical error can always be decreased so long as one can divide the original data sample into subsamples with different background fractions. The improvement in error is larger the greater the difference in the background fractions.

The example above involved two subsamples. However, the result generalizes to an arbitrary number of subsamples. One can, in principle, continue the process of subdivision *ad infinitum*. In practice, a point of diminishing return is reached pretty quickly because the estimated background fractions do have errors, so that every time a new subsample is created, a new source of systematic uncertainty is introduced. In this thesis, the experimental data are divided into eight parts. First, the data are categorized according to the number of jets passing tight cuts (see chapter 4 for the definition of the cuts). If an event has four or more tight jets, it belongs to the "4-jets" sample, denoted as *NJ4*. If an event has only three such jets, it belongs to the "3.5-jets" sample, denoted *NJ3.5*.[2] These subsamples are further subdivided into four subsamples, categorized according to the nature of the b-tags in the event. They are the following:

SVX only

The event has only SVX tagged jets. This category is denoted by the symbol XO, for SV<u>X</u> <u>O</u>nly.

SLT only

The event has only SLT tagged jets. This category is denoted by the symbol TO, for SL<u>T</u> <u>O</u>nly.

SVX and SLT

The event has both an SVX and SLT tagged jets. The tags are allowed to be on a single jet. This category is denoted by the symbol XT, for SV<u>X</u> and SL<u>T</u>.

[2]The notation "3.5-jets" is just a suggestive way of saying that there are only three jets passing the tight cuts and one or more jets passing the looser cuts.

No Tags

No b-tagged jets exist. This category is denoted by the symbol NT, for No Tags.

These subsamples are mutually exclusive. The log-likelihood expression for this case is the following:

$$\mathcal{L}(\alpha) = \sum_{\lambda} \mathcal{L}_{\lambda}(\alpha) \tag{8.36}$$

The summation index λ is over the eight subsamples. Each subsample has a log-likelihood function $\mathcal{L}_{\lambda}$ given by equation 8.24.

8.3 An Estimate of the Statistical Error in the Experimental Data

8.3.1 Introduction

In this section, the results of the last section are used to obtain an estimate of the statistical error in the top quark spin polarization and W helicity measurement in the experimental data. In the first section, the estimated statistical error, before the analysis is optimized, is given. Following this, a couple of optimizations in the analysis are described, and the error after the optimizations is presented. In the final section, the results from the previous sections are corrected to take account the fact that $N_{ev} < \infty$.

8.3.2 Estimated Statistical Error before Optimization

Given the log-likelihood expression in equation 8.36, and given the fact that $1/\sigma^2$ is equal to the $\partial^2\mathcal{L}/\partial\alpha^2$ evaluated at $\alpha = \alpha_0$, the statistical error of the parameter α is given by the following equations:

$$\frac{1}{\sigma^2} = \sum_{\lambda=1}^{8} \frac{1}{\sigma_{\lambda}^2} \tag{8.37}$$

$$\frac{1}{\sigma_{\lambda}^2} = \frac{(1-\beta_{\lambda})^2\, r_{\lambda} \cdot N_{ev}}{K_{\lambda}^2} \tag{8.38}$$

UNOPTIMIZED									
					K_λ				
SUBSAMPLE		N_λ	r_λ	β_λ	$U1+$	$U1-$	$U2$	$U3$	E_ℓ
NJ3.5	XO	4	0.0252	0.29	3.48	3.22	3.33	4.87	1.76
	TO	5	0.0314	0.66	3.69	3.14	3.19	4.73	1.81
	XT	4	0.0252	0.18	3.78	3.63	3.32	3.90	1.84
	NT	77	0.4843	0.90	4.00	3.52	3.20	5.25	1.68
NJ4	XO	11	0.0692	0.07	3.31	3.17	2.89	3.98	1.73
	TO	9	0.0566	0.25	3.62	3.51	3.48	5.08	1.78
	XT	4	0.0252	0.04	3.41	3.30	2.96	4.13	1.72
	NT	45	0.2830	0.58	3.75	3.57	3.16	4.92	1.77

Table 8.2: The numbers necessary for calculating the combined estimated statistical error of the parameters. These numbers are those obtained before the analysis is optimized.

The summation index λ is over the eight subsamples of the experimental data. The quantity σ_λ is the statistical error of the parameter when subsample λ alone is used in the measurement, r_λ is the fraction of events belonging to subsample λ, and β_λ and K_λ are the background fraction and K-factor, respectively, for subsample λ. Table 8.2 gives, for each observable, the numbers necessary to calculate the combined statistical error. In the table, the symbols $U1$, $U2$, and $U3$ stand for $\cos U_i$, $i = 1, 2, 3$. The '1'-direction is forward-backward asymmetric, so the situation where the $\cos U_1$ distribution is asymmetric along and against the axis must be treated separately. $U1+$ stands for the situation where the polarized distribution favors positive values of $\cos U_1$, while $U1-$ stands for the situation where negative $\cos U_1$ is favored.

Table 8.3 shows $1/\sigma_\lambda^2$ for all of the observables. The final row is the sum of $1/\sigma_\lambda^2$. In the right-hand part of the table is the fractional contribution of $1/\sigma_\lambda^2$ to the total. The larger this fraction is, the more statistically important the subsample is. From this point of view, the subsample (NJ4, XO) is the most important subsample in the experimental data. It should be noted that the statistical importance of a subsample is only very loosely correlated with the number of events in the subsample. For example, the subsample (NJ3.5, NT) accounts for 48.4% of all events, but its statistical importance is only about 2%. Simi-

UNOPTIMIZED											
		$1/\sigma_\lambda^2$					$fr(1/\sigma_\lambda^2)$ (%)				
SUBSAMPLE		$U1+$	$U1-$	$U2$	$U3$	E_ℓ	$U1+$	$U1-$	$U2$	$U3$	E_ℓ
NJ3.5	XO	0.167	0.194	0.181	0.084	0.655	6.5	6.8	5.5	5.1	6.3
	TO	0.043	0.059	0.057	0.025	0.176	1.7	2.1	1.7	1.6	1.7
	XT	0.189	0.205	0.243	0.178	0.798	7.3	7.2	7.3	10.7	7.6
	NT	0.049	0.062	0.075	0.027	0.272	1.9	2.2	2.3	1.7	2.6
NJ4	XO	0.867	0.943	1.132	0.599	3.169	33.6	33.2	34.1	36.2	30.4
	TO	0.385	0.410	0.415	0.197	1.595	14.9	14.4	12.5	11.9	15.3
	XT	0.318	0.340	0.421	0.218	1.248	12.3	12.0	12.7	13.2	12.0
	NT	0.563	0.626	0.793	0.326	2.523	21.8	22.1	23.9	19.6	24.2
Total:		2.581	2.839	3.317	1.654	10.436					

Table 8.3: $1/\sigma_\lambda^2$ for all observables, before optimization. The sum of all the terms in each column is shown in the last row. The right-hand side of the table, under $fr(1/\sigma_\lambda^2)$, is the ratio of $1/\sigma_\lambda^2$ to the total, in %. This fraction is a measure of the statistical importance of a subsample.

UNOPTIMIZED					
	$U1+$	$U1-$	$U2$	$U3$	E_ℓ
σ	0.62	0.59	0.55	0.78	0.31

Table 8.4: The statistical error for each observable, before optimization.

larly, although the subsample (NJ4, XO) accounts for only 6.9% of all events, its statistical importance is about 33%.

The statistical error of the parameter is obtained by taking the square root of the inverse of the numbers in the last row of table 8.3. This is shown for each observable in table 8.4. The following are some remarks concerning the values in this table:

- The statistical errors for $\cos U_i$ are about twice as large as that for E_ℓ. This is mostly due to the fact that, at the analytic level, the component functions for E_ℓ are more distinct from each other than they are for $\cos U_i$. See chapter 7 for a discussion about this.

- Among the observables $\cos U_i$, $\cos U_2$ has the smallest error, while $\cos U_3$ has the

largest error. This trend can be understood on the basis of the relationship between the '1'-'2'-'3' coordinate system (figure 7.8) and the detector. See chapter 7 for more on this.

- For each observable, the physically meaningful range of parameter values is $\alpha \in [0, 1]$. This shows that, for the observables $\cos U_i$, the errors cover most of this range. Therefore, one cannot expect a very meaningful measurement from the amount of data available from Tevatron Run I. Compared to $\cos U_i$, the error for E_ℓ covers a considerably smaller region; nevertheless, the error is too large to conclude anything definitive from the experimental data.

- The errors in table 8.4 can be made smaller by optimizing the analysis. This optimization is discussed in the next section.

- The error in table 8.4 is, in fact, a lower-bound on the true error. This is because the analytic formulas used to obtain them are valid only in the limit $N_{ev} \to \infty$. In section 8.3.4, this issue is discussed, and a correction factor is obtained in order to obtain the true error.

8.3.3 Estimated Statistical Error after Optimization

The statistical error discussed in the last section can be made smaller by a couple of optimizations in the analysis. These optimizations involve: (1) cutting on the quantity H_T; and (2) applying a top mass constraint to the $t\bar{t}$ reconstruction algorithm.

Let us first consider the H_T cut. The quantity H_T is defined as the sum of the transverse energy of all of the decay products of the $t\bar{t}$ system. It is known that background processes have considerably smaller H_T compared to that from $t\bar{t}$ decay, assuming that $m_{top} = 175\ GeV$. Because of this, H_T is very effective at separating the signal from the background. By choosing H_T judiciously, one can decrease the statistical error of the parameters describing the top quark spin and helicity properties.

In order to decide what value of H_T to cut on, it is useful to express the statistical error for the subsample λ as follows:

$$\sigma_\lambda = K_\lambda \left(\frac{\sqrt{S_\lambda + B_\lambda}}{S_\lambda} \right) \tag{8.39}$$

The quantities S_λ and B_λ are the number of signal and background events, and K_λ the K-factor, all for the subsample λ. All of these quantities vary with the H_T cut. However, the variation of K_λ with the cut value is small, so it can be taken as a constant. Then, the value of the H_T cut to choose is that which minimizes $\sqrt{S_\lambda + B_\lambda}/S_\lambda$. Alternatively, one can choose to maximize $S_\lambda^2/(S_\lambda + B_\lambda)$. This can be re-expressed as follows:

$$\frac{S_\lambda^2}{S_\lambda + B_\lambda} = N_\lambda \left[\frac{[(1-\beta)\cdot\epsilon_s(U)]^2}{(1-\beta)\cdot\epsilon_s(U) + \beta\cdot\epsilon_b(U)} \right] \tag{8.40}$$

The quantity β is the background fraction before the H_T cut is applied, $\epsilon_s(U)$ and $\epsilon_b(U)$ are the fraction of signal and background events surviving the cut $H_T > U$, and N_λ is the number of events in subsample λ. The quantity inside of the square brackets is referred to as the *significance*. The optimal value of the H_T cut is that which maximizes the significance. Figure 8.1 shows the significance as a function of the H_T cut for each subsample. It is seen that a clear maximum exists only for the subsamples (NJ3.5, XO), (NJ3.5, NT), and (NJ4, NT). In the optimized analysis, the cuts $H_T > 237.5\ GeV$, $266.9\ GeV$, and $286.3\ GeV$ are applied to the subsamples (NJ3.5, XO), (NJ3.5, NT), and (NJ4, NT), respectively.

The second optimization in the analysis that can improve the statistical error is the application of a mass constraint in the $t\bar{t}$ reconstruction algorithm (see chapter 6 for details on the algorithm). In the default setting of the algorithm, the terms in χ^2 for the semileptonically and hadronically decaying top quark (t_ℓ and t_h, respectively) have the following form:

$$\chi^2(t_\ell) + \chi^2(t_h) = \left(\frac{m_{\ell\nu b} - MTOP}{\Gamma_{top}} \right)^2 + \left(\frac{m_{jjb} - MTOP}{\Gamma_{top}} \right)^2 \tag{8.41}$$

The quantity $m_{\ell\nu b}$ is the invariant mass of the ℓ-ν_ℓ-jet system presumed to originate from the decay of t_ℓ, while m_{jjb} is the invariant mass of the 3-jet system presumed to originate from the decay of t_h. The quantity Γ_{top} is equal to the width of the top quark mass obtained from tree-level calculations, and is equal to $2.5\ GeV$ (assuming $m_{top} = 175\ GeV$), and $MTOP$ is one of the parameters used to minimize the total χ^2 (see appendix D for more on χ^2).

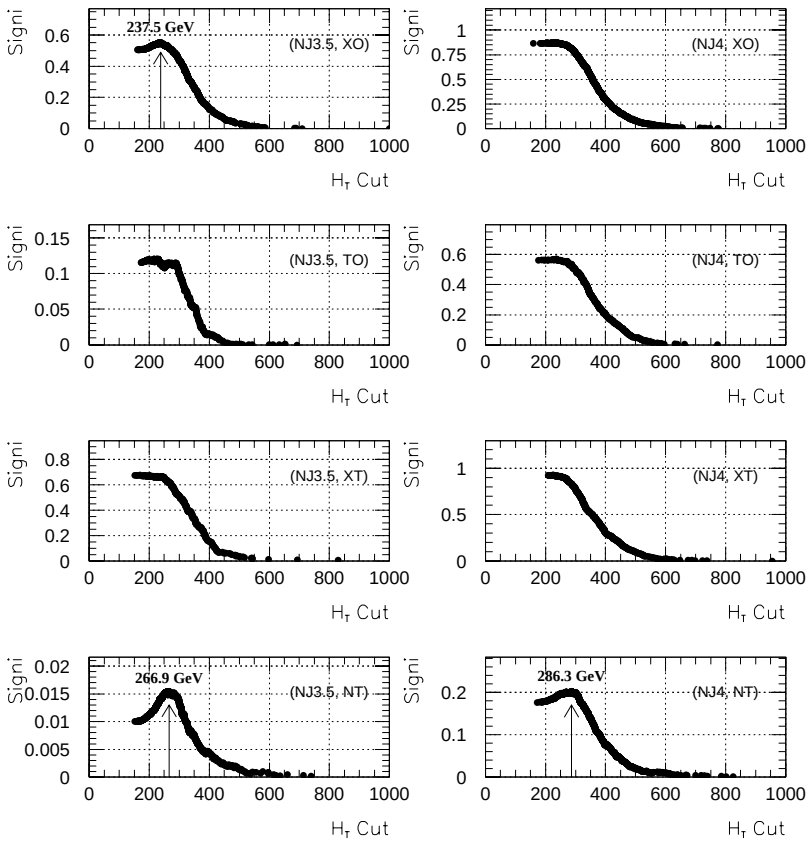


Figure 8.1: The significance as a function on U, where U is the cut on H_T given by the relation $H_T > U$.

		$f_{b\ell}$ (%)	
SUBSAMPLE		BEFORE	AFTER
NJ3.5	XO	48.6	53.3
	TO	42.0	42.7
	XT	50.6	54.7
	NT	37.3	41.4
NJ4	XO	53.5	58.5
	TO	44.5	48.4
	XT	53.4	56.9
	NT	41.0	44.9

Table 8.5: The fraction of events $f_{b\ell}$ with the label b_ℓ applied to the correct jet, before and after applying the top mass constraint.

These terms in χ^2 favor configurations with the mass of t_ℓ and t_h with similar values. What those values are, however, is not constrained by these terms. In other words, as long as the masses are similar, it does not matter whether the masses are about 140 GeV, 200 GeV, or whatever value.

In the optimized setting of the $t\bar{t}$ reconstruction algorithm, one makes use of the fact that the top quark mass m_{top} has been measured to be 175 GeV. Since the top mass has already been measured, the parameter $MTOP$ no longer needs to be free, so $MTOP$ is fixed at 175 GeV. When this is done, the χ^2 terms in equation 8.41 favor configurations with the mass of t_ℓ and t_h that are not only close to each other, but close to 175 GeV.

The direct benefit of applying the top quark mass constraint is the improvement in the fraction of events where the output of the $t\bar{t}$ reconstruction algorithm has the quark label b_ℓ applied to the correct jet. This fraction is referred to as $f_{b\ell}$. Table 8.5 compares $f_{b\ell}$ before and after the top mass constraint is applied to the algorithm. An indirect consequence of applying the top mass constraint is this. In chapter 9, it is demonstrated that the single greatest source of smearing in the measurement is the matching of the quark label b_ℓ to the wrong jet. Thus, an improvement in $f_{b\ell}$ results in decreased smearing. This decreased smearing is reflected in the decreased value of the K-factors, which, in turn, result in decreased statistical error.

Tables 8.6 through 8.8 show the numbers necessary to calculate the statistical errors,

H_T CUT, NO TOP MASS CONSTRAINT									
					K_λ				
SUBSAMPLE		N_λ	r_λ	β_λ	$U1+$	$U1-$	$U2$	$U3$	E_ℓ
NJ3.5	XO	4	0.0351	0.19	3.48	3.19	3.21	4.80	1.76
	TO	5	0.0439	0.66	3.69	3.14	3.19	4.73	1.81
	XT	4	0.0351	0.18	3.78	3.63	3.32	3.90	1.84
	NT	45	0.3947	0.81	3.86	3.22	3.01	5.03	1.74
NJ4	XO	11	0.0965	0.07	3.31	3.17	2.89	3.98	1.73
	TO	9	0.0789	0.25	3.62	3.51	3.48	5.08	1.78
	XT	4	0.0351	0.04	3.41	3.30	2.96	4.13	1.72
	NT	32	0.2807	0.46	3.77	3.51	3.12	4.84	1.78

Table 8.6: The numbers necessary for calculating the combined estimated statistical error of the parameters. H_T cuts *are* applied, top mass *unconstrained.*

$1/\sigma_\lambda^2$ for each subsample, and the statistical error when the H_T cut is applied, but before applying the top mass constraint. Tables 8.9 through 8.11 show the corresponding numbers for the case where the top mass constraint is applied, but the H_T cut is not applied. Finally, tables 8.12 through 8.14 show the numbers when both optimization techniques are applied. Table 8.15 summarizes statistical errors.

8.3.4 Correcting the Errors for the Fact that $N_{ev} < \infty$

The statistical errors obtained so far (table 8.15) underestimate the true value because the formulas used to obtain them (equations 8.37 and 8.38) are valid only in the limit $N_{ev} \to \infty$. In order to obtain the true value of the errors, one must use the pseudo-experiment method discussed in page 147. The true value of the errors is shown in table 8.16. Table 8.17 shows the percent difference between the true and approximate errors. As claimed earlier, the approximate error underestimates the true error, but only by a few percent. In a small number of cases, the true error is smaller than the approximate error; this is most likely due to statistical fluctuation in determining the true error. See appendix O for further discussion on true versus approximate statistical error.

H_T CUT, NO TOP MASS CONSTRAINT											
		$1/\sigma_\lambda^2$					$fr(1/\sigma_\lambda^2)$ (%)				
SUBSAMPLE		$U1+$	$U1-$	$U2$	$U3$	E_ℓ	$U1+$	$U1-$	$U2$	$U3$	E_ℓ
NJ3.5	XO	0.218	0.258	0.254	0.113	0.848	7.8	8.2	7.0	6.3	7.5
	TO	0.043	0.059	0.057	0.025	0.176	1.7	2.1	1.7	1.6	1.7
	XT	0.189	0.205	0.243	0.178	0.798	7.3	7.2	7.3	10.7	7.6
	NT	0.108	0.156	0.180	0.065	0.538	3.9	5.0	4.9	3.6	4.7
NJ4	XO	0.867	0.943	1.132	0.599	3.169	33.6	33.2	34.1	36.2	30.4
	TO	0.385	0.410	0.415	0.197	1.595	14.9	14.4	12.5	11.9	15.3
	XT	0.318	0.340	0.421	0.218	1.248	12.3	12.0	12.7	13.2	12.0
	NT	0.653	0.756	0.961	0.401	2.958	23.5	24.2	26.2	22.3	26.1
Total:		2.781	3.127	3.663	1.796	11.330					

Table 8.7: $1/\sigma_\lambda^2$ for all observables. The sum of all the terms in each column is shown in the last row. The right-hand side of the table, under $fr(1/\sigma_\lambda^2)$, is the ratio of $1/\sigma_\lambda^2$ to the total, in %. The H_T cuts *are* applied, top mass *unconstrained.*

H_T CUT, NO TOP MASS CONSTRAINT						
	$U1+$	$U1-$	$U2$	$U3$	E_ℓ
σ	0.60	0.57	0.52	0.75	0.30

Table 8.8: The statistical error for each observable. H_T cuts *are* applied, top mass *unconstrained.*

NO H_T CUT, TOP MASS CONSTRAINED									
					K_λ				
SUBSAMPLE		N_λ	r_λ	β_λ	$U1+$	$U1-$	$U2$	$U3$	E_ℓ
NJ3.5	XO	4	0.0252	0.29	3.30	3.05	2.99	4.25	1.56
	TO	5	0.0314	0.66	3.56	2.93	3.44	5.35	1.12
	XT	4	0.0252	0.18	3.15	3.00	3.01	4.13	1.68
	NT	77	0.4843	0.90	3.65	3.23	3.24	5.01	1.67
NJ4	XO	11	0.0692	0.07	3.20	3.06	2.63	3.79	1.63
	TO	9	0.0566	0.25	3.51	3.43	3.12	4.47	1.69
	XT	4	0.0252	0.04	3.25	3.14	2.86	6.48	1.57
	NT	45	0.2830	0.58	3.51	3.35	3.12	4.54	1.75

Table 8.9: The numbers necessary for calculating the combined estimated statistical error of the parameters. *No* H_T cuts applied, top mass *constrained.*

NO H_T CUT, TOP MASS CONSTRAINED											
		$1/\sigma_\lambda^2$					$fr(1/\sigma_\lambda^2)$ (%)				
SUBSAMPLE		$U1+$	$U1-$	$U2$	$U3$	E_ℓ	$U1+$	$U1-$	$U2$	$U3$	E_ℓ
NJ3.5	XO	0.186	0.218	0.226	0.111	0.828	6.4	6.8	5.9	6.5	6.9
	TO	0.046	0.067	0.048	0.021	0.458	1.6	2.1	1.3	1.2	3.8
	XT	0.272	0.299	0.296	0.159	0.959	9.4	9.4	7.8	9.3	8.0
	NT	0.057	0.073	0.073	0.030	0.277	2.0	2.3	1.9	1.8	2.3
NJ4	XO	0.933	1.018	1.380	0.666	3.568	32.2	31.9	36.2	39.0	29.8
	TO	0.410	0.431	0.522	0.253	1.778	14.2	13.5	13.7	14.8	14.9
	XT	0.351	0.377	0.450	0.089	1.503	12.1	11.8	11.8	5.2	12.6
	NT	0.642	0.706	0.817	0.382	2.587	22.2	22.1	21.4	22.3	21.6
Total:		2.897	3.189	3.812	1.711	11.958					

Table 8.10: $1/\sigma_\lambda^2$ for all observables. The sum of all the terms in each column is shown in the last row. The right-hand side of the table, under $fr(1/\sigma_\lambda^2)$, is the ratio of $1/\sigma_\lambda^2$ to the total, in %. *No* H_T cuts applied, top mass *constrained.*

NO H_T CUT, TOP MASS CONSTRAINED						
	$U1+$	$U1-$	$U2$	$U3$	E_ℓ
σ	0.59	0.56	0.51	0.76	0.29

Table 8.11: The statistical error for each observable. *No* H_T cuts applied, top mass *constrained.*

H_T CUT, TOP MASS CONSTRAINED									
					K_λ				
SUBSAMPLE		N_λ	r_λ	β_λ	$U1+$	$U1-$	$U2$	$U3$	E_ℓ
NJ3.5	XO	4	0.0351	0.19	3.37	3.10	2.86	4.22	1.57
	TO	5	0.0439	0.66	3.56	2.93	3.44	5.35	1.12
	XT	4	0.0351	0.18	3.15	3.00	3.01	4.13	1.68
	NT	45	0.3947	0.81	3.67	3.10	2.95	4.74	1.58
NJ4	XO	11	0.0965	0.07	3.20	3.06	2.63	3.79	1.63
	TO	9	0.0789	0.25	3.51	3.43	3.12	4.47	1.69
	XT	4	0.0351	0.04	3.25	3.14	2.86	6.48	1.57
	NT	32	0.2807	0.46	3.52	3.31	3.00	4.34	1.66

Table 8.12: The numbers necessary for calculating the combined estimated statistical error of the parameters. H_T cuts applied *and* top mass constrained.

H_T CUT, TOP MASS CONSTRAINED											
		$1/\sigma_\lambda^2$					$fr(1/\sigma_\lambda^2)$ (%)				
SUBSAMPLE		$U1+$	$U1-$	$U2$	$U3$	E_ℓ	$U1+$	$U1-$	$U2$	$U3$	E_ℓ
NJ3.5	XO	0.231	0.274	0.320	0.147	1.068	7.4	7.8	7.6	7.7	8.0
	TO	0.046	0.067	0.049	0.021	0.458	1.5	1.9	1.1	1.1	3.4
	XT	0.271	0.299	0.296	0.158	0.958	8.7	8.6	7.0	8.4	7.2
	NT	0.120	0.169	0.187	0.072	0.649	3.9	4.9	4.4	3.8	4.8
NJ4	XO	0.933	1.018	1.379	0.666	3.568	29.9	29.2	32.6	35.1	26.7
	TO	0.409	0.430	0.521	0.253	1.776	13.1	12.4	12.3	13.3	13.3
	XT	0.350	0.376	0.450	0.089	1.501	11.2	10.8	10.6	4.7	11.2
	NT	0.756	0.849	1.036	0.495	3.387	24.2	24.4	24.4	26.0	25.3
Total:		3.116	3.482	4.238	1.901	13.365					

Table 8.13: $1/\sigma_\lambda^2$ for all observables, before optimization. The sum of all the terms in each column is shown in the last row. The right-hand side of the table, under $fr(1/\sigma_\lambda^2)$, is the ratio of $1/\sigma_\lambda^2$ to the total, in %. H_T cuts applied *and* top mass constrained.

H_T CUT, TOP MASS CONSTRAINED					
	$U1+$	$U1-$	$U2$	$U3$	E_ℓ
σ	0.57	0.54	0.49	0.73	0.27

Table 8.14: The statistical error for each observable. H_T cuts applied and top mass constrained.

H_T CUT	TOP MASS CONSTRAINT	$U1+$	$U1-$	$U2$	$U3$	E_ℓ
NO	NO	0.62	0.59	0.55	0.78	0.31
YES	NO	0.60	0.57	0.52	0.75	0.30
NO	YES	0.59	0.56	0.51	0.76	0.29
YES	YES	0.57	0.54	0.49	0.73	0.27

Table 8.15: Summary of the statistical errors, before the analysis is optimized, and with various combinations of optimizing techniques.

H_T CUT	TOP MASS CONSTRAINT	$U1+$	$U1-$	$U2$	$U3$	E_ℓ
NO	NO	0.636	0.609	0.558	0.798	0.319
YES	NO	0.611	0.573	0.522	0.764	0.299
NO	YES	0.595	0.578	0.520	0.790	0.299
YES	YES	0.581	0.541	0.489	0.737	0.280

Table 8.16: The true statistical error for each observable. In each case, the values are accurate to the third decimal place.

H_T CUT	TOP MASS CONSTRAINT	$U1+$	$U1-$	$U2$	$U3$	E_ℓ
NO	NO	$+2.5\%$	$+3.1\%$	$+1.4\%$	$+2.3\%$	$+2.8\%$
YES	NO	$+1.8\%$	$+0.5\%$	$+0.4\%$	$+1.8\%$	-0.3%
NO	YES	$+0.8\%$	$+3.1\%$	$+2.0\%$	$+3.8\%$	$+3.0\%$
YES	YES	$+1.9\%$	$+0.2\%$	-0.2%	$+0.9\%$	$+3.6\%$

Table 8.17: The percent difference between the true errors (table 8.16) and the approximate errors (table 8.15).

Chapter 9

Sources of Degradation of the Measurement Resolution

9.1 Introduction

In chapter 7, it is shown that the distribution of the observables $\cos U_i$ $(i = 1, 2, 3)$ and E_ℓ is a linear combination of two component functions:

$$f_\alpha(x) = \alpha \cdot f_1(x) + (1 - \alpha) \cdot f_0(x) \tag{9.1}$$

The symbol x stands for $\cos U_i$ or E_ℓ. When $x = \cos U_i$, $f_0(x)$ and $f_1(x)$ are the distribution of x from a 0% and 100% polarized top quark, and α is the degree of polarization. When $x = E_\ell$, $f_0(x)$ and $f_1(x)$ are the soft and medium E_ℓ distribution, and α is the soft transverse helicity fraction r_T. In chapter 8, the two-component log-likelihood method is introduced. This method allows one to obtain an estimate of the true value of the parameter α, given a set of observable values $\{x_i\}$. The statistical error σ of the estimated parameter value is related to the log-likelihood function according to the following relation:

$$\frac{1}{\sigma^2} = \left.\frac{\partial^2 \mathcal{L}}{\partial \alpha^2}\right|_{\alpha=\alpha_0} \tag{9.2}$$

Using a continuum approximation of the log-likelihood function, σ can be approximated as follows:

$$\sigma = K \cdot \frac{1}{(1 - \beta)} \cdot \frac{1}{\sqrt{N_{ev}}} \tag{9.3}$$

N_{ev} is the number of events in an experiment, β is the estimated background fraction in the event sample, and K is a geometric factor that depends mostly on the shape of the component functions $f_0(x)$ and $f_1(x)$:

$$\frac{1}{K^2} = \int dx \, \frac{[f_1(x) - f_0(x)]^2}{f_{\alpha_0}(x)} \tag{9.4}$$

$$f_{\alpha_0}(x) = \beta \cdot f_b(x) + (1-\beta) \cdot [\alpha_0 \cdot f_1(x) + (1-\alpha_0) \cdot f_0(x)] \tag{9.5}$$

The function $f_{\alpha_0}(x)$ is the theoretically expected distribution of x, and $f_b(x)$ is the distribution of x in the background.

The use of the continuum approximation of $\mathcal{L}$ allows one to express the statistical error in a modular form — *i.e.* the statistical error can be factorized into a statistical term $1/\sqrt{N_{ev}}$, a background term $1/(1-\beta)$, and a geometric term K. The focus of this chapter is on the geometric term. By examining the size of K under various conditions, one can deduce what aspect of the $t\bar{t}$ reconstruction process is most responsible for degrading the measurement resolution. The main objective of this chapter is the demonstration of the fact that the matching of the wrong jet to b_ℓ (b-quark originating from t_ℓ, the semileptonically decaying top quark) is by far the most important source of degradation the measurement resolution. In addition to this, sundry observations on the factors that affect the size of the statistical error for all of the observables are made.

9.2 Analyzing the Degradation of Measurement Resolution

The analysis of the degradation of measurement resolution is based on a monte carlo simulation of $t\bar{t}$ events in the lepton + jets channel. In this analysis, the $t\bar{t}$ system output by the $t\bar{t}$ reconstruction algorithm is compared with the parton-level system. Changes in the $t\bar{t}$ system in going form the parton-level to the output results in the the smearing of the observable distributions. This smearing, in turn, causes degradation in the measurement resolution. The objective of this analysis is to determine what aspects of the $t\bar{t}$ reconstruction algorithm are most responsible for the degradation.

The ultimate source of degradation is the mismeasurement of the momentum and energy of the physics objects. In the lepton + jets decay channel of $t\bar{t}$ events, the possible physics objects are: (1) charged lepton; (2) jets; and (3) unclustered energy. Of these, the mismeasurement of the jet momentum accounts for most of the degradation in the measurement resolution. The mismeasurement of the jet momentum affects the following aspects of the $t\bar{t}$ reconstruction process:

- The measurement of the neutrino momentum
- The matching of jets to quarks

The second item can be further subdivided into the following:

- The matching of a jet to b_ℓ
- The matching of jets to the other quarks

These three aspects of the $t\bar{t}$ reconstruction process are correlated with each other, so one cannot meaningfully say which one is most important in degrading the measurement resolution. One can, however, gauge the importance of one of the factors by comparing the statistical error when that aspect is successfully accomplished to that when it is unsuccessfully accomplished. In what follows, the following is demonstrated:

> The assignment of a wrong jet to b_ℓ is the most important source of degradation of the measurement resolution in the following sense: when the assignment is correct, the K-factor (and, therefore, the statistical error) is not much larger than that at the parton-level, while when the assignment is incorrect, it is much larger.

The first step in demonstrating this is to take the monte carlo sample of $t\bar{t}$ events passing the lepton + jets event selection cuts and divide it into two mutually exclusive sets, where one set has $\Delta R(b_\ell) < 0.4$, while the other has $\Delta R(b_\ell) \geq 0.4$, where $\Delta R(b_\ell)$ is the η-ϕ separation between the true b_ℓ momentum and the momentum of the jet assigned to b_ℓ by the reconstruction algorithm. In other words, the first set has the correct jet associated with b_ℓ, and the second set has an incorrect jet associated with it.

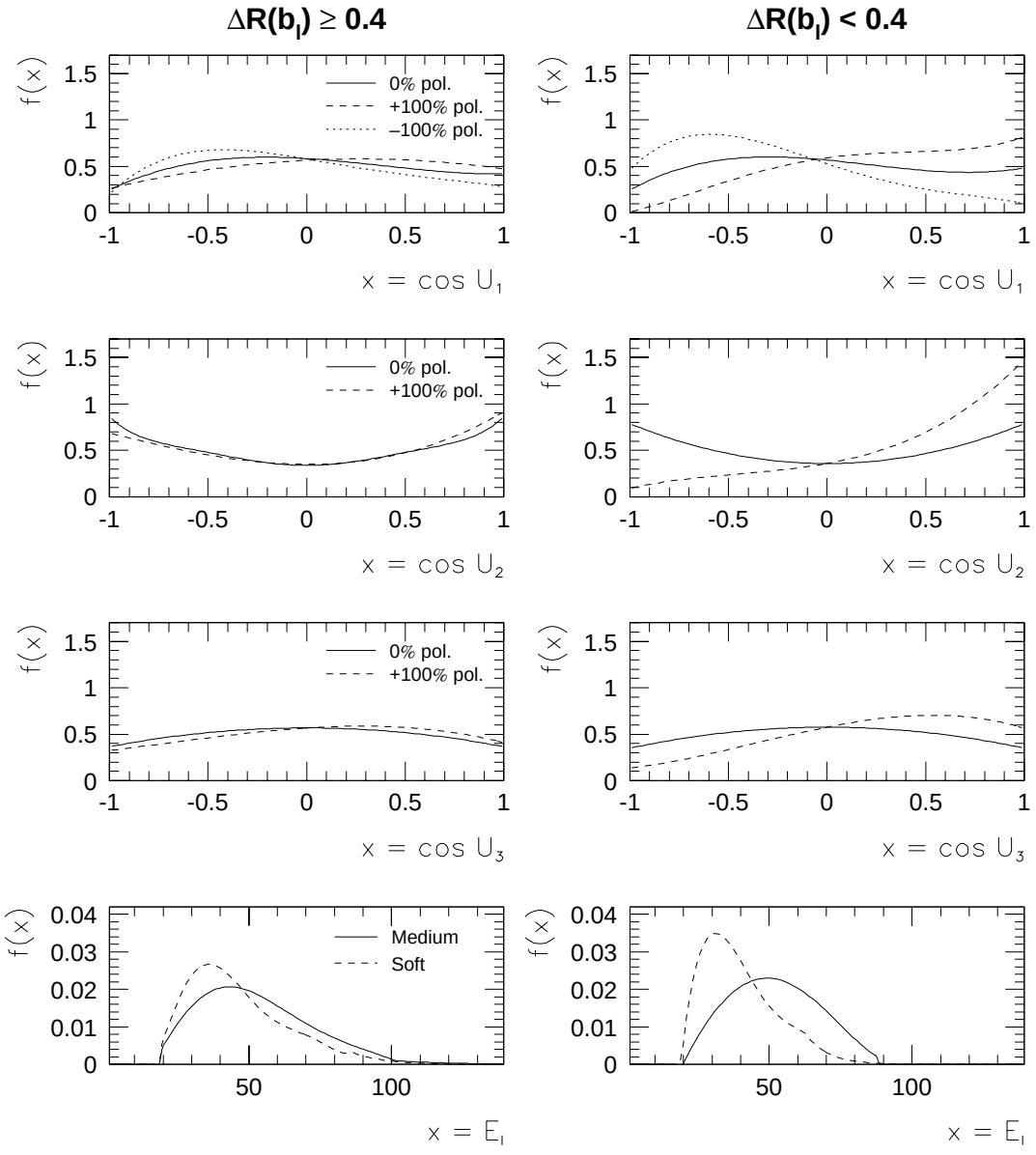


Figure 9.1: The distribution of the component functions $f_0(x)$ and $f_1(x)$ for the observables $\cos U_i$ and E_ℓ when $\Delta R(b_\ell) \geq 0.4$ (left column) and $\Delta R(b_\ell) < 0.4$ (right column).

The next step is to obtain the component functions $f_0(x)$ and $f_1(x)$ in the two subsets. See figure 9.1. The figure shows that when the jet-b_ℓ matching is done correctly, the component functions are very much distinct from one another, whereas when the matching is incorrect, the distinction is, to varying degrees, lost because of smearing. Equation 9.4 shows that the greater the degree of distinction between the component functions, the smaller K is, and, therefore, the smaller the statistical error is. The figure, therefore, illustrates qualitatively that jet-b_ℓ mismatch is a large source of degradation of the measurement resolution. In order to demonstrate this quantitatively, the following strategy is used:

1. Calculate the K-factor at the parton-level.[1] This will be denoted $K(parton)$.

2. Calculate the K-factor for the set of events satisfying $\Delta R(b_\ell) < 0.4$. This value of the K-factor corresponds to the situation where the fraction $f_{b\ell}$ of events with correct jet-b_ℓ matching is 100% — *i.e.* $f_{b\ell} = 1$. The K-factor in this situation is denoted $K(1)$.

3. Obtain the component functions in the general case where a fraction $f_{b\ell}$ of events has $\Delta R(b_\ell) < 0.4$, and a fraction $(1 - f_{b\ell})$ has $\Delta R(b_\ell) \geq 0.4$. This is done by simply taking a linear combination of the functions in the sets $\Delta R(b_\ell) < 0.4$ and $\Delta R(b_\ell) \geq 0.4$:

$$f_i^{f_{b_\ell}}(x) = f_{b\ell} \cdot f_i^{<}(x) + (1 - f_{b\ell}) \cdot f_i^{\geq}(x) \tag{9.6}$$

   The index $i = 0$ and 1. The function $f_i^{<}(x)$ is the component function for events satisfying $\Delta R(b_\ell) < 0.4$, and $f_i^{\geq}(x)$ is defined similarly.

4. Using $f_i^{f_{b_\ell}}(x)$, obtain the K-factor as a function of $f_{b\ell}$. This is denoted as $K(f_{b\ell})$.

5. Compare $K(1)$ with $K(parton)$. Show that $K(1)$ is not much larger than $K(parton)$. This demonstrates that the mismeasurement of the neutrino momentum and the mismatching of the other quarks to jets is not an important source of measurement degradation when $\Delta R(b_\ell) < 0.4$.

[1] In the calculation of the K-factor for this and all other cases in this section, the background fraction β is set to zero in equation 9.4. The β-dependence of K is through $f_{\alpha_0}(x)$ in the denominator. For the observables considered in this thesis, K depends only weakly on β, so the exact choice of β is immaterial. Also, the monte carlo sample is not divided into eight subsets as described in section 8.2.4. Instead, the shape of the component functions $f_0(x)$ and $f_1(x)$ are approximated as being the same across all the eight subsets. This approximation does not affect the conclusions reached in this section.

	U1+	U1−	U2	U3	E_ℓ
ANALYTIC	1.73	1.73	1.73	1.73	0.90
PARTON	1.90	1.78	1.54	1.87	1.02
OUT, $f_{b\ell} = 1$	2.20	2.13	1.69	2.73	1.19
OUT, $f_{b\ell} = 0.5$	3.41	3.23	3.08	4.36	1.76

Table 9.1: The K-factor at various stages.

	U1+	U1−	U2	U3	E_ℓ
PARTON → OUT, $f_{b\ell} = 1$	15.8	20.0	9.7	46.0	16.7
PARTON → OUT, $f_{b\ell} = 0.5$	79.5	81.5	100	133	72.5

Table 9.2: The % change in the K-factor going from the parton-level to the output with $f_{b\ell} = 1$, and to the output with $f_{b\ell} = 0.5$.

6. Compare $K(1)$ to $K(f_{b\ell})$ with $f_{b\ell}$ at typical values — anywhere between $f_{b\ell} = 0.45$ to 0.65. Show that $K(f_{b\ell})$ is much larger than $K(1)$. This demonstrates that the loss of measurement resolution is mostly due to the b_ℓ-jet mismatch.

Figure 9.2 shows $K(f_{b\ell})$ and $K(parton)$ for all of the observables. Figure 9.3(a) shows $K(f_{b\ell})$ for each observable superposed in the same plot, while figure 9.3(b) shows the ratio $K(f_{b\ell})/K(parton)$. Table 9.1 shows the K-factor at the analytic stage, parton-level stage, at the output stage with $f_{b\ell} = 1$, and the output stage with $f_{b\ell} = 0.5$. Table 9.2 shows the percent change in going from the parton-level to the output with $f_{b\ell} = 1$, and to the output with $f_{b\ell} = 0.5$. The following are some observations on these plots and table:

Change in K from PARTON → OUT

Table 9.2 shows that, for all observables except $\cos U_3$, the change in the K-factor in going from the parton-level to the output stage with $f_{b\ell} = 1$ is not very large — 10 to 20%. The corresponding change in K when $f_{b\ell}$ has a typical value of 0.5 is very large — between 70 to 100%. This proves the assertion that jet-b_ℓ mismatch is the primary source of degradation in the measurement resolution. The observable $\cos U_3$ is somewhat exceptional because the worsening of K in going from PARTON → OUT, $f_{b\ell} = 1$ is rather large. However, compared to the change in K going from PARTON

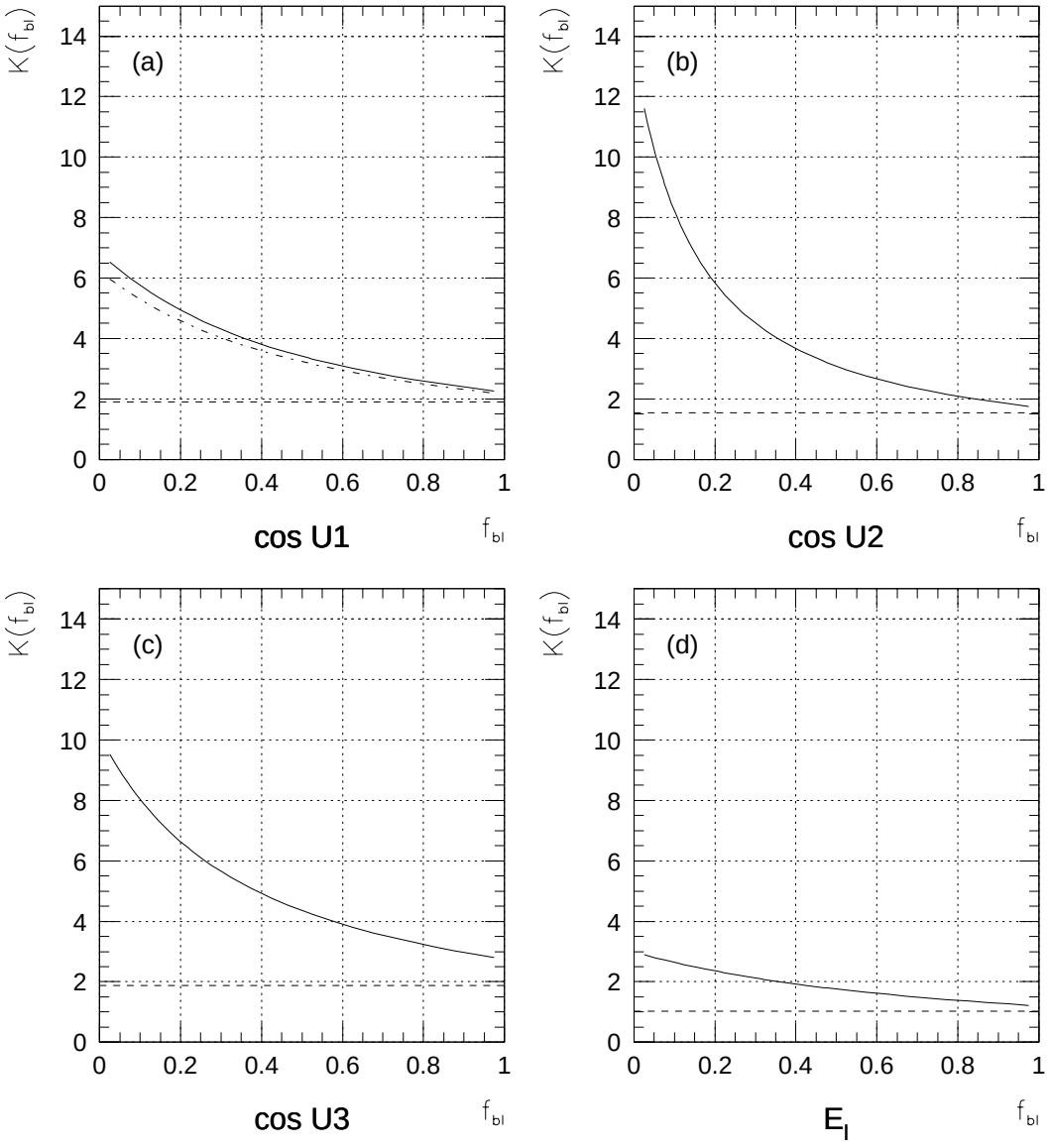


Figure 9.2: $K(f_{b\ell})$ vs. $f_{b\ell}$ for the observables $\cos U_i$ and E_ℓ. The dashed line in each frame shows $K(parton)$. For $\cos U_1$, the solid curve is $K(f_{b\ell})$ for U1+, and the dot-dashed curve is for U1−, where U1+ represents a top quark 100% spin polarized such that the $\cos U_1$ distribution is asymmetric along the '1'-direction, and U1− is that for which the distribution is asymmetric against the '1'-direction.

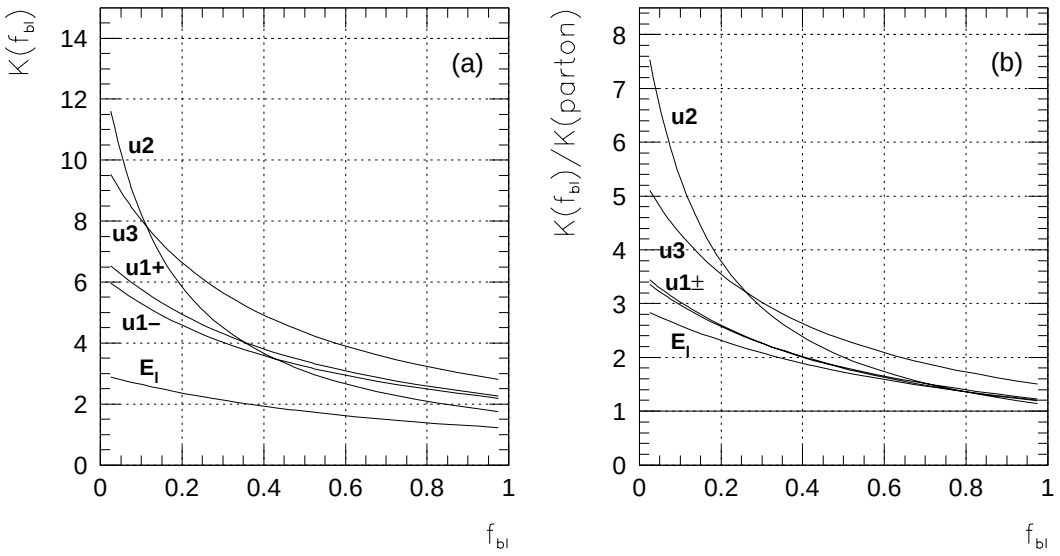


Figure 9.3: (a) $K(f_{b\ell})$ for each observable superposed in the same plot. (b) $K(f_{b\ell})/K(parton)$ for each observable.

$\rightarrow$ OUT, $f_{b\ell} = 0.5$, this change is relatively small, so that the conclusion regarding the role of jet-b_ℓ mismatch in the degradation of the measurement resolution is still valid.

The Value of K at the Analytic Stage

At the analytic stage, the K-factor for $\cos U_i$ are all the same. The difference between the K-factor at the parton and output stages are due to differences between the efficiency curve and smearing matrix for each observable (see chapter 7 on efficiency curves and smearing matrices). The fact that the analytic K-factor for $\cos U_i$ is about twice as large as that for E_ℓ implies the following: the W helicity measurement has an intrinsically better measurement resolution compared to the top spin polarization measurement. The intrinsic superiority of the W helicity resolution is due to the fact that the component functions for E_ℓ at the analytic level are more distinguishable from each other compared to those for $\cos U_i$. It should be noted that this intrinsic superiority of the W helicity resolution survives the acceptance and smearing effects, as is seen in table 9.1, where the K-factor for E_ℓ is roughly a factor of two smaller compared to those of $\cos U_i$ at all stages.

Why is K for $\cos U_3$ Large?

Earlier, it was noted that $\cos U_3$ is exceptional among the observables because the change in the K-factor in going from the PARTON $\rightarrow$ OUT, $f_{b\ell} = 1$ is rather large compared to that in other observables. The reason for this is that the '3'-direction is strongly dependent on the longitudinal motion of the $t\bar{t}$ system, whereas the '1'- and '2'-directions and E_ℓ are only weakly dependent on it (see figure 7.8 on the definition of the '1'-, '2'-, and '3'-directions). Since the longitudinal momentum of the $t\bar{t}$ system has poor resolution, the fact that the '3'-direction depends on it results in relatively poor resolution for $\cos U_3$, even when the jet-b_ℓ matching is correct.

Why does K for $\cos U_2$ Increase Rapidly as $f_{b\ell} \rightarrow 0$?

One striking feature of the observable $\cos U_2$ is the fact that its K-factor is excellent at $f_{b\ell} = 1$, but it rapidly worsens as $f_{b\ell} \rightarrow 0$. In figure 9.2, $K(1)$ is under 2, while $K(0)$ is almost 12. This is in contrast to the other observables, where $K(f_{b\ell})$ does not worsen nearly as rapidly. This behavior is explainable by the following fact: no lab-frame observable has significant correlation with $\cos U_2$. This is in contrast to the other observables, for which such lab-frame observables exist (see section 7.5.3). To illustrate how the existence of such a lab-frame observable keeps $K(0)$ from being very large, let us take E_ℓ. Figure 7.18 shows that E_ℓ is strongly correlated with the lab-frame observable $P_T(\ell)$. Thus, for example, if the true value of E_ℓ were large, this implies that $P_T(\ell)$ will tend to be large. On the other hand, if $P_T(\ell)$ is large, then the value of E_ℓ output by the $t\bar{t}$ reconstruction algorithm will tend to be large, even if the jet-b_ℓ matching is incorrect. This last observation is due to the fact that if the lab frame energy of an object is large, then its energy after boosting in a random direction tends to be large. Thus an observable such as E_ℓ is not smeared completely when the jet-b_ℓ matching is unsuccessful.

Change in K from ANALYTIC $\rightarrow$ PARTON

Table 9.1 shows that the K-factor changes by just a little in going from ANALYTIC $\rightarrow$ PARTON. K can either increase or decrease. The change in K is due to the modification of the analytic distribution by the acceptance effect. The fact that K for $\cos U_2$ decreases (*i.e.* the statistical error improves) is due to the fact that the

acceptance curve for $\cos U_2$ is largest where the difference between the component functions is largest, and smallest where the difference is smallest. This means that the difference between the two component functions is enhanced by the acceptance effect. In contrast, the other observables have acceptance curves that have small values where the difference between the component functions is large. Thus the K-factors worsen somewhat going from ANALYTIC $\rightarrow$ PARTON. See figures 7.11 through 7.13.

Chapter 10

Analysis of the Experimental Data

10.1 Introduction

In this chapter, the measurement of the degree of spin polarization of the top quark and of the soft transverse W helicity fraction are described. The measurements are made using the minimum log-likelihood technique described in chapter 8.

10.2 The Data Sample

Details about the event selection for $t\bar{t}$ in the lepton + jets channel are given in chapter 4. That information is summarized here.

Before any H_T cut is made, there are 159 candidate events. Table 10.1(a) shows a breakdown of these events into eight subsets (see 8.2.4 for a definition of the subsets). Also shown is the expected background fraction for each subset. From these numbers, one obtains 55 as the expected number of signal events contributing to the 159 candidate events, giving a signal fraction of 35%.

In section 8.2.4, it is shown that a cut of $H_T > 237.5\ GeV$, $H_T > 266.9\ GeV$, and $H_T > 286.3\ GeV$ on the subsets (NJ3.5, XO), (NJ3.5, NT), and (NJ4, NT), respectively, optimizes the statistical error on the top polarization and W helicity measurements. After these cuts are made, 114 candidate events survive. Table 10.1(b) shows a breakdown of these events into the eight subsets. The expected number of signal events in this case is, again, 55, giving a signal fraction of 48%.

		(a) NO HT CUT		(b) HT CUT	
SUBSAMPLE		N_λ	β_λ	N_λ	β_λ
NJ3.5	XO	4	0.29	4	0.19
	TO	5	0.66	5	0.66
	XT	4	0.18	4	0.18
	NT	77	0.90	45	0.81
NJ4	XO	11	0.07	11	0.07
	TO	9	0.25	9	0.25
	XT	4	0.04	4	0.04
	NT	45	0.58	32	0.46

Table 10.1: The breakdown of the number of candidate events in each subsample (N_λ), and the estimated background fraction in each subsample (β_λ).

		(a) NO HT CUT		(b) HT CUT	
SUBSAMPLE		N_{ev} $(q_\ell = -)$	N_{ev} $(q_\ell = +)$	N_{ev} $(q_\ell = -)$	N_{ev} $(q_\ell = +)$
NJ3.5	XO	2	2	2	2
	TO	4	1	4	1
	XT	2	2	2	2
	NT	35	42	20	25
NJ4	XO	5	6	5	6
	TO	4	5	4	5
	XT	2	2	2	2
	NT	18	27	13	19
TOTAL:		72	87	52	62

Table 10.2: The breakdown of the event sample according to the sign of the charge of ℓ.

Table 10.2 shows the breakdown of the candidate events according to the sign of the charge of ℓ.

10.3 The Observable Distributions

Figures 10.1 to 10.3 show the distribution of the observables $\cos U_i$ $(i = 1, 2, 3)$ before optimization — *i.e.* no H_T cut and no top mass constraint, or (HTCUT, MTCON) = (NO, NO). Separate plots are shown for $q_\ell = -$, $+$, $-$ & $+$, and charge-weighted. The

charge-weighted plot shows the distribution of $q_\ell \cdot \cos U_i$ — in other words, it combines the histogram for $q_\ell = +$ with the parity-inversion of the histogram for $q_\ell = -$. This plot ought to be examined because of the possibility of the following situation. Suppose that some physical process causes both t and $\bar{t}$ to be 100% spin polarized along the direction i. In events where t decays semileptonically, ℓ^+ has a strong tendency to travel *along* the i direction, whereas in events where $\bar{t}$ decays semileptonically, ℓ^- has a strong tendency to travel *against* the i direction. Thus, in the combined plot $-$ & $+$, the asymmetries in the two distributions cancel each other out, resulting in the obscuring of the presence of spin polarization. The presence of spin polarization in this sort of situation can be detected by examining the charge-weighted distribution.

In figures 10.1 to 10.3 , the points show the experimental data. The solid line shows the monte carlo distribution of the signal + background, with the signal events having 0% spin polarization — this is the theoretically expected distribution. The dashed curve shows the monte carlo distribution of the signal + background with the signal 100% polarized such that the distribution favors positive values of $\cos U_i$. The dotted curve shows the corresponding distribution when the signal is 100% polarized such that the distribution favors negative values of $\cos U_i$.[1] It should be noted that the monte carlo distributions are obtained by combining the signal and background distributions in each of the eight categories in the proportions shown in tables 10.1 and 10.2.

Figure 10.4 shows the distribution of E_ℓ. Separate plots are shown for $q_\ell = -$, $+$, and $-$ & $+$. A charge-weighted distribution is not applicable in this case. The points show the distribution in the experimental data. The solid line shows the signal + background distribution, with the signal distributed according to the standard model — *i.e.* $r_T = 25\%$.[2] The dashed line shows the distribution distribution of signal + background where the signal has $r_T = 100\%$ (*i.e.* 100% soft E_ℓ distribution). The dotted line shows the corresponding distribution with $r_T = 0\%$ (*i.e.* 100% medium E_ℓ distribution).

[1] For $q_\ell = -$, the dashed curve shows the distribution where $\bar{t}$ has 100% spin polarization *against* the i direction, while the dotted curve shows the distribution where the spin polarization is 100% *along* the i direction. The sense of the spin is opposite for $q_\ell = +$. For $q_\ell = -$ & $+$, the dashed curve shows the distribution where t has spin polarization *along* the i direction and $\bar{t}$ has spin polarization *against* this direction. Arguing in this manner, one can obtain the spin configurations for the rest of the curves.

[2] This is after taking account of the acceptance effect. Before acceptance, $r_T = 30\%$, assuming $m_{top} = 175\ GeV$. See section 7.4.2.

Figures 10.5 to 10.8 show the background-subtracted distributions for $\cos U_i$ and E_ℓ.

All of the plots described above show the distributions before optimizations are applied — *i.e.* (HTCUT, MTCON) = (NO, NO). Figures 10.9 to 10.32 show corresponding plots for the optimizations (HTCUT, MTCON) = (YES, NO), (NO, YES), and (YES, YES).

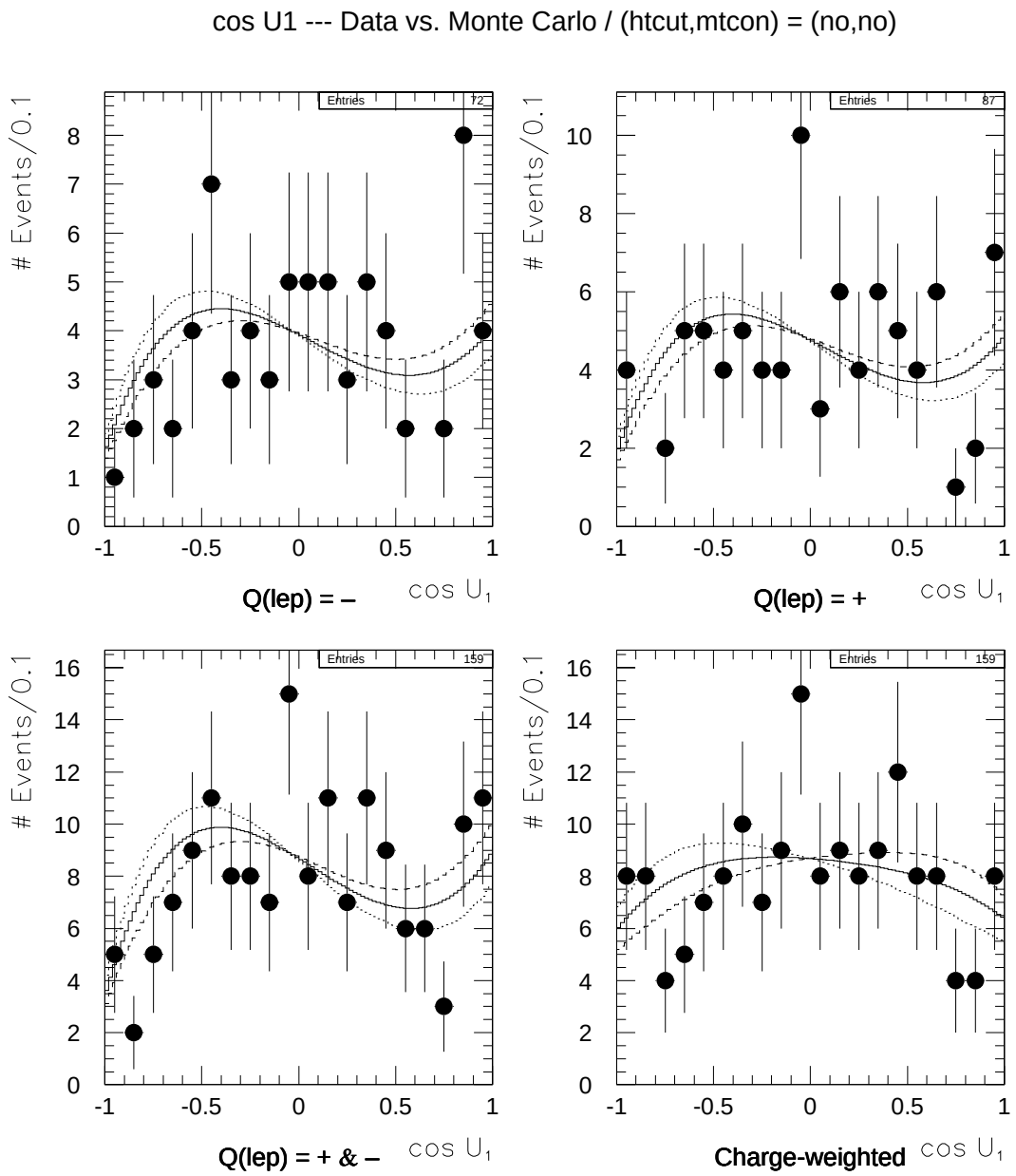


Figure 10.1: The distribution of $\cos U_1$. Points = experimental data; solid curve = monte carlo background + 0% polarized signal; dashed curve = background + 100% polarized signal, polarization favoring positive $\cos U_1$; dotted curve = background + 100% polarized signal, polarization favoring negative $\cos U_1$. The upper-left plot is for primary lepton charge $-$, the upper-right for charge $+$. The lower-left plot is a combination of the upper plots. The lower-right plot is a combination of the plot for $q_\ell = +$ with the parity-inversion of that for $q_\ell = -$.

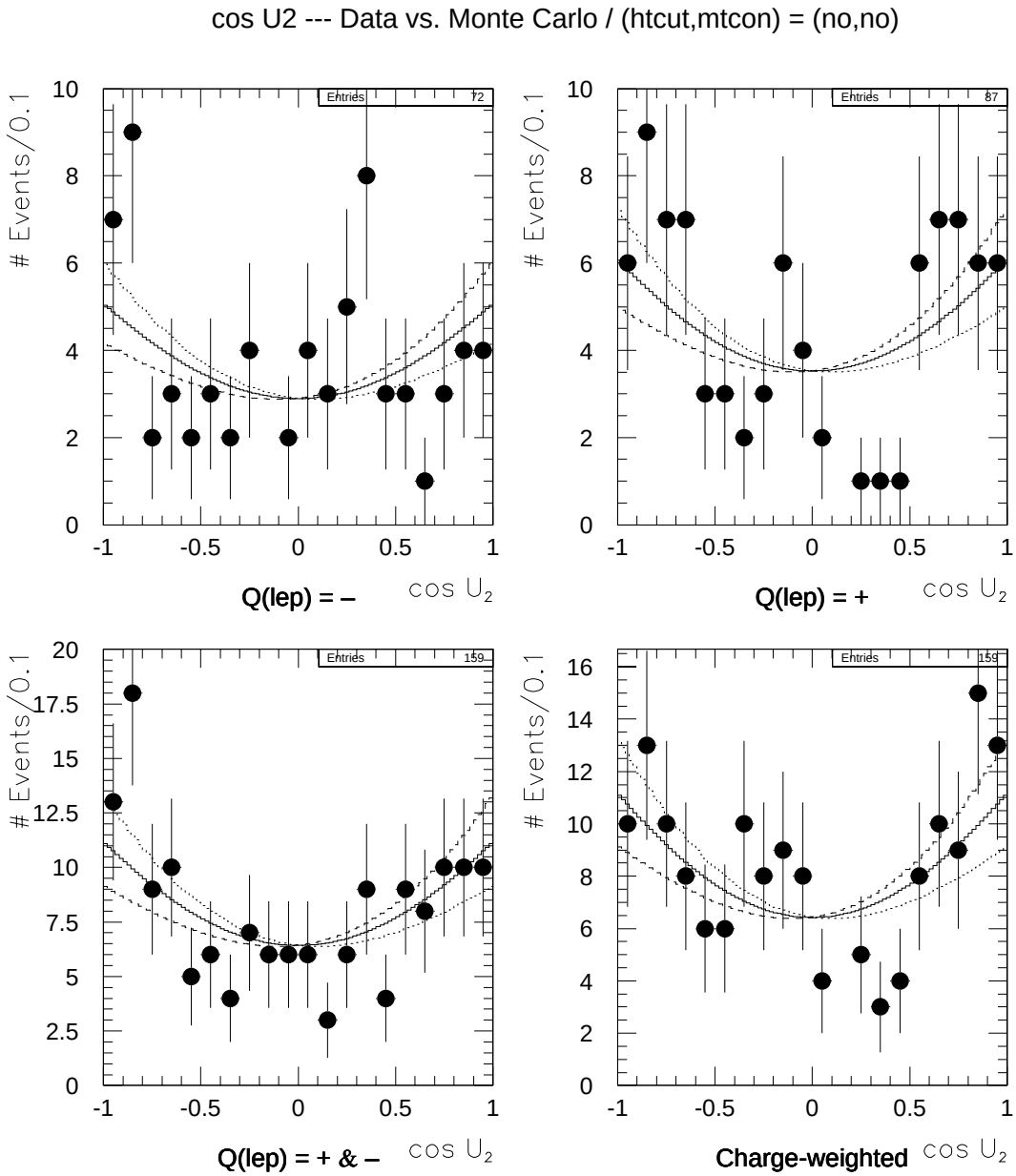


Figure 10.2: Same as figure 10.1, for the observable $\cos U_2$.

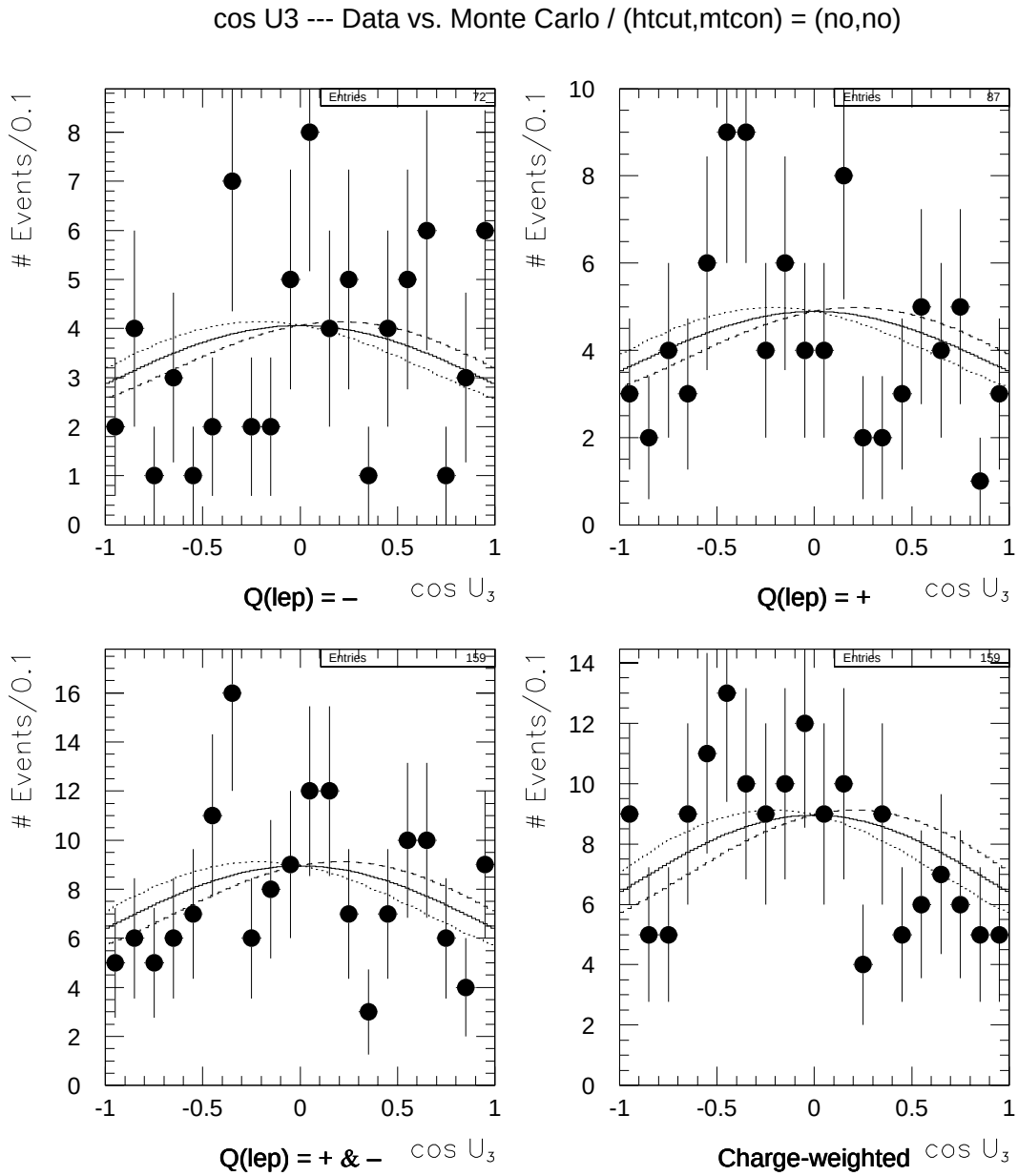


Figure 10.3: Same as figure 10.1, for the observable $\cos U_3$.

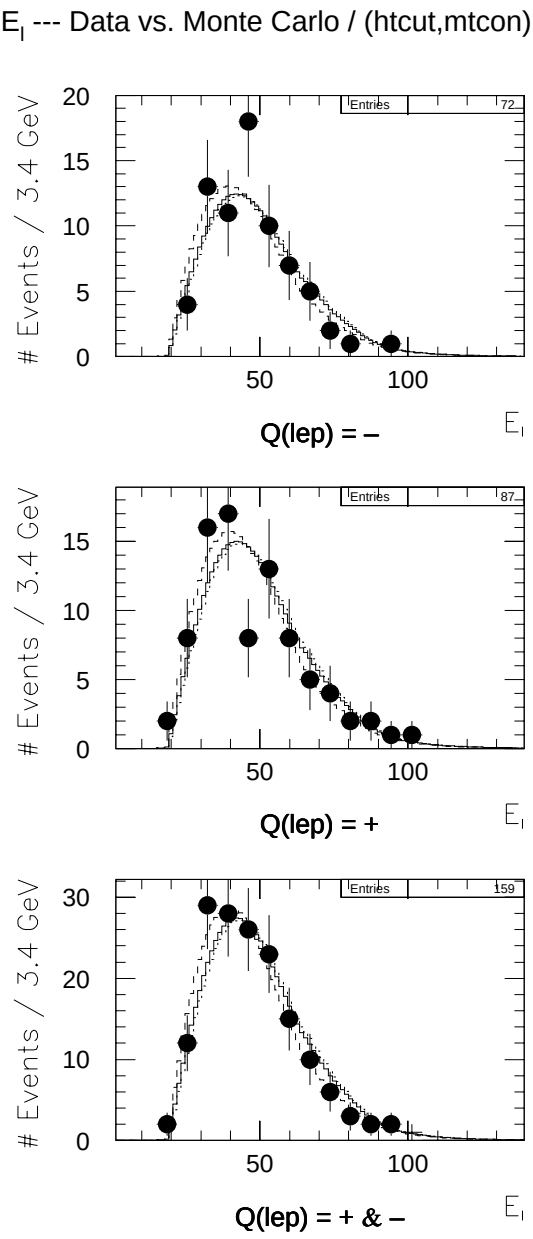


Figure 10.4: The distribution of E_ℓ. Points = experimental data; solid curve = monte carlo background + signal with $r_T = 25\%$; dashed curve = background + signal with $r_T = 100\%$; dotted curve = background + signal with $r_T = 0\%$. The two upper plots are for primary lepton charge $-$ and $+$; the bottom plot is a combination of the first two.

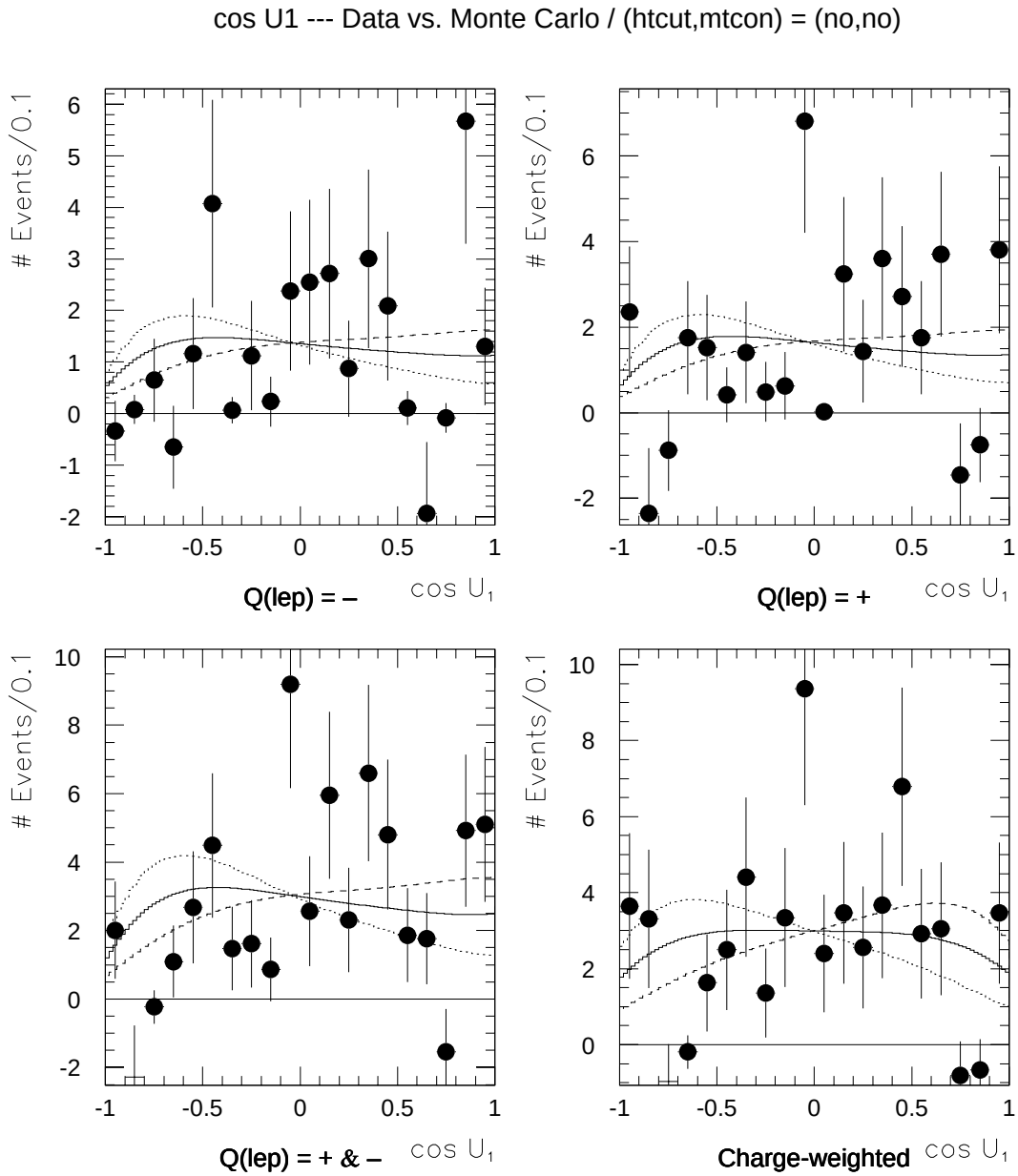


Figure 10.5: The background-subtracted distribution of $\cos U_1$.

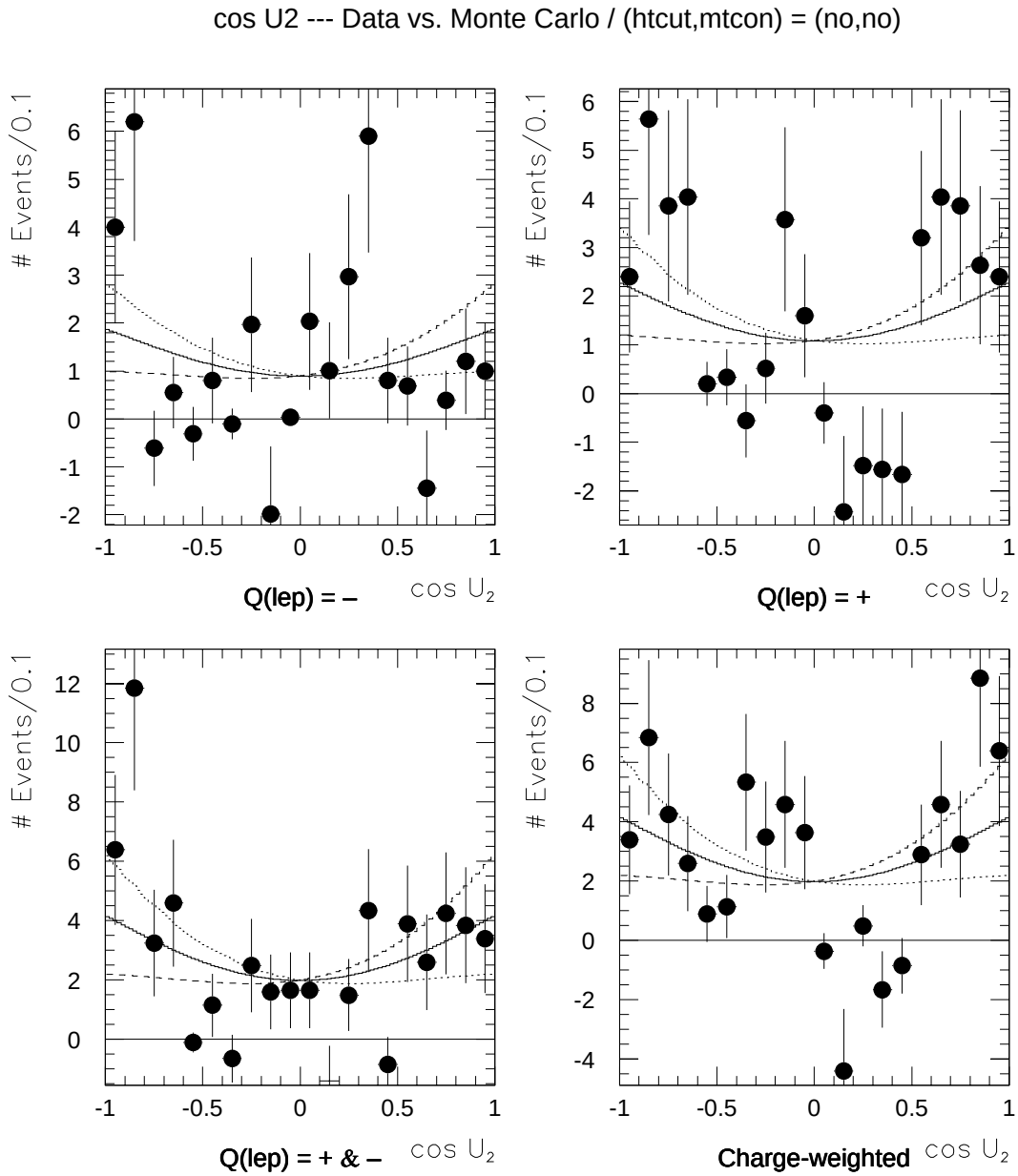


Figure 10.6: The background-subtracted distribution of $\cos U_2$.

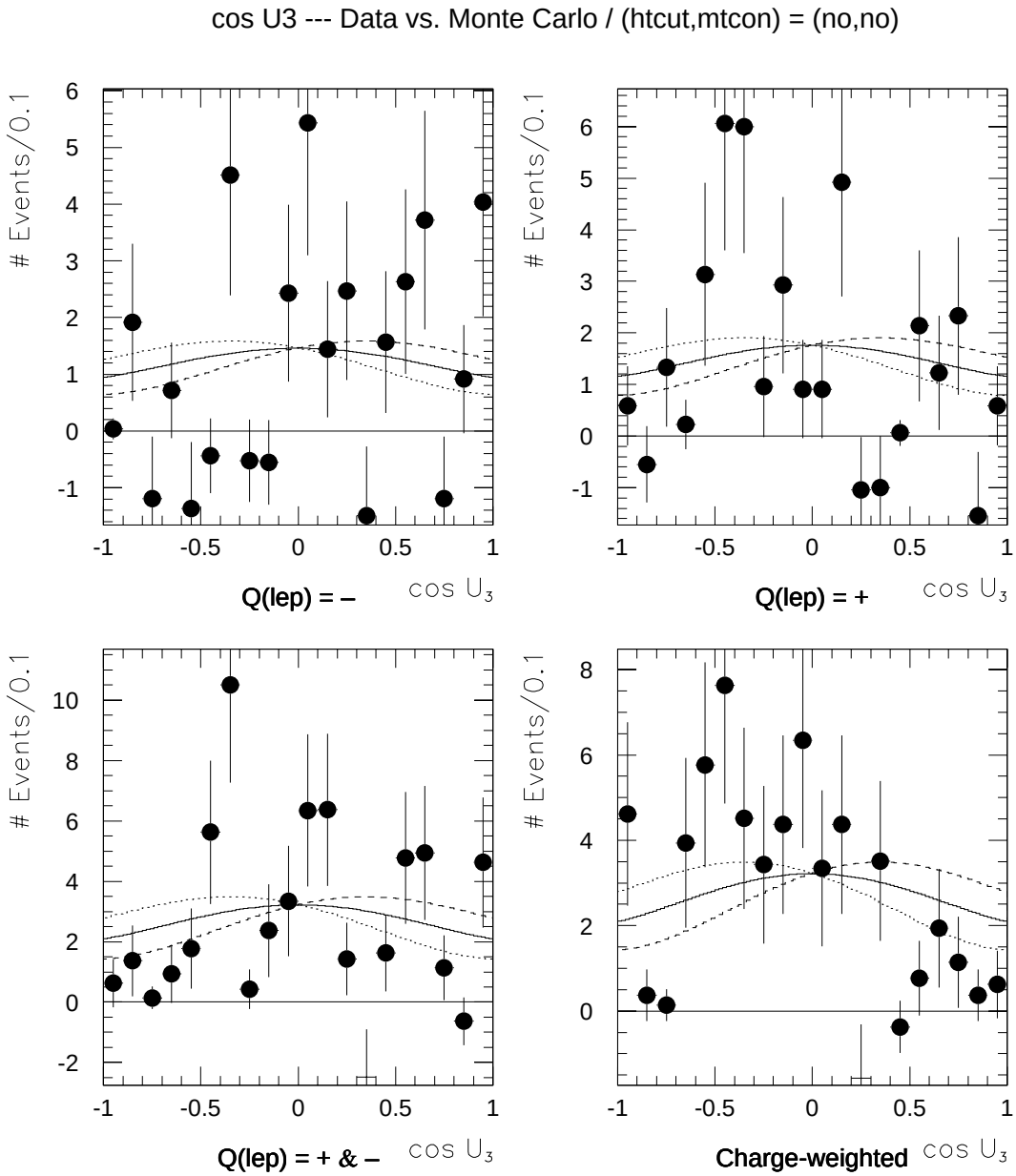


Figure 10.7: The background-subtracted distribution of $\cos U_3$.

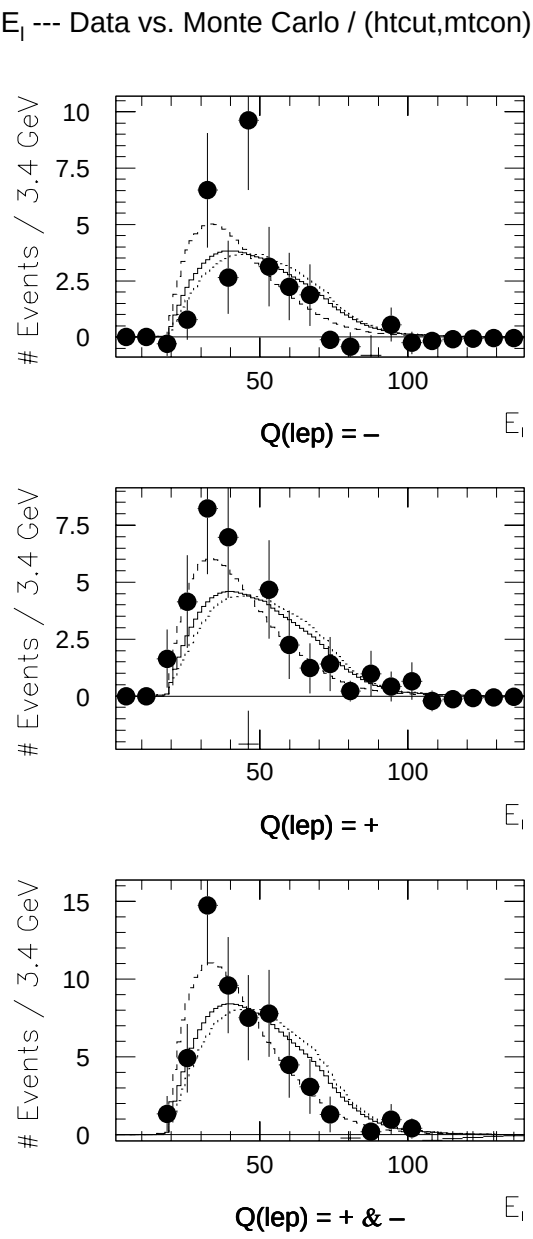


Figure 10.8: The background-subtracted distribution of E_ℓ.

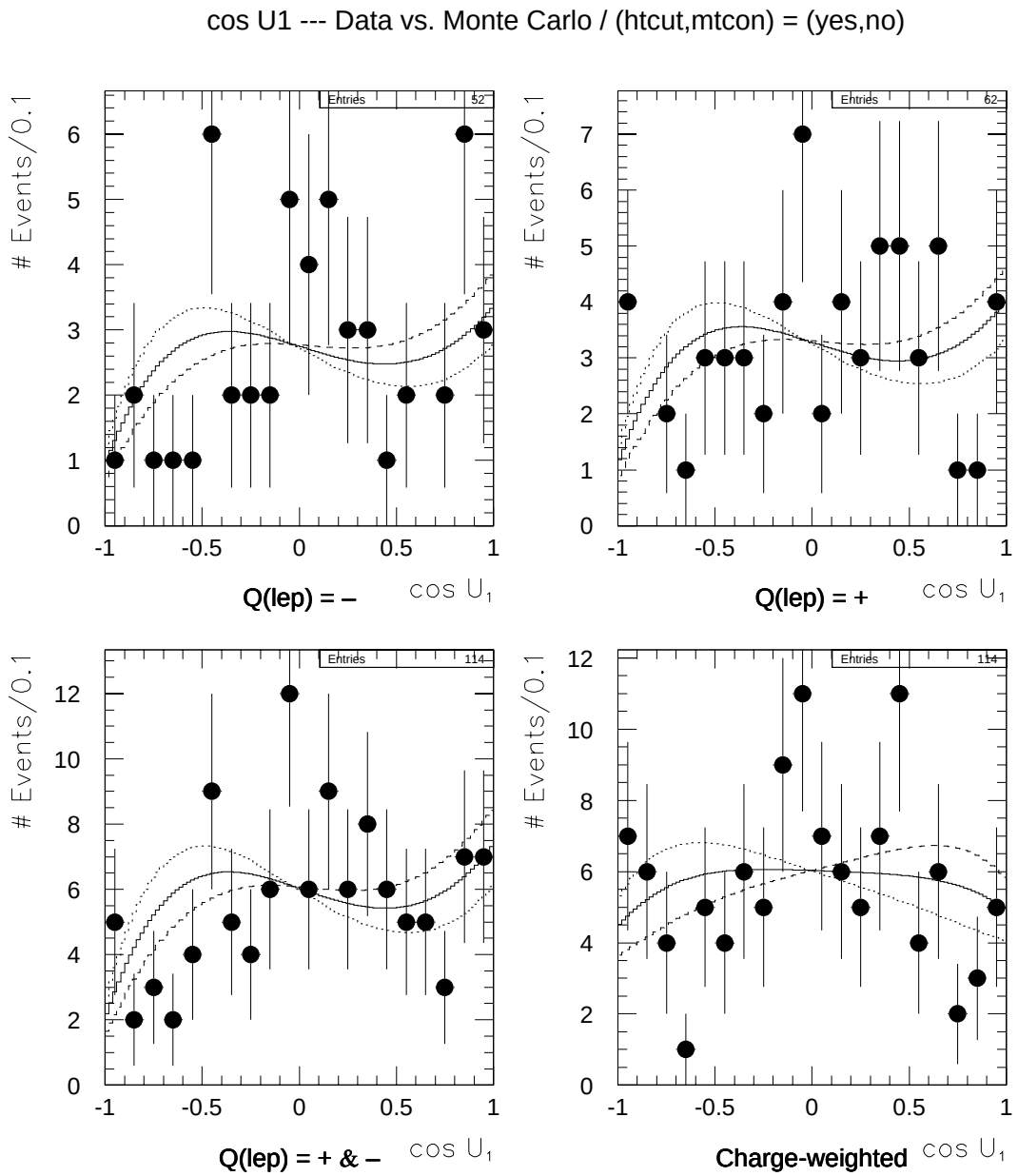


Figure 10.9: Same as figure 10.1, but with optimization (HTCUT, MTCON) = (YES, NO).

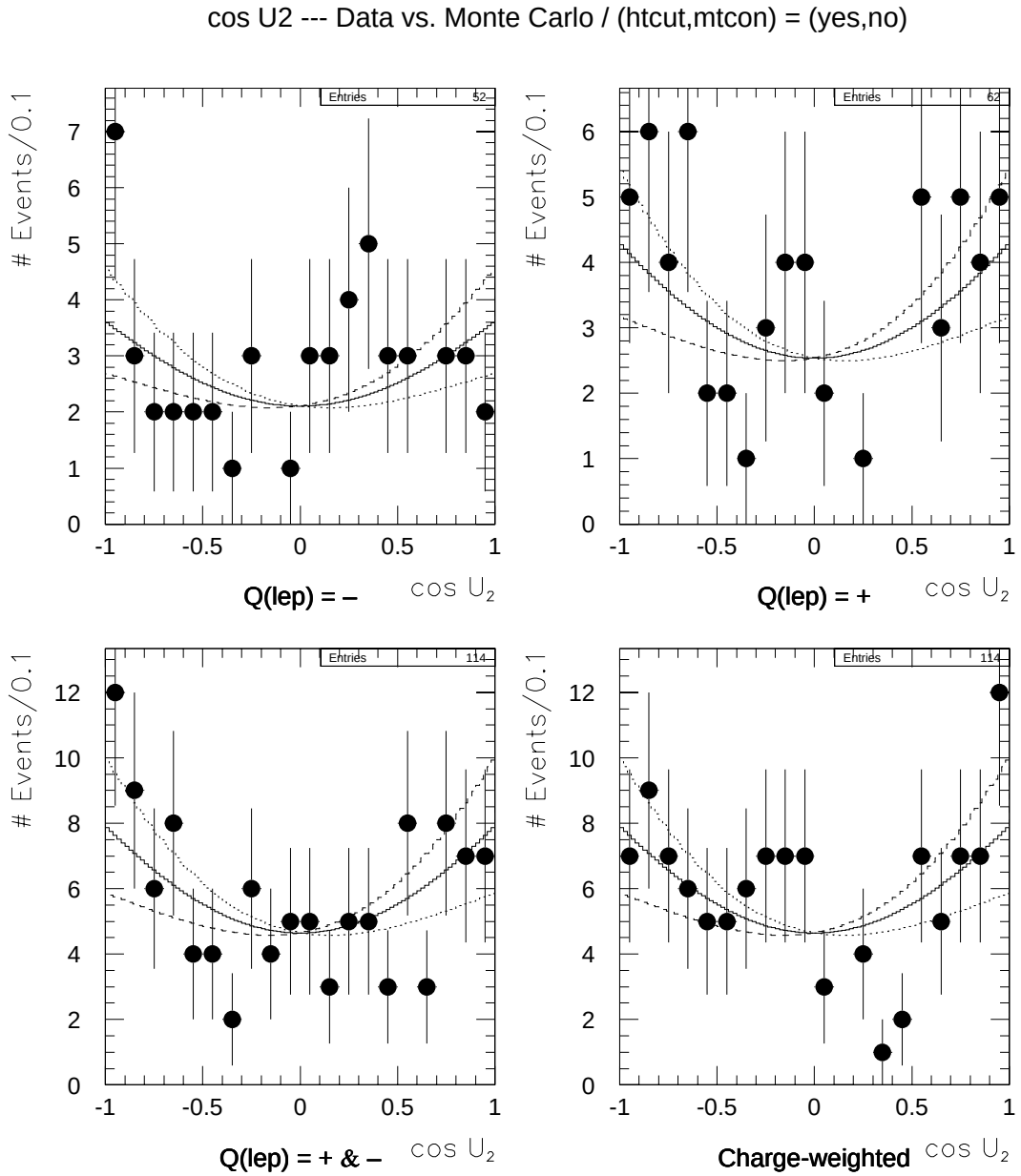


Figure 10.10: Same as figure 10.2, but with optimization (HTCUT, MTCON) = (YES, NO).

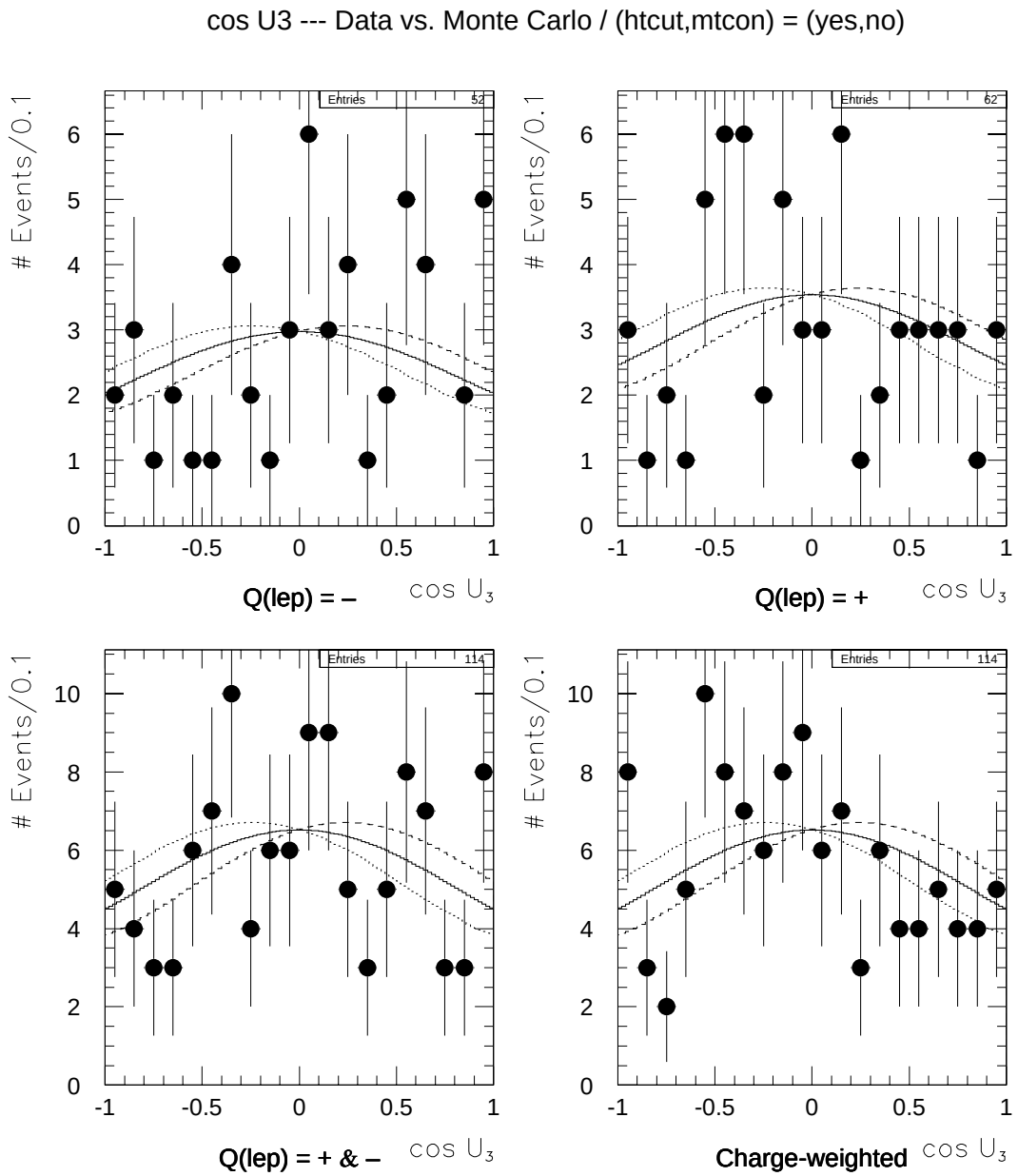


Figure 10.11: Same as figure 10.3, but with optimization (HTCUT, MTCON) = (YES, NO).

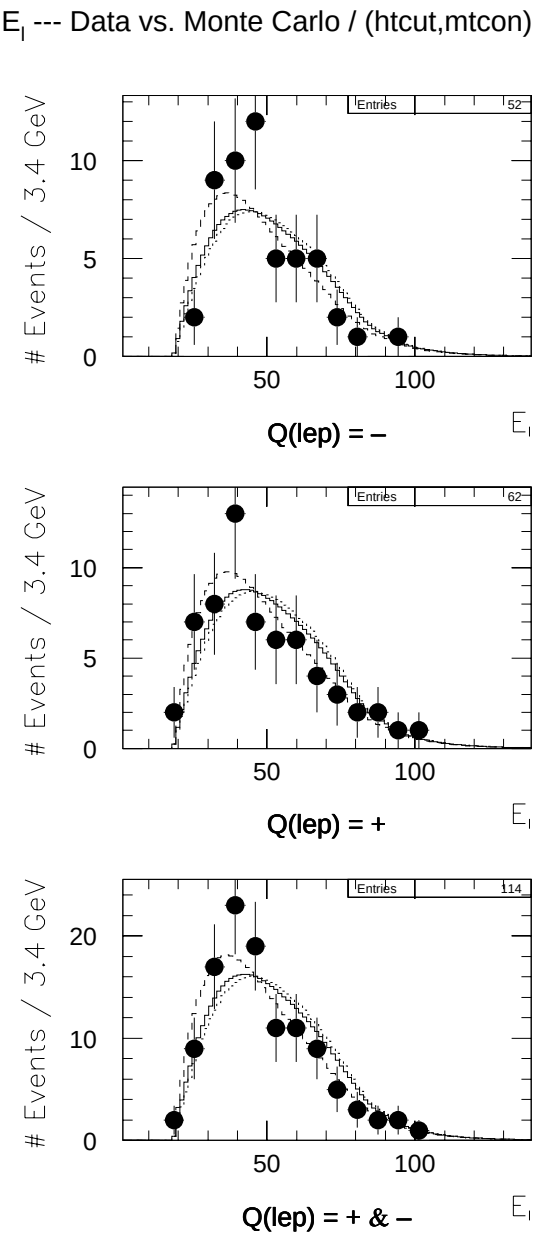


Figure 10.12: Same as figure 10.4, but with optimization (HTCUT, MTCON) = (YES, NO).

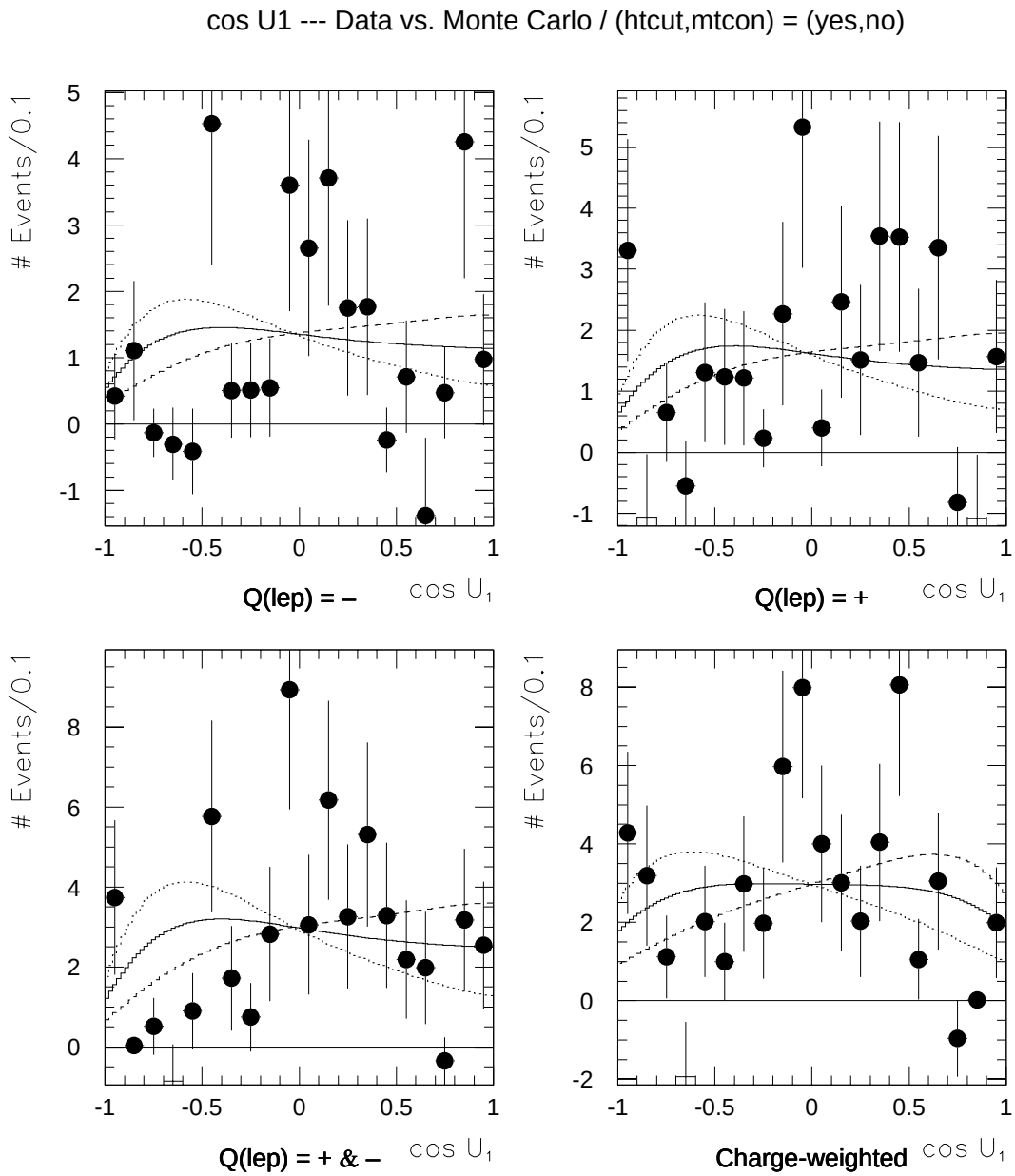


Figure 10.13: Same as figure 10.5, but with optimization (HTCUT, MTCON) = (YES, NO).

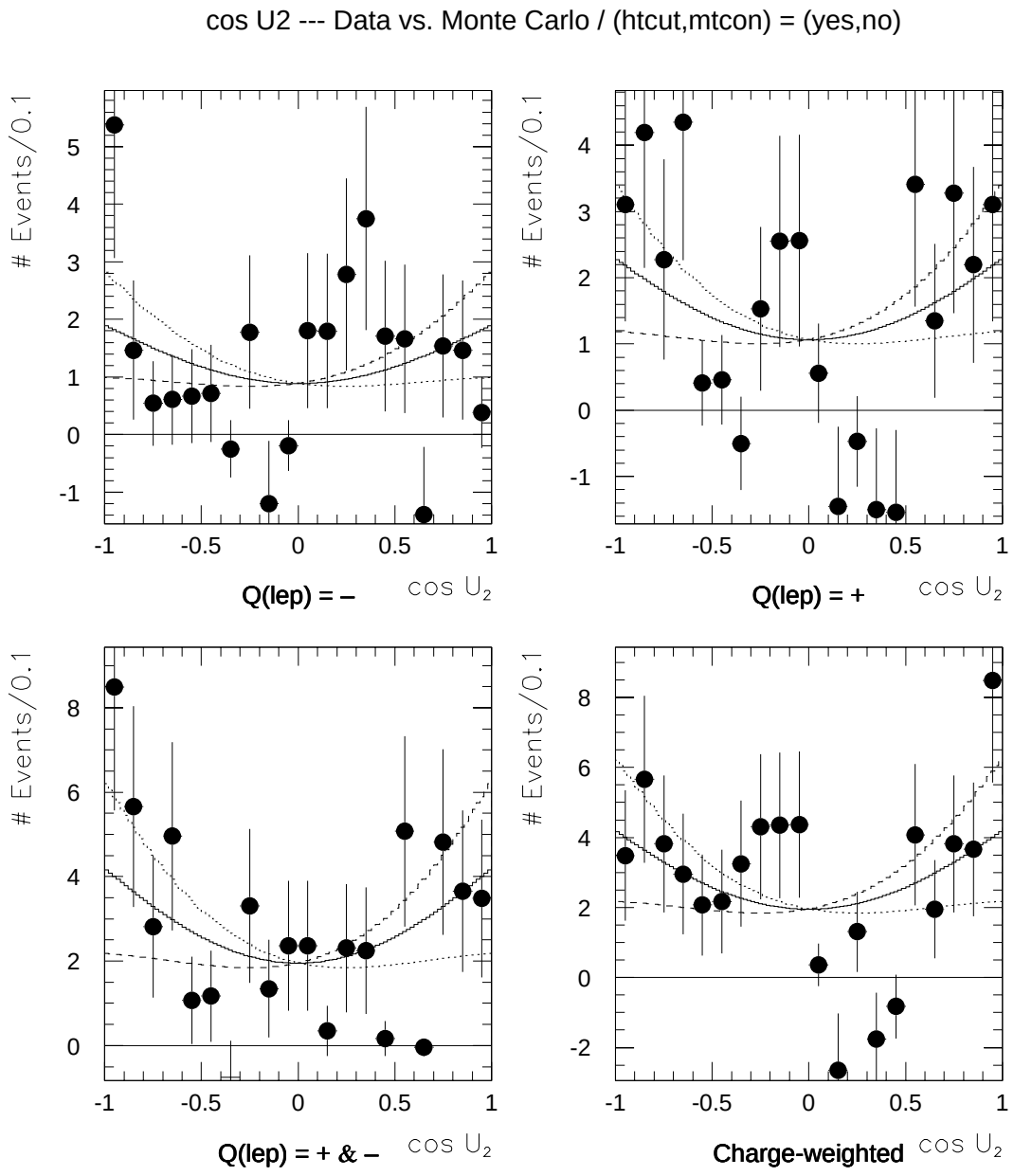


Figure 10.14: Same as figure 10.6, but with optimization (HTCUT, MTCON) = (YES, NO).

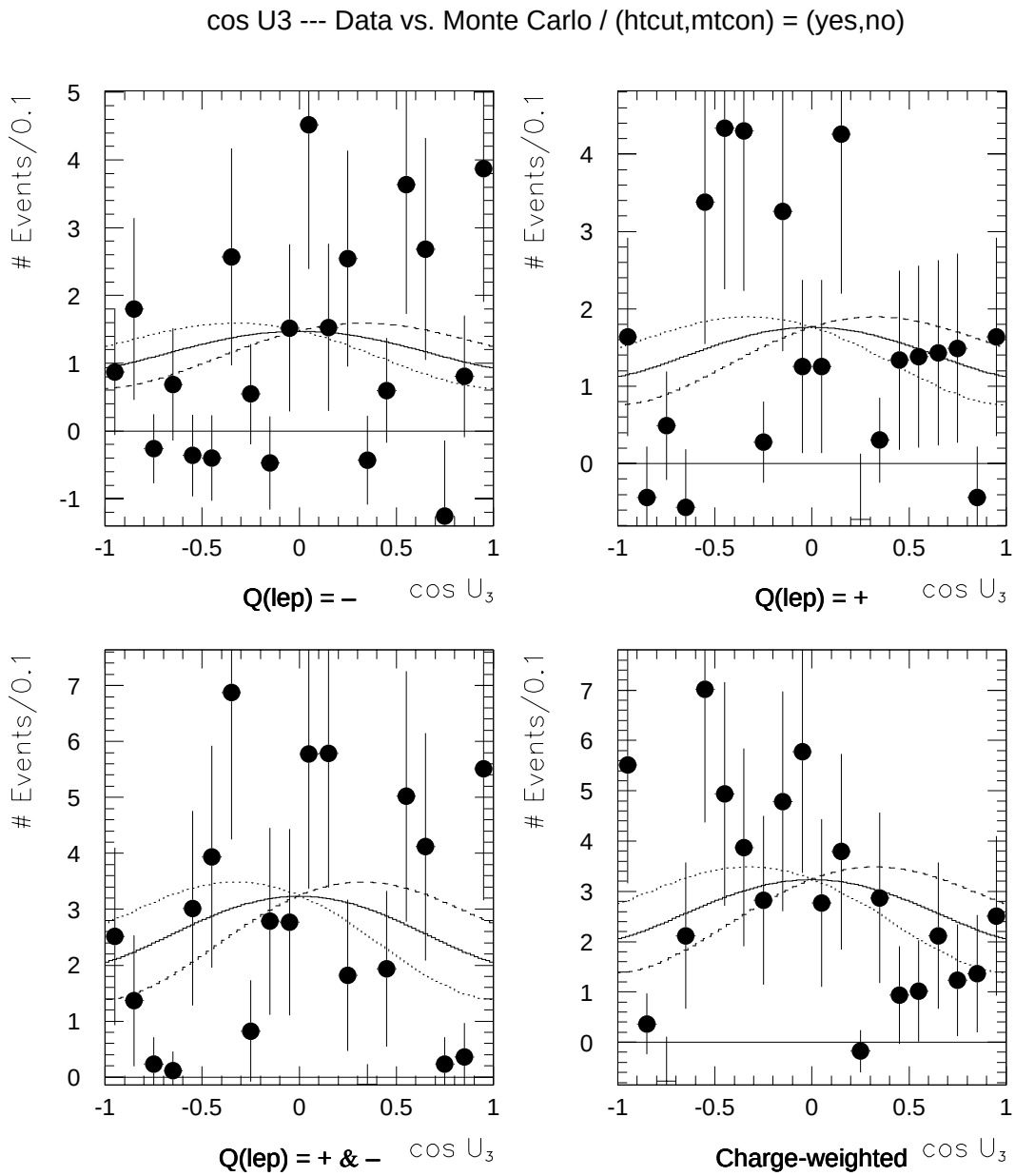


Figure 10.15: Same as figure 10.7, but with optimization (HTCUT, MTCON) = (YES, NO).

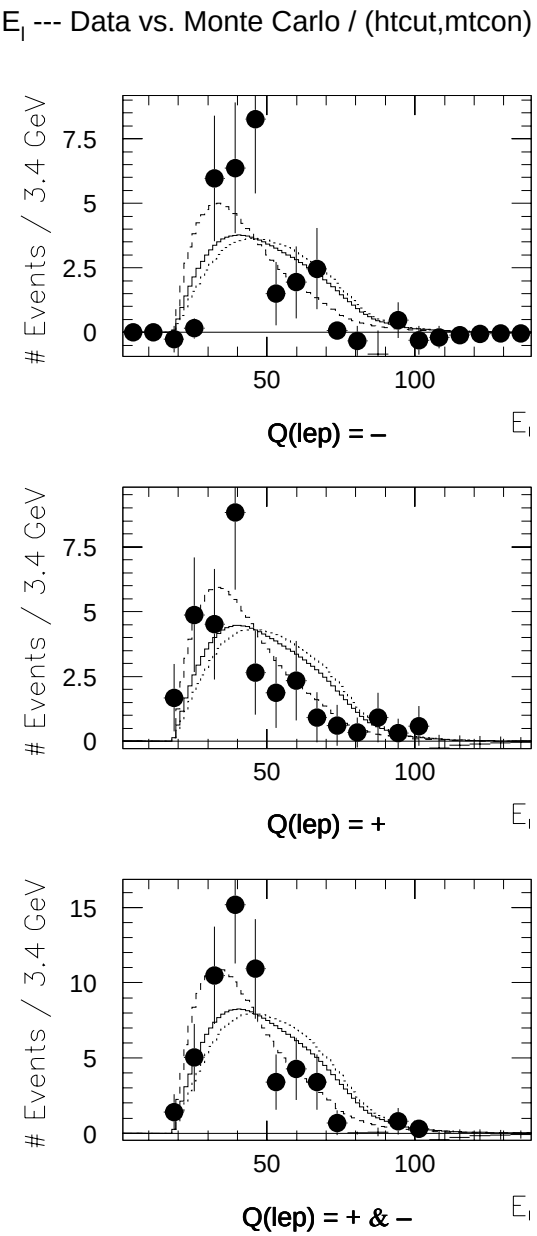


Figure 10.16: Same as figure 10.8, but with optimization (HTCUT, MTCON) = (YES, NO).

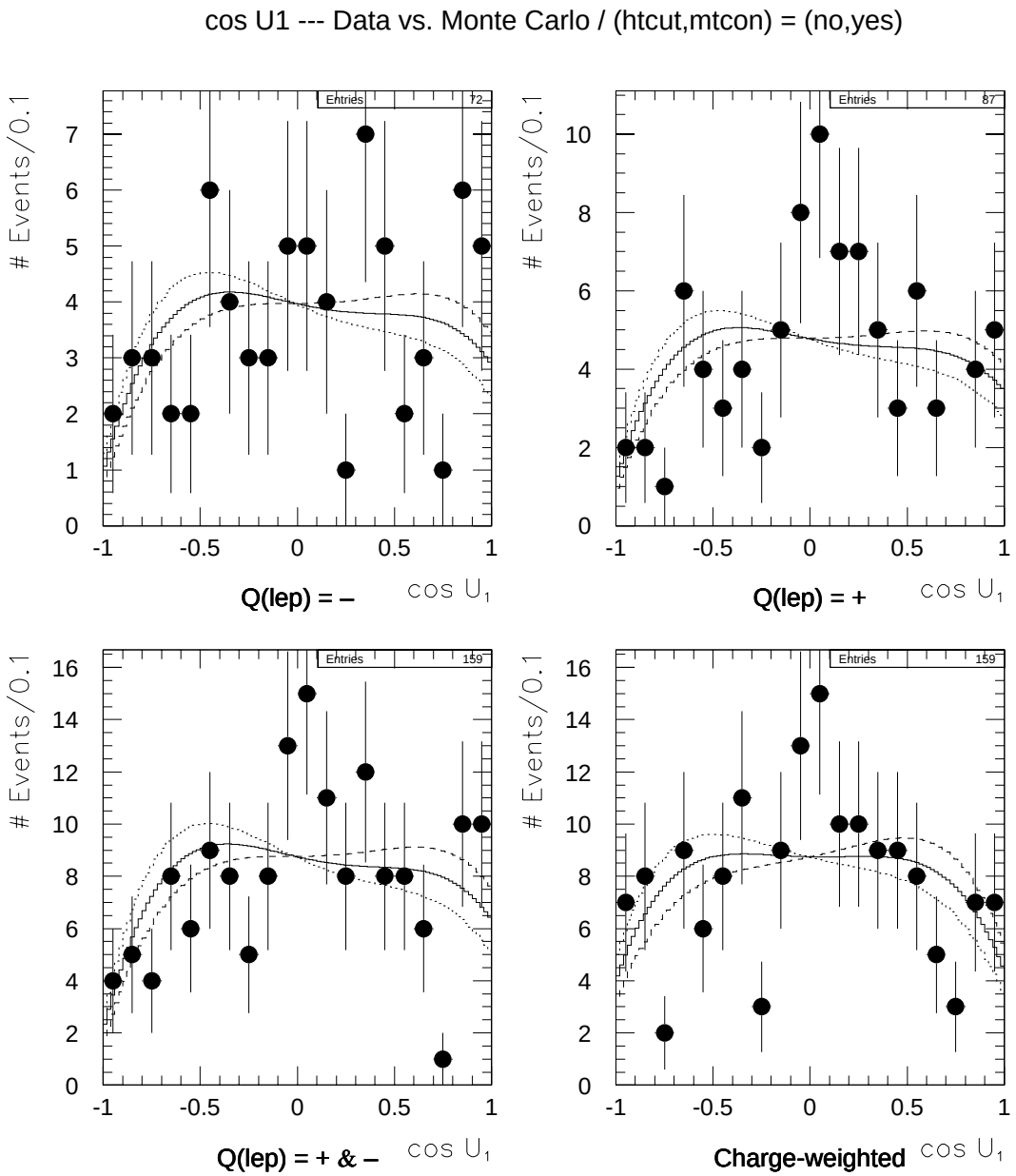


Figure 10.17: Same as figure 10.1, but with optimization (HTCUT, MTCON) = (NO, YES).

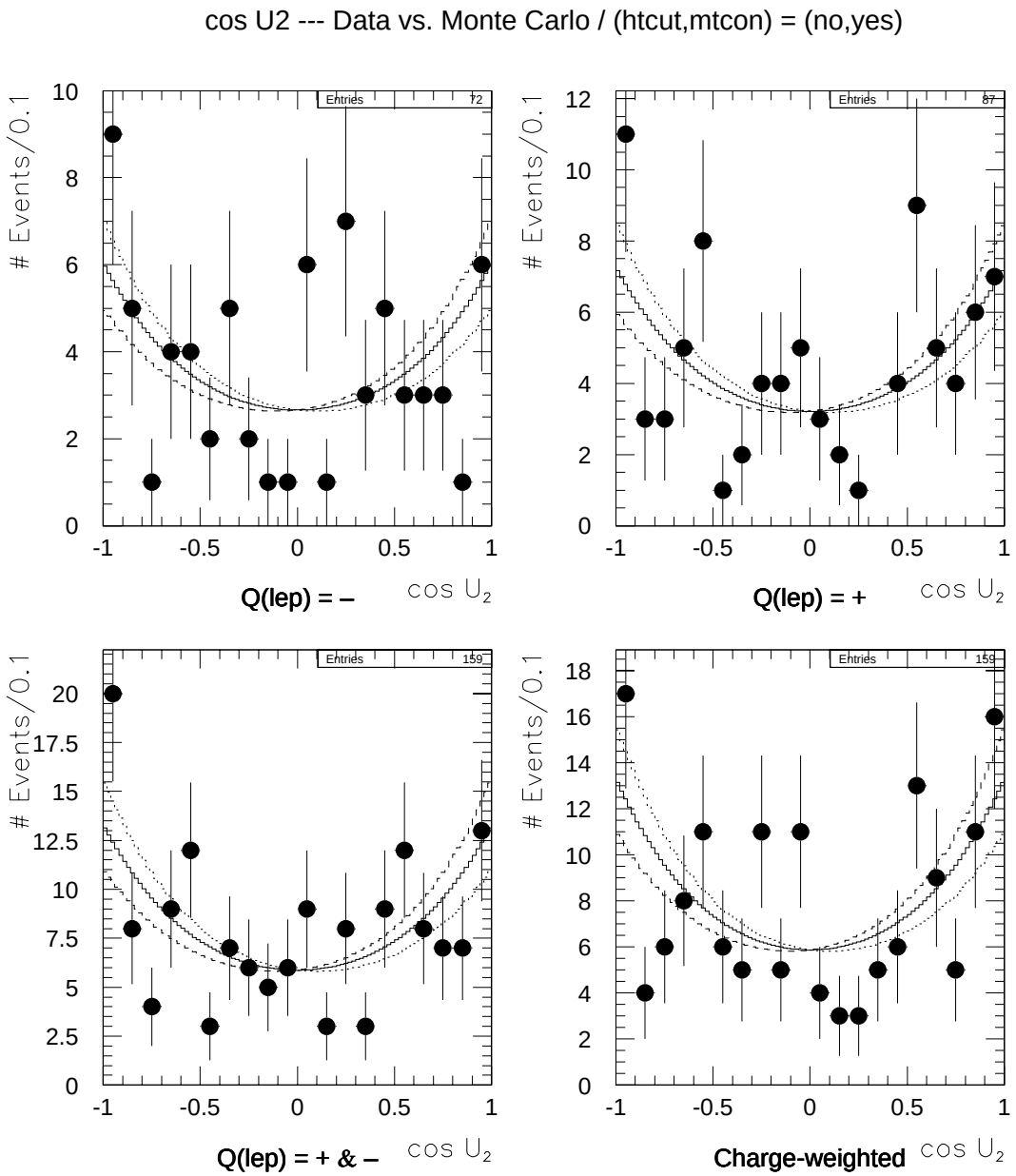


Figure 10.18: Same as figure 10.2, but with optimization (HTCUT, MTCON) = (NO, YES).

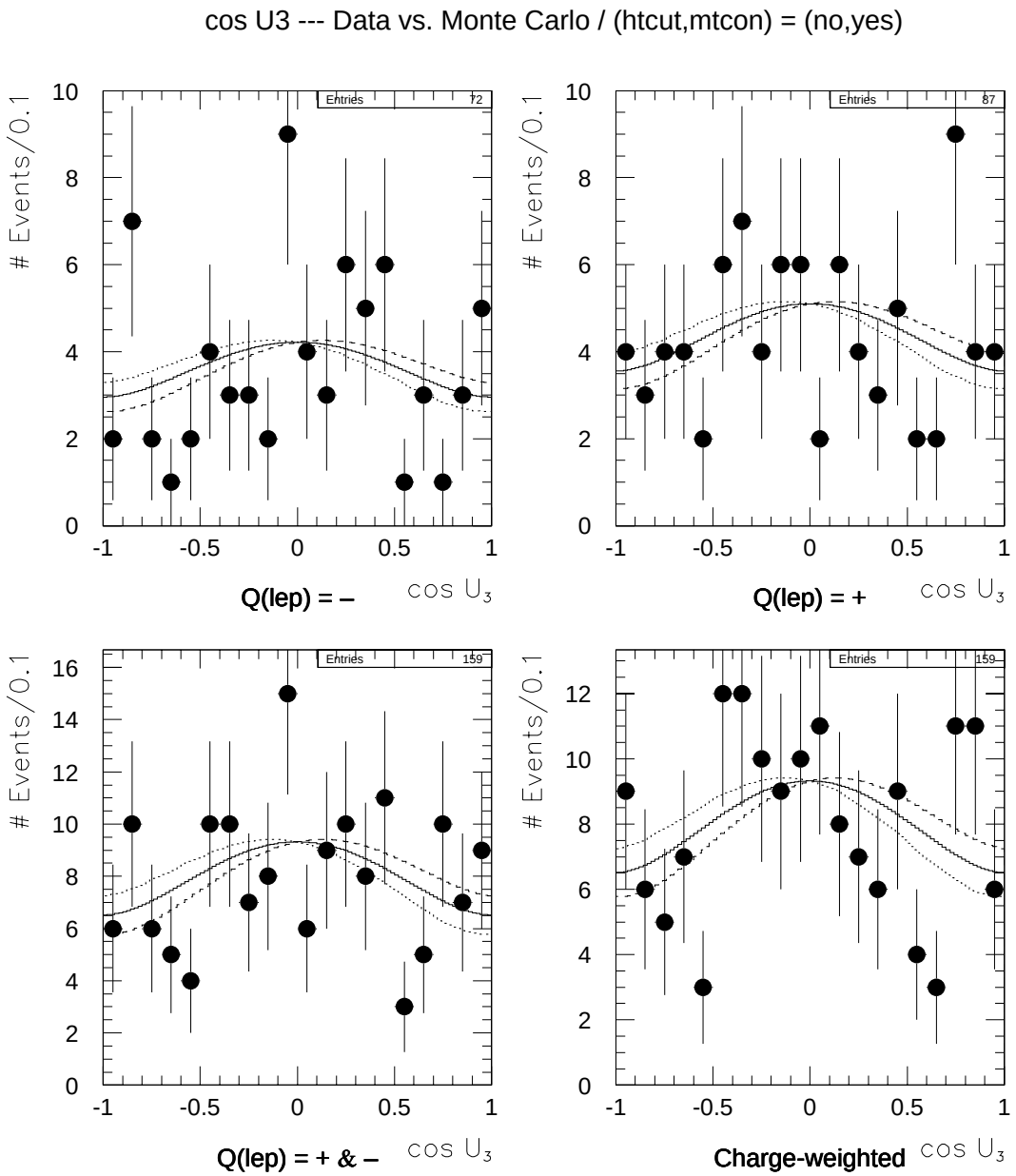


Figure 10.19: Same as figure 10.3, but with optimization (HTCUT, MTCON) = (NO, YES).

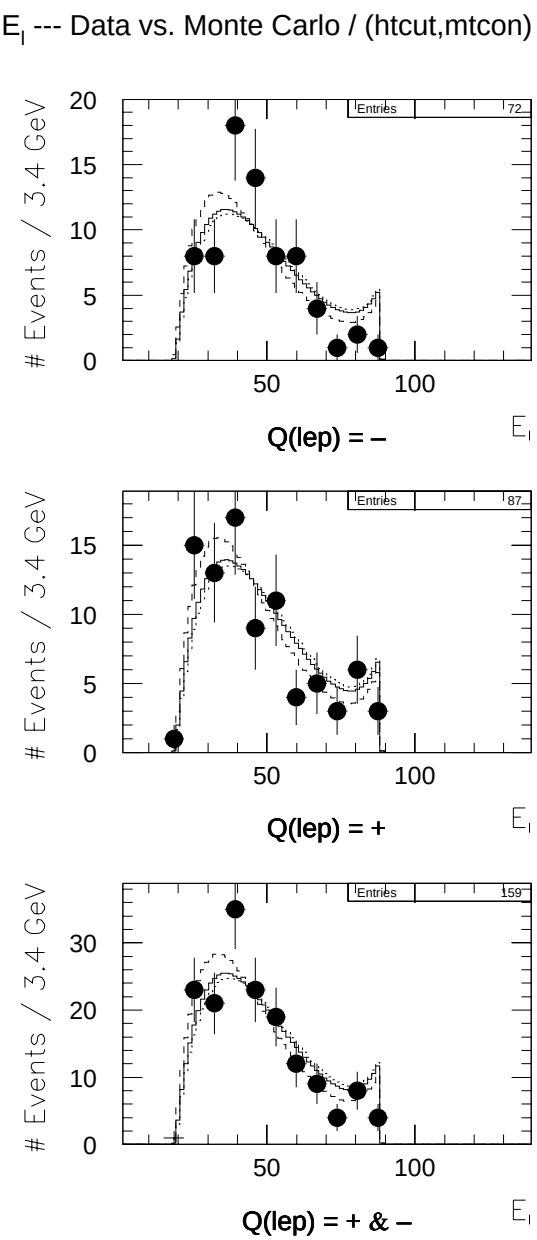


Figure 10.20: Same as figure 10.4, but with optimization (HTCUT, MTCON) = (NO, YES).

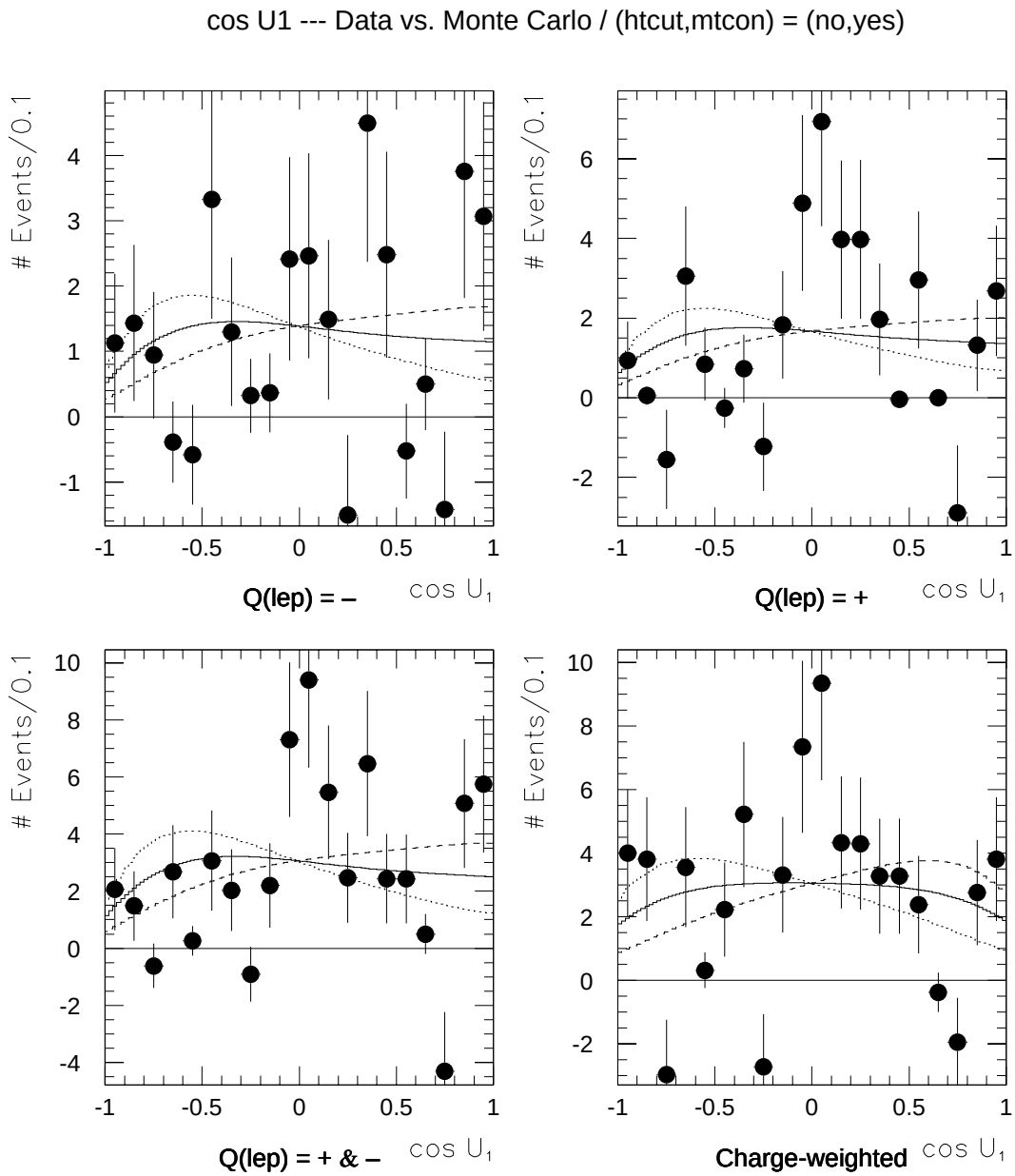


Figure 10.21: Same as figure 10.5, but with optimization (HTCUT, MTCON) = (NO, YES).

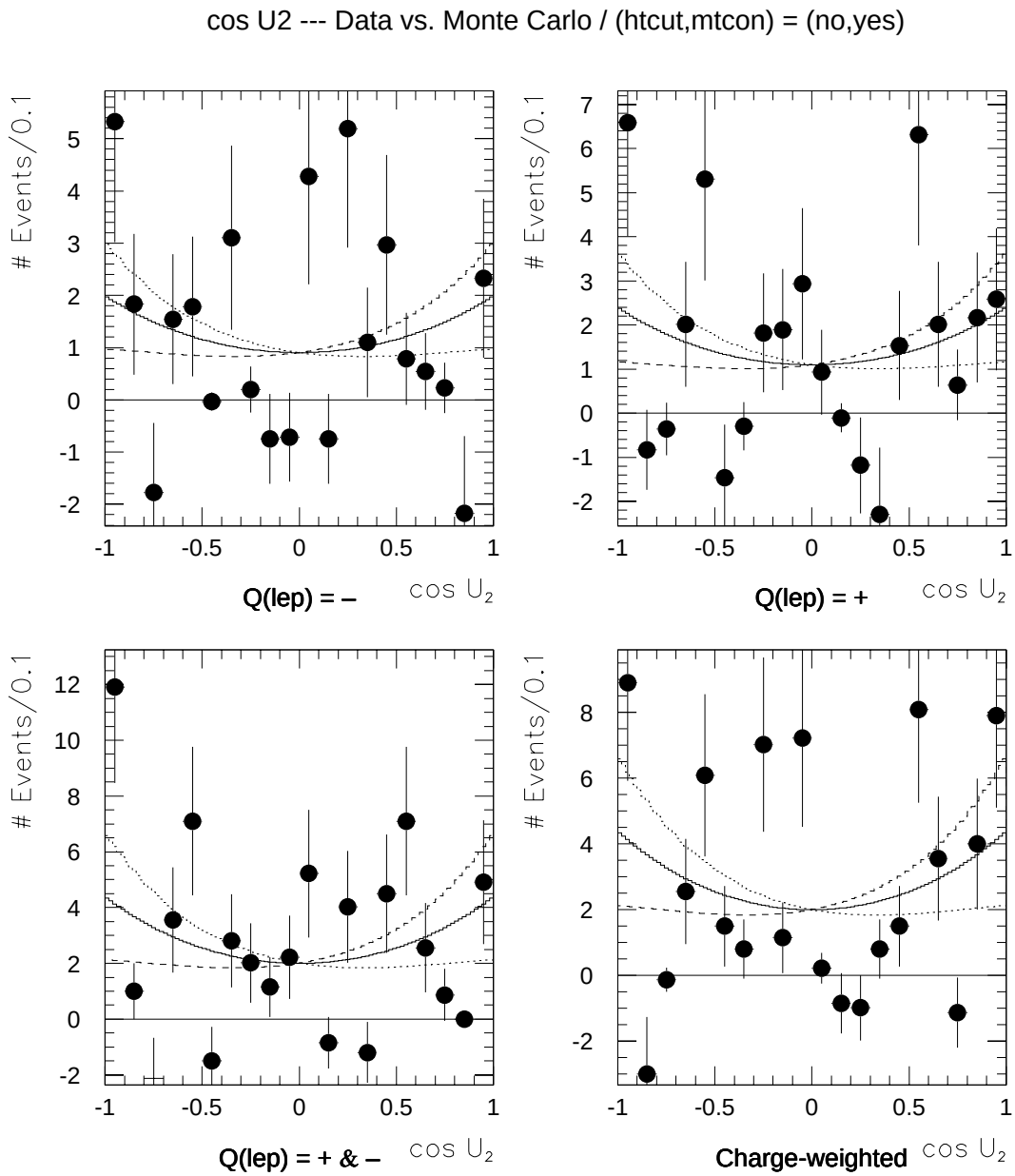


Figure 10.22: Same as figure 10.6, but with optimization (HTCUT, MTCON) = (NO, YES).

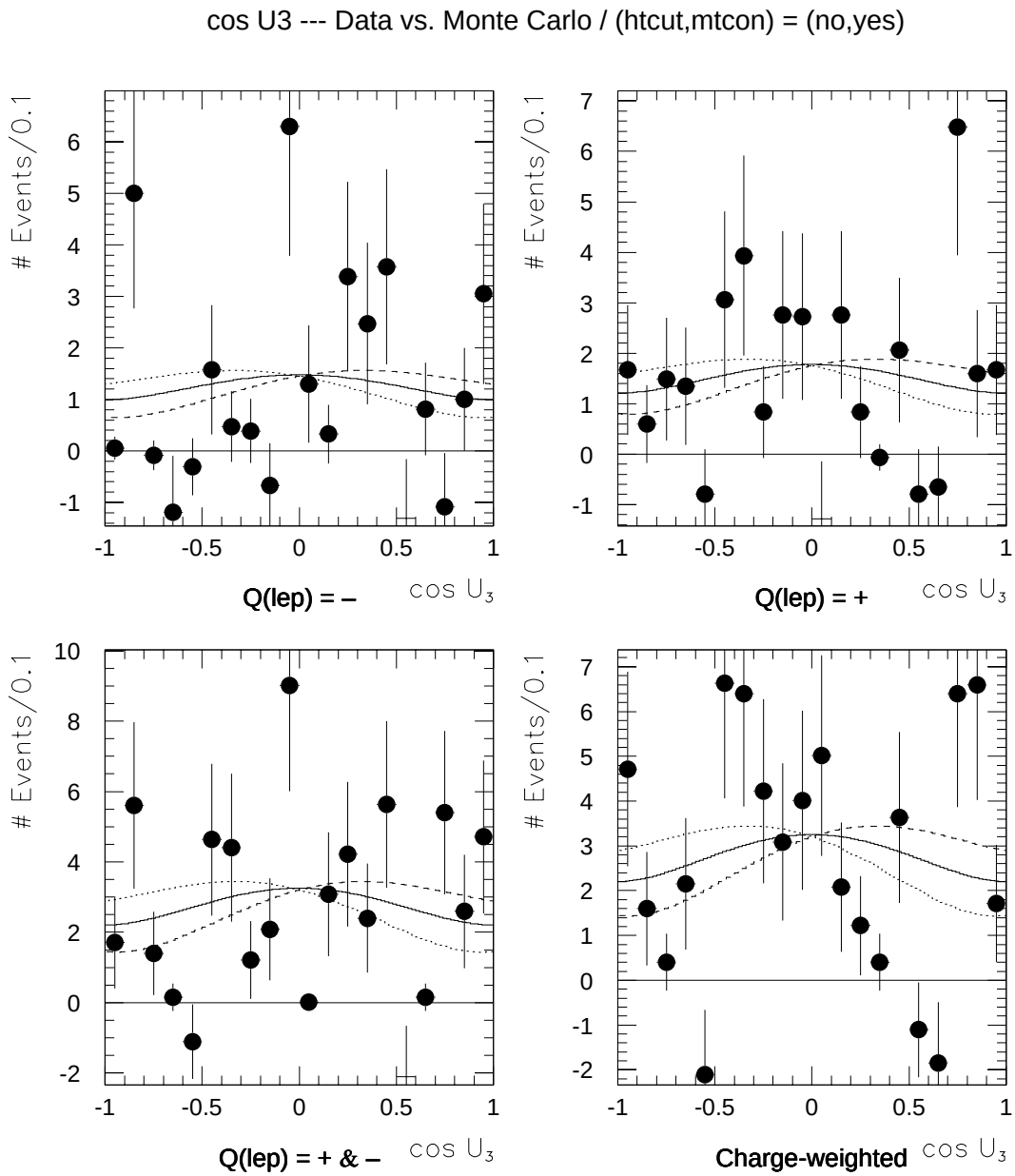


Figure 10.23: Same as figure 10.7, but with optimization (HTCUT, MTCON) = (NO, YES).

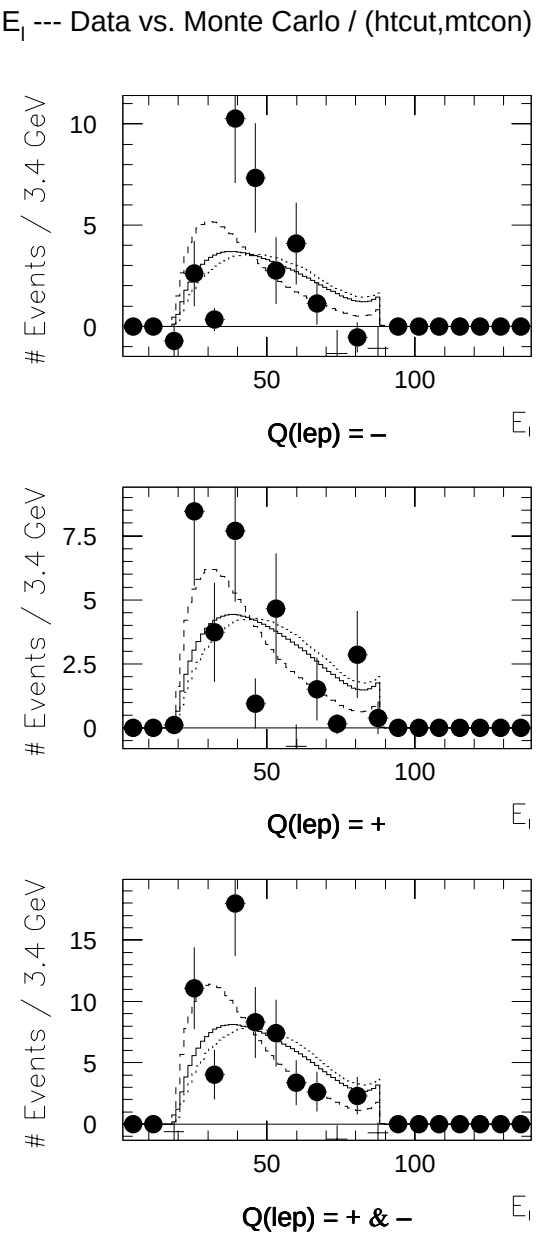


Figure 10.24: Same as figure 10.8, but with optimization (HTCUT, MTCON) = (NO, YES).

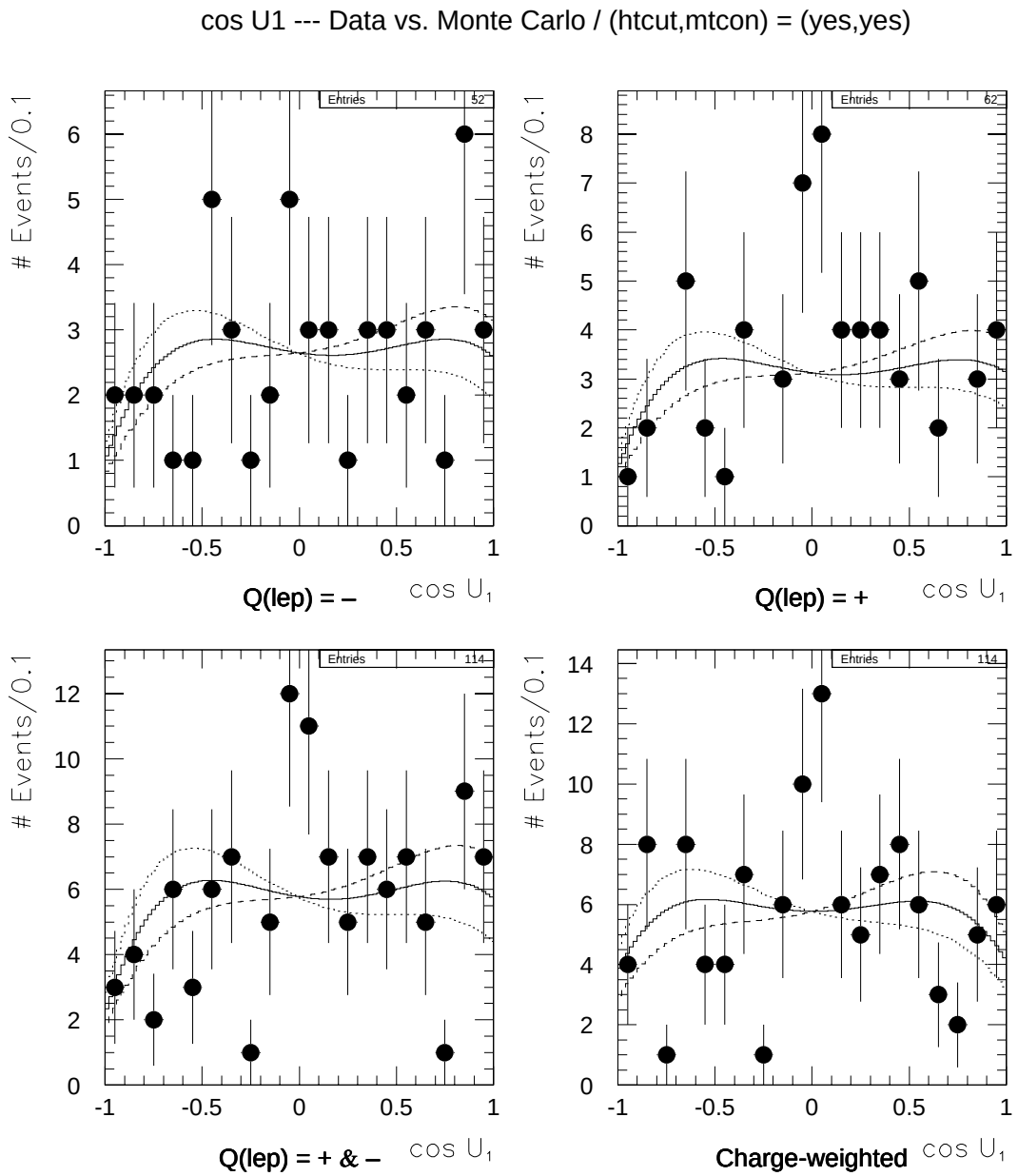


Figure 10.25: Same as figure 10.1, but with optimization (HTCUT, MTCON) = (YES, YES).

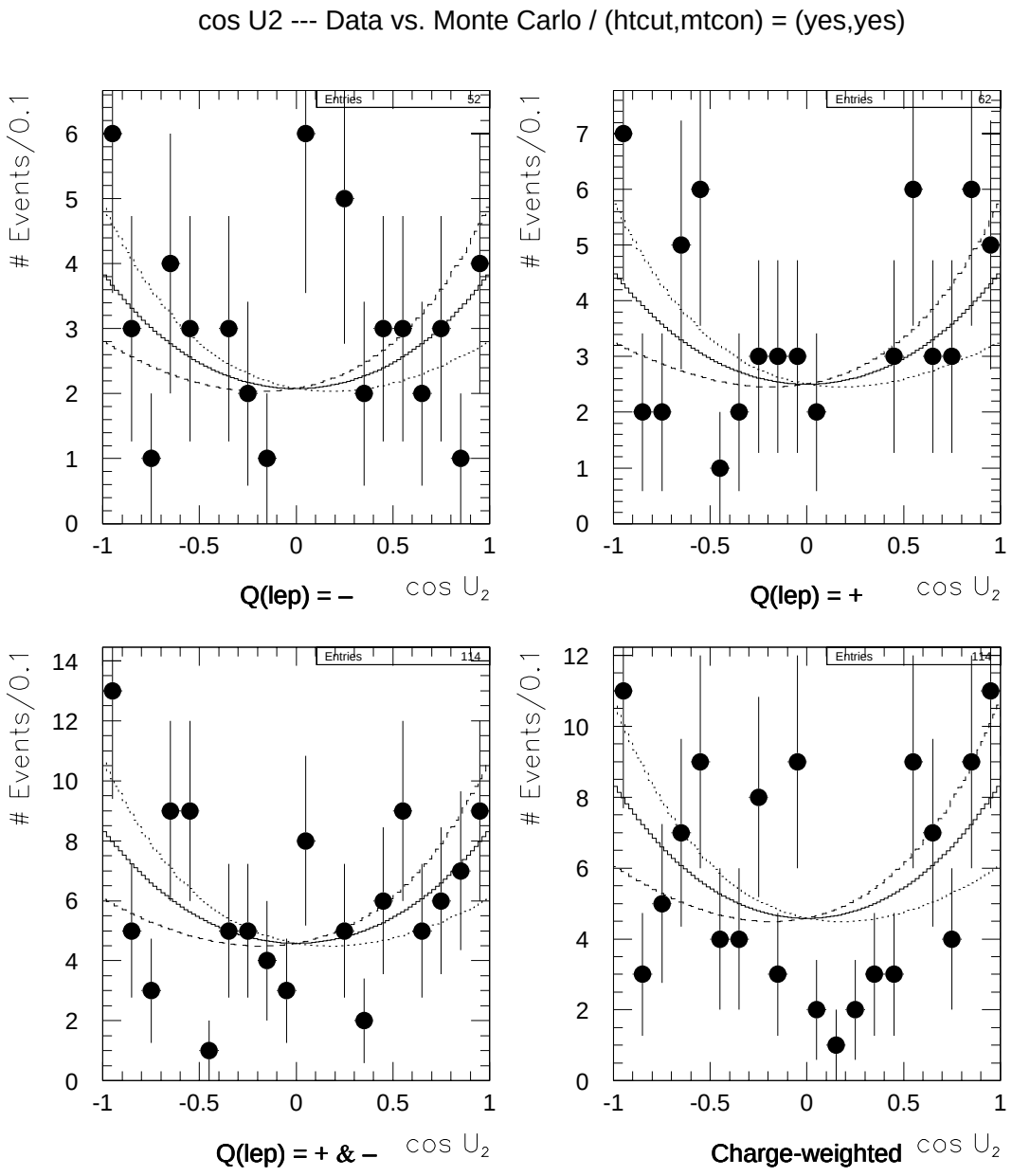


Figure 10.26: Same as figure 10.2, but with optimization (HTCUT, MTCON) = (YES, YES).

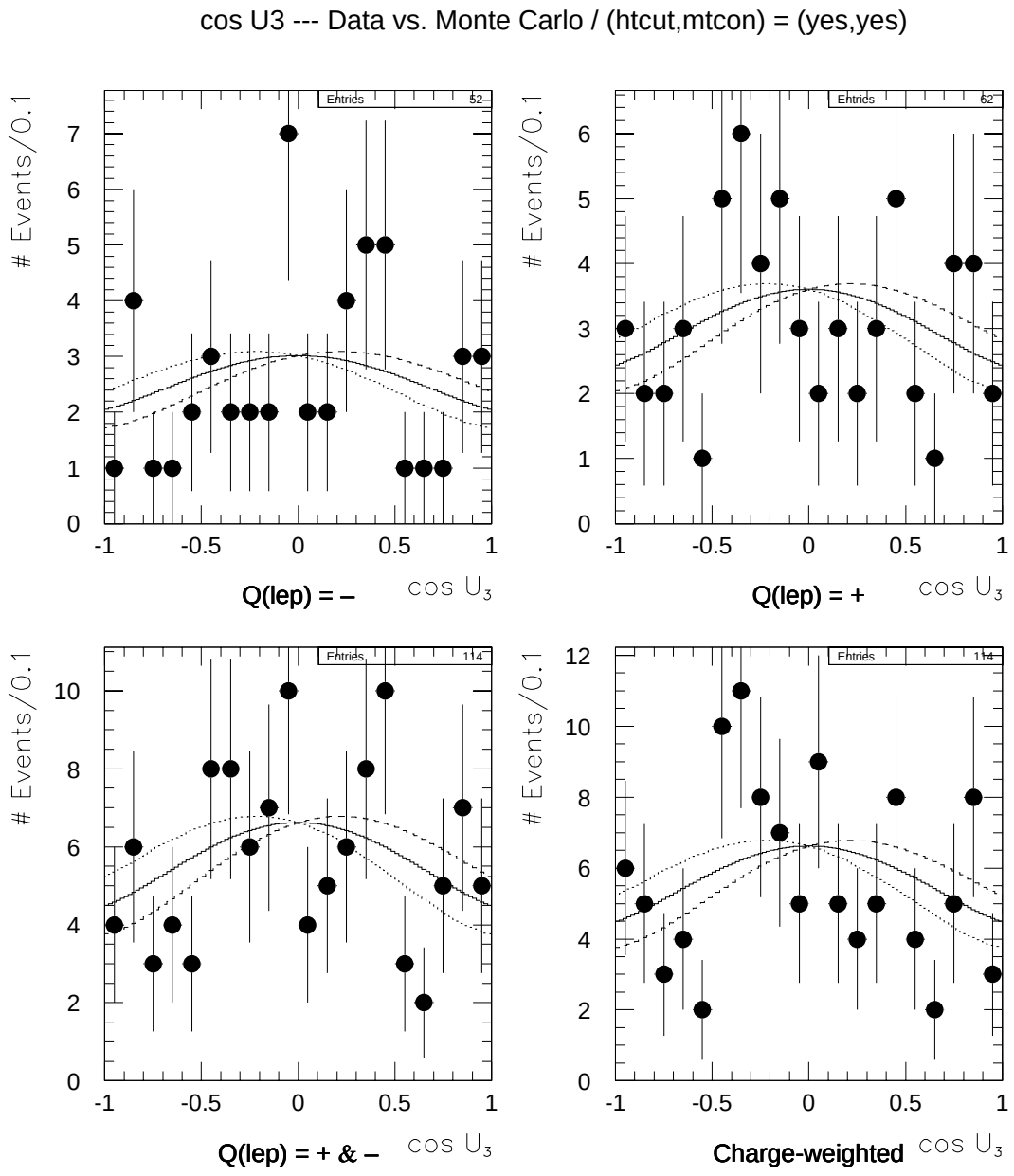


Figure 10.27: Same as figure 10.3, but with optimization (HTCUT, MTCON) = (YES, YES).

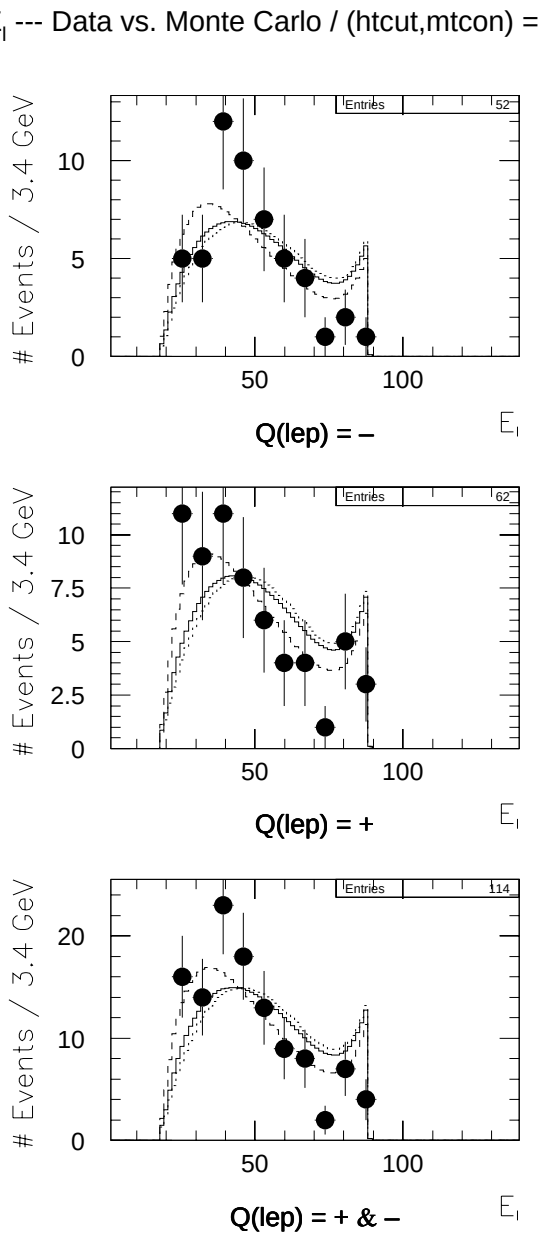


Figure 10.28: Same as figure 10.4, but with optimization (HTCUT, MTCON) = (YES, YES).

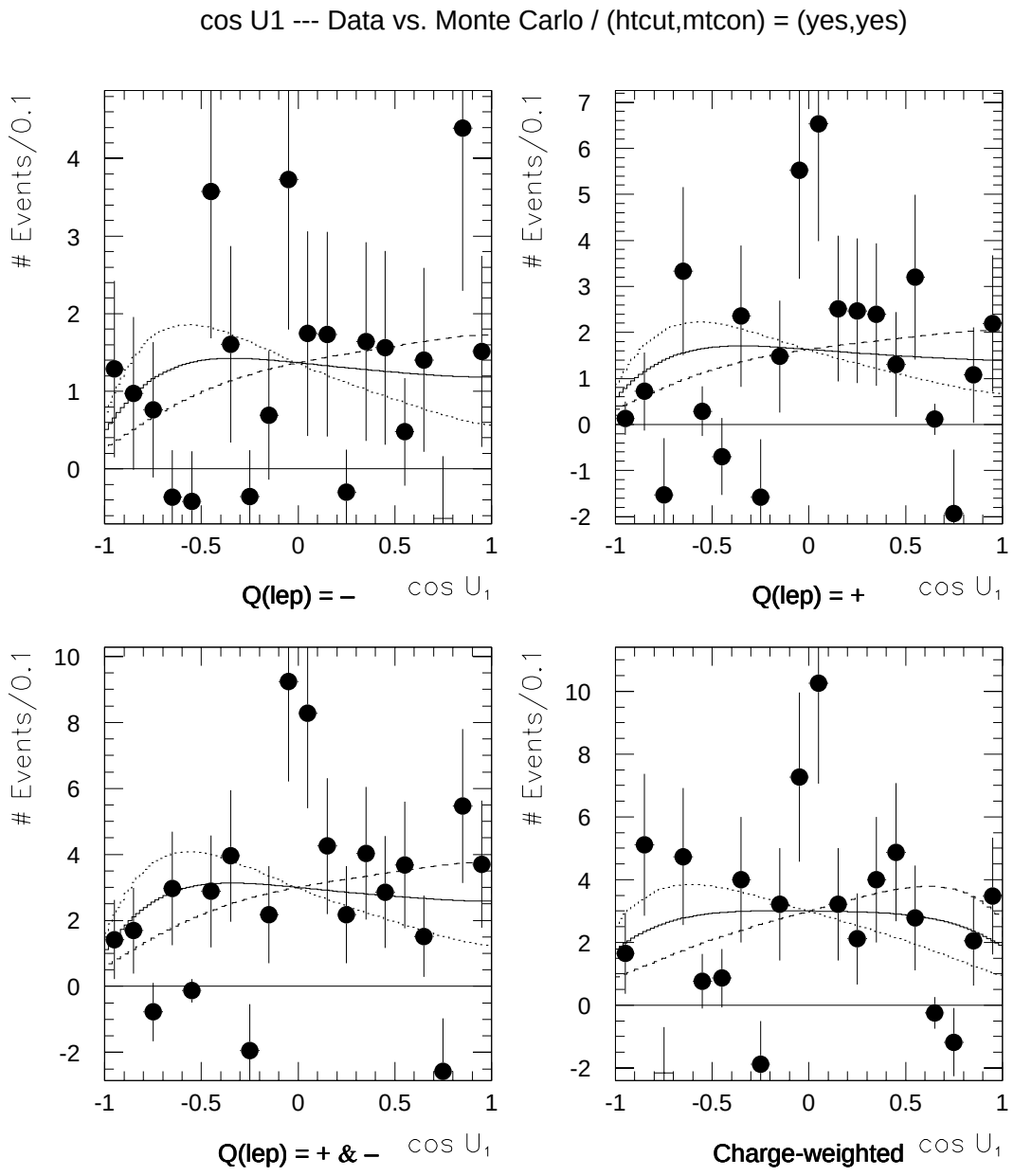


Figure 10.29: Same as figure 10.5, but with optimization (HTCUT, MTCON) = (YES, YES).

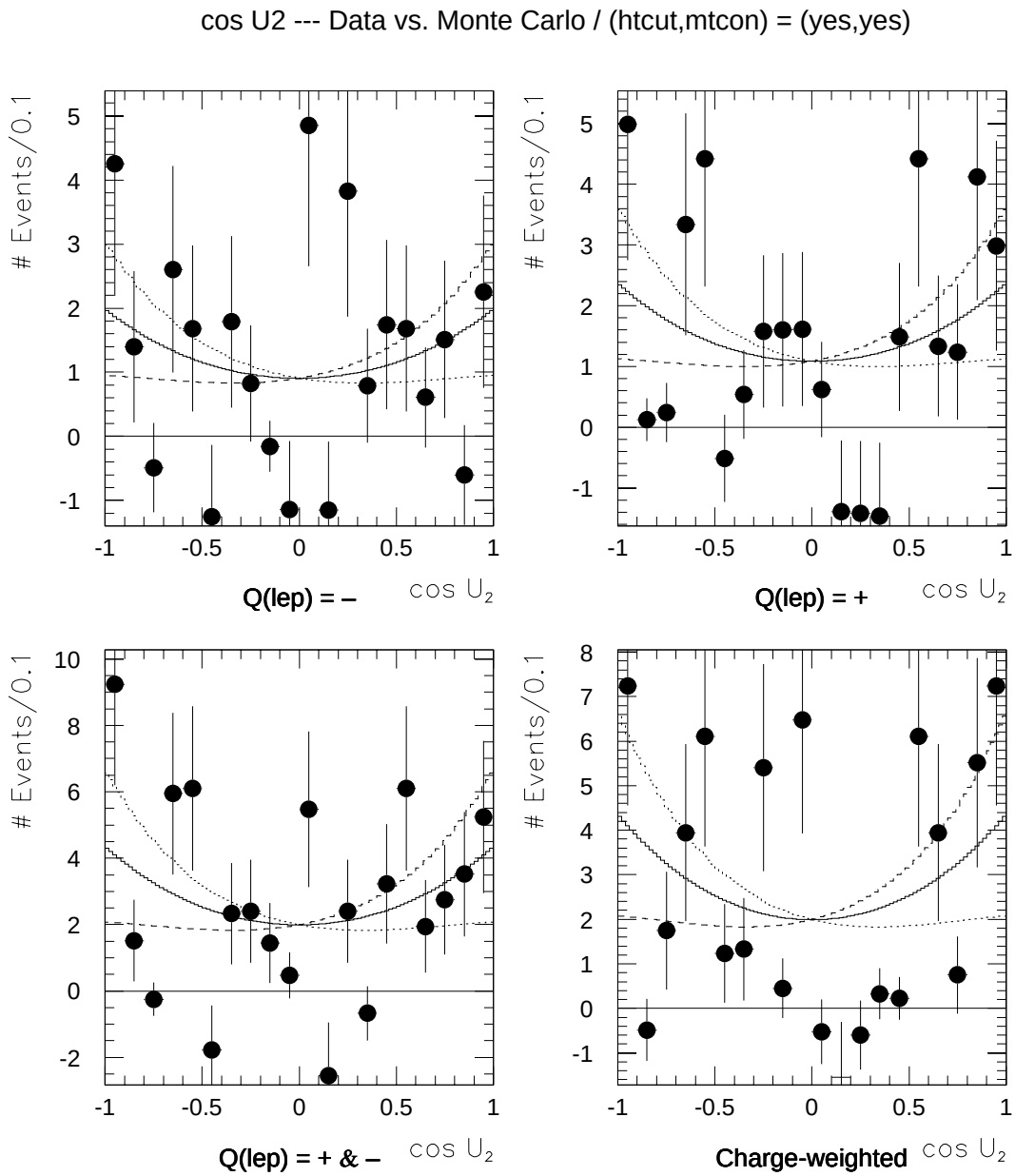


Figure 10.30: Same as figure 10.6, but with optimization (HTCUT, MTCON) = (YES, YES).

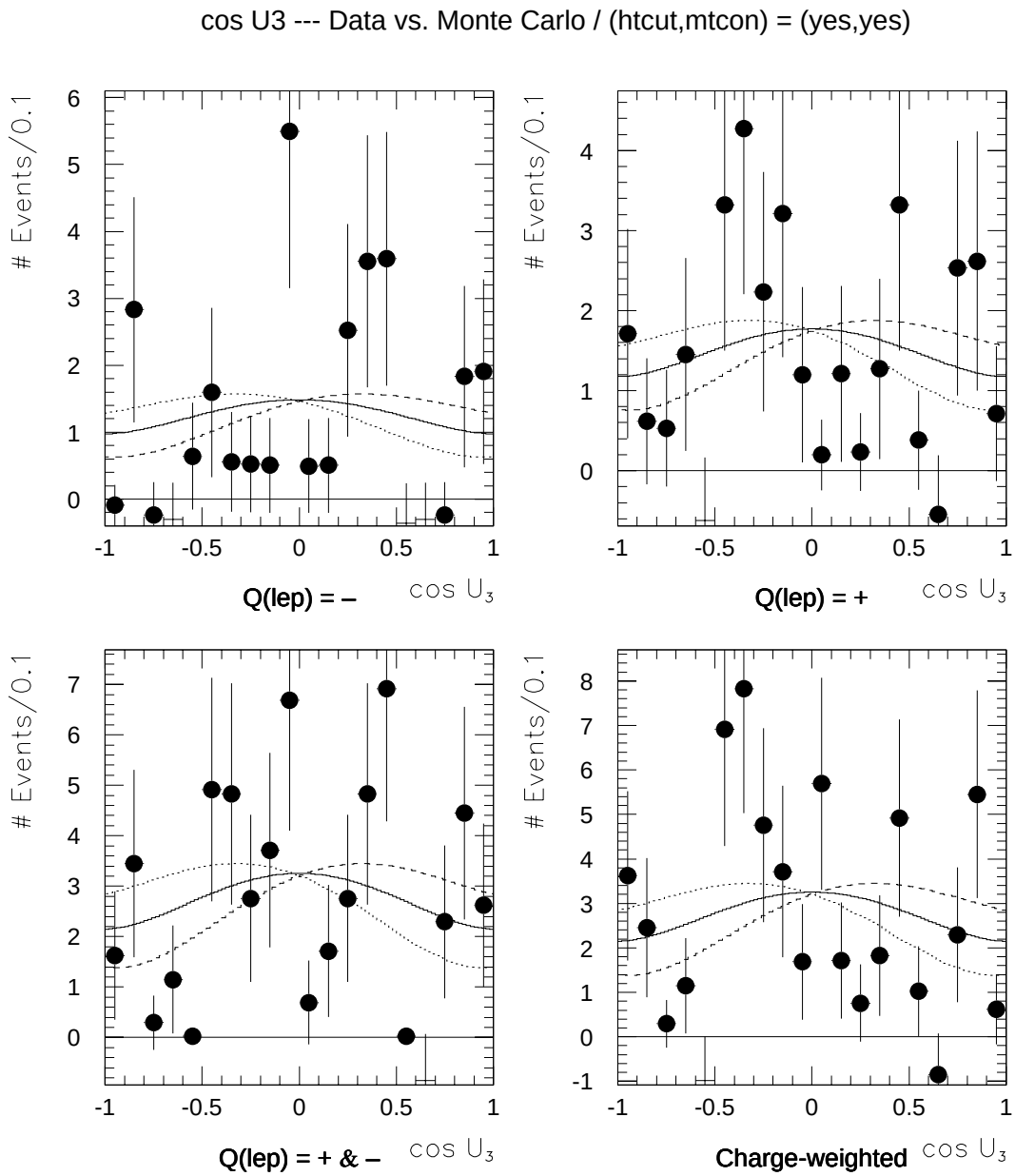


Figure 10.31: Same as figure 10.7, but with optimization (HTCUT, MTCON) = (YES, YES).

10.4 The Measurements

The result of the measurement of the degree of top spin polarization and of the soft transverse W helicity fraction is presented in this section.

10.4.1 The Top Spin Polarization Measurement

Table 10.3 shows the degree of top spin polarization extracted from the experimental data before any optimizations — *i.e.* (HTCUT, MTCON) = (NO, NO). Figure 10.33 is a graphical representation of the numbers in the table. The following are some explanatory remarks about the table and figure:

- The numbers labeled "POL$(-)$" are obtained using, for the 100% polarized distributions, that which favors negative values of $\cos U_i$; the numbers labeled "POL$(+)$", on the other hand, are obtained using distributions favoring positive values of $\cos U_i$. The two sets of numbers are related to each other in the following manner: a measured value of top spin polarization for POL$(-)$ is approximately equal in magnitude, but opposite in sign, compared to the corresponding measurement for POL$(+)$. Similarly, the positive (negative) error for a given measurement in POL$(-)$ is approximately equal to the negative (positive) error for the corresponding error in POL$(+)$.

- The labels Q$(-)$ and Q$(+)$ represent the subset of data with the charge of the primary lepton $q_\ell = -$ and $+$, respectively. The label Q$(- \ \& \ +)$ represents the combination of Q$(-)$ and Q$(+)$, while Q(wgt) represents the charge-weighted combination of the two sets.

- The measured value of the top spin polarization is, by definition, the parameter value that minimize the log-likelihood function. The positive error is defined as the positive displacement from the minimum value that changes the value of the log-likelihood function by 0.5 from the minimum value, and the negative error is defined similarly.

Tables 10.4 to 10.6 show the results of the measurement of the top spin polarization for the optimization configuration (HTCUT, MTCON) = (YES, NO), (NO, YES), and (YES,

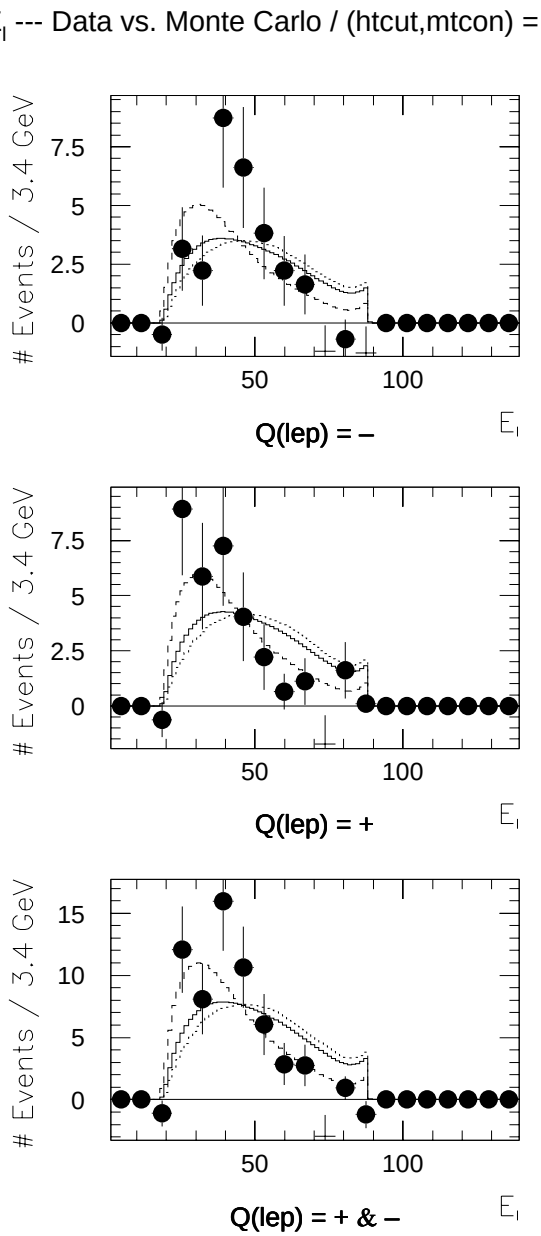


Figure 10.32: Same as figure 10.8, but with optimization (HTCUT, MTCON) = (YES, YES).

		Q(−)	Q(+)	Q(− & +)	Q(wgt)
POL(−)	U1	$-1.001^{+0.882}_{-0.808}$	$-1.376^{+0.788}_{-0.678}$	$-1.209^{+0.587}_{-0.537}$	$-0.269^{+0.612}_{-0.583}$
	U2	$-0.066^{+0.899}_{-0.851}$	$0.105^{+0.642}_{-0.666}$	$0.045^{+0.529}_{-0.533}$	$0.145^{+0.533}_{-0.549}$
	U3	$-1.148^{+0.964}_{-0.778}$	$0.969^{+0.989}_{-1.109}$	$-0.227^{+0.773}_{-0.762}$	$1.132^{+0.620}_{-0.706}$
POL(+)	U1	$1.122^{+0.841}_{-0.930}$	$1.252^{+0.684}_{-0.806}$	$1.198^{+0.549}_{-0.606}$	$0.172^{+0.584}_{-0.613}$
	U2	$0.229^{+0.930}_{-0.962}$	$-0.041^{+0.668}_{-0.650}$	$0.047^{+0.548}_{-0.545}$	$-0.050^{+0.537}_{-0.524}$
	U3	$1.242^{+0.728}_{-0.924}$	$-0.736^{+1.045}_{-0.878}$	$0.368^{+0.739}_{-0.754}$	$-0.949^{+0.698}_{-0.609}$

Table 10.3: The result of the top polarization measurement in the '1'-, '2'-, and '3'-directions before optimization. The section labeled "POL(−)" show results obtained using 100% polarized distributions favoring negative values of $\cos U_i$, while those for "POL(+)" show results obtained using distributions favoring positive values of $\cos U_i$. The numbers in POL(−) are approximately equal in magnitude but opposite in sign to the corresponding numbers in POL(+). The numbers under Q(−) and Q(+) are from events with primary lepton charge q_ℓ = − and +, respectively. The numbers under Q(− & +) are obtained from the combination of Q(−) and Q(+), while those under Q(wgt) are from the charge-weighted combination of the two sets.

		Q(−)	Q(+)	Q(− & +)	Q(wgt)
POL(−)	U1	$-0.804^{+0.853}_{-0.811}$	$-1.159^{+0.805}_{-0.714}$	$-0.994^{+0.588}_{-0.551}$	$-0.227^{+0.603}_{-0.578}$
	U2	$0.212^{+0.873}_{-0.861}$	$0.042^{+0.613}_{-0.630}$	$0.100^{+0.503}_{-0.510}$	$-0.014^{+0.516}_{-0.523}$
	U3	$-1.143^{+0.936}_{-0.766}$	$0.806^{+0.988}_{-1.075}$	$-0.291^{+0.741}_{-0.727}$	$1.060^{+0.619}_{-0.692}$
POL(+)	U1	$0.869^{+0.856}_{-0.904}$	$1.055^{+0.724}_{-0.828}$	$0.973^{+0.569}_{-0.611}$	$0.124^{+0.578}_{-0.602}$
	U2	$-0.138^{+0.926}_{-0.921}$	$0.023^{+0.630}_{-0.620}$	$-0.028^{+0.523}_{-0.517}$	$0.087^{+0.513}_{-0.508}$
	U3	$1.235^{+0.720}_{-0.899}$	$-0.612^{+1.020}_{-0.892}$	$0.418^{+0.706}_{-0.723}$	$-0.890^{+0.683}_{-0.608}$

Table 10.4: The result of the top polarization measurement in the '1'-, '2'-, and '3'-directions, with optimization (HTCUT, MTCON) = (YES, NO).

YES). Figures 10.34 to 10.36 show graphical representations of the measurements for each configuration.

10.4.2 The Soft Transverse W Helicity Measurement

Table 10.7 shows the results of the measurement of the soft transverse W helicity fraction from the experimental data. Figure 10.37 is a graphical representation of the numbers in the table.

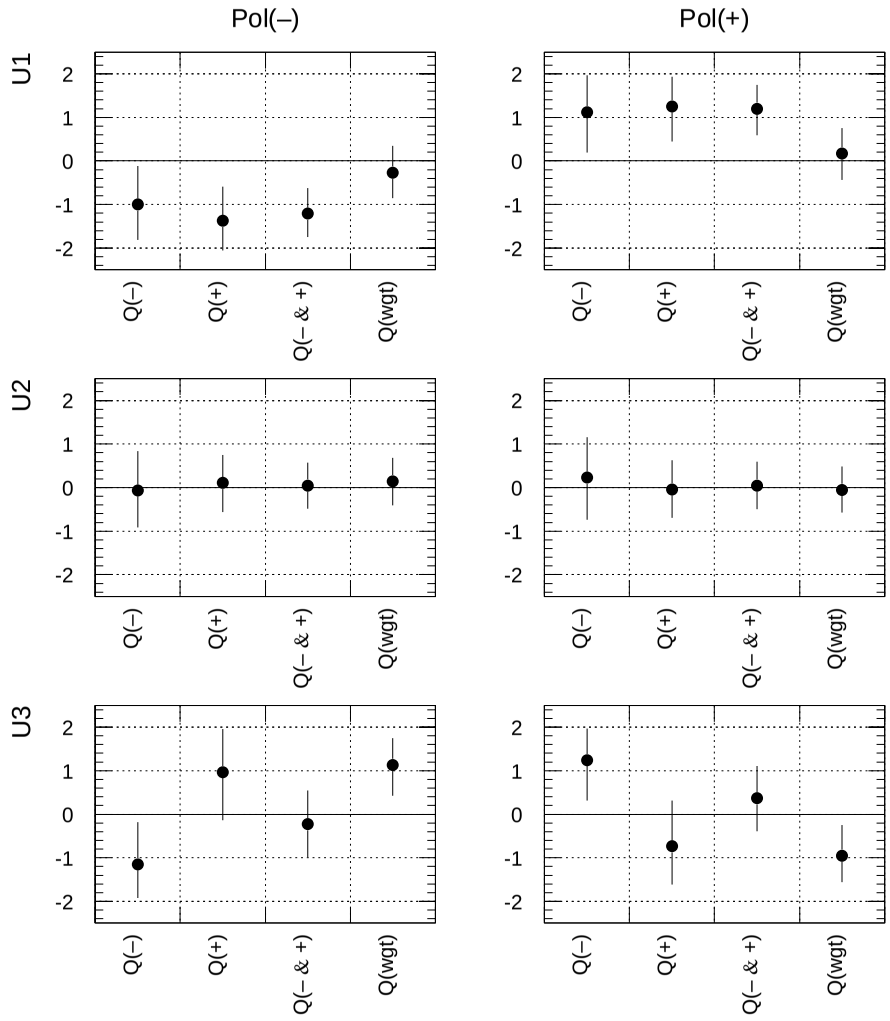


Figure 10.33: Graphical representation of the numbers shown in table 10.3.

		Q(−)	Q(+)	Q(− & +)	Q(wgt)
POL(−)	U1	$-0.383^{+0.753}_{-0.713}$	$-0.750^{+0.780}_{-0.657}$	$-0.563^{+0.546}_{-0.505}$	$-0.163^{+0.574}_{-0.548}$
	U2	$0.152^{+0.799}_{-0.777}$	$-0.467^{+0.626}_{-0.581}$	$-0.229^{+0.490}_{-0.474}$	$-0.336^{+0.500}_{-0.487}$
	U3	$-0.555^{+1.140}_{-1.027}$	$0.500^{+1.014}_{-1.080}$	$0.001^{+0.783}_{-0.786}$	$0.685^{+0.732}_{-0.786}$
POL(+)	U1	$0.441^{+0.776}_{-0.798}$	$0.814^{+0.676}_{-0.783}$	$0.637^{+0.531}_{-0.565}$	$0.187^{+0.533}_{-0.559}$
	U2	$-0.105^{+0.798}_{-0.817}$	$0.526^{+0.569}_{-0.622}$	$0.293^{+0.475}_{-0.495}$	$0.388^{+0.479}_{-0.494}$
	U3	$0.893^{+0.961}_{-1.148}$	$-0.546^{+1.095}_{-1.021}$	$0.138^{+0.800}_{-0.802}$	$-0.551^{+0.788}_{-0.742}$

Table 10.5: The result of the top polarization measurement in the '1'-, '2'-, and '3'-directions, with optimization (HTCUT, MTCON) = (NO, YES).

		Q(−)	Q(+)	Q(− & +)	Q(wgt)
POL(−)	U1	$-0.463^{+0.727}_{-0.679}$	$-0.810^{+0.752}_{-0.629}$	$-0.633^{+0.525}_{-0.482}$	$-0.166^{+0.565}_{-0.537}$
	U2	$0.055^{+0.756}_{-0.732}$	$-0.293^{+0.582}_{-0.567}$	$-0.162^{+0.460}_{-0.451}$	$-0.190^{+0.466}_{-0.464}$
	U3	$-0.849^{+1.086}_{-0.935}$	$0.603^{+0.971}_{-1.038}$	$-0.089^{+0.758}_{-0.757}$	$0.749^{+0.695}_{-0.748}$
POL(+)	U1	$0.527^{+0.767}_{-0.796}$	$0.913^{+0.668}_{-0.791}$	$0.726^{+0.528}_{-0.567}$	$0.153^{+0.531}_{-0.553}$
	U2	$-0.004^{+0.754}_{-0.778}$	$0.333^{+0.564}_{-0.584}$	$0.211^{+0.456}_{-0.466}$	$0.230^{+0.459}_{-0.462}$
	U3	$0.905^{+0.918}_{-1.082}$	$-0.552^{+1.042}_{-0.982}$	$0.147^{+0.756}_{-0.762}$	$-0.696^{+0.751}_{-0.703}$

Table 10.6: The result of the top polarization measurement in the '1'-, '2'-, and '3'-directions, with optimization (HTCUT, MTCON) = (YES, YES).

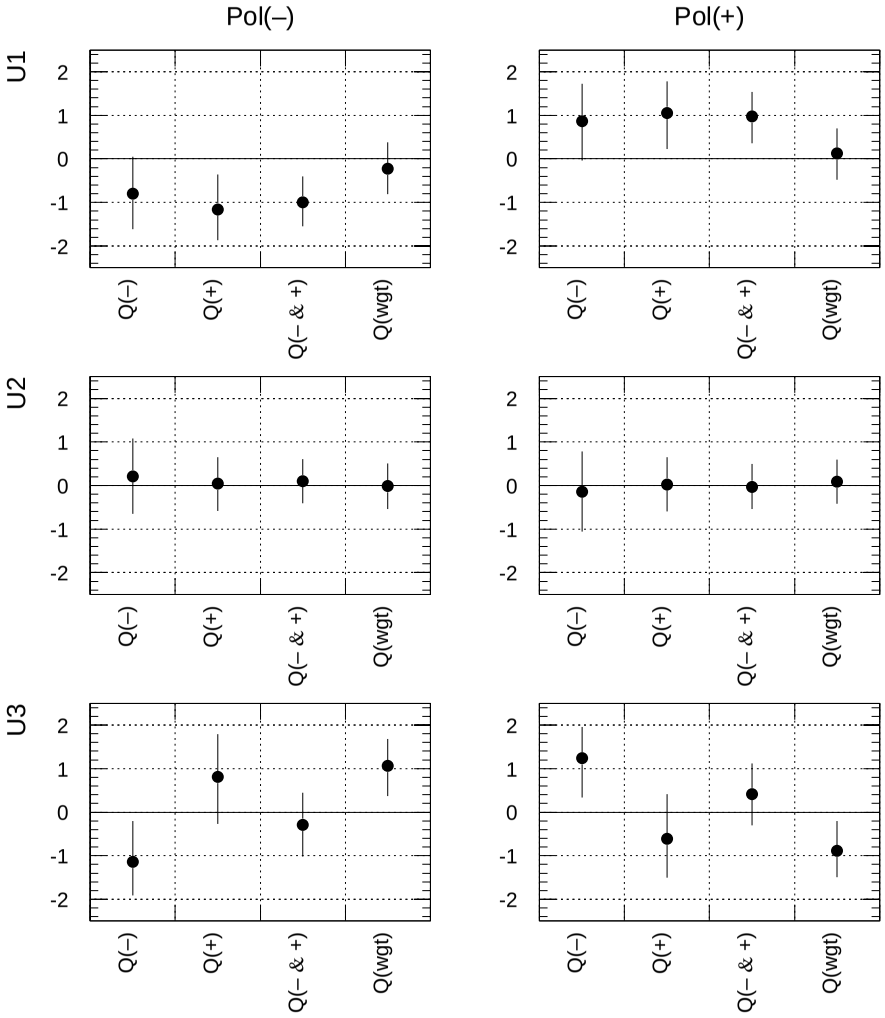


Figure 10.34: Graphical representation of the numbers shown in table 10.4.

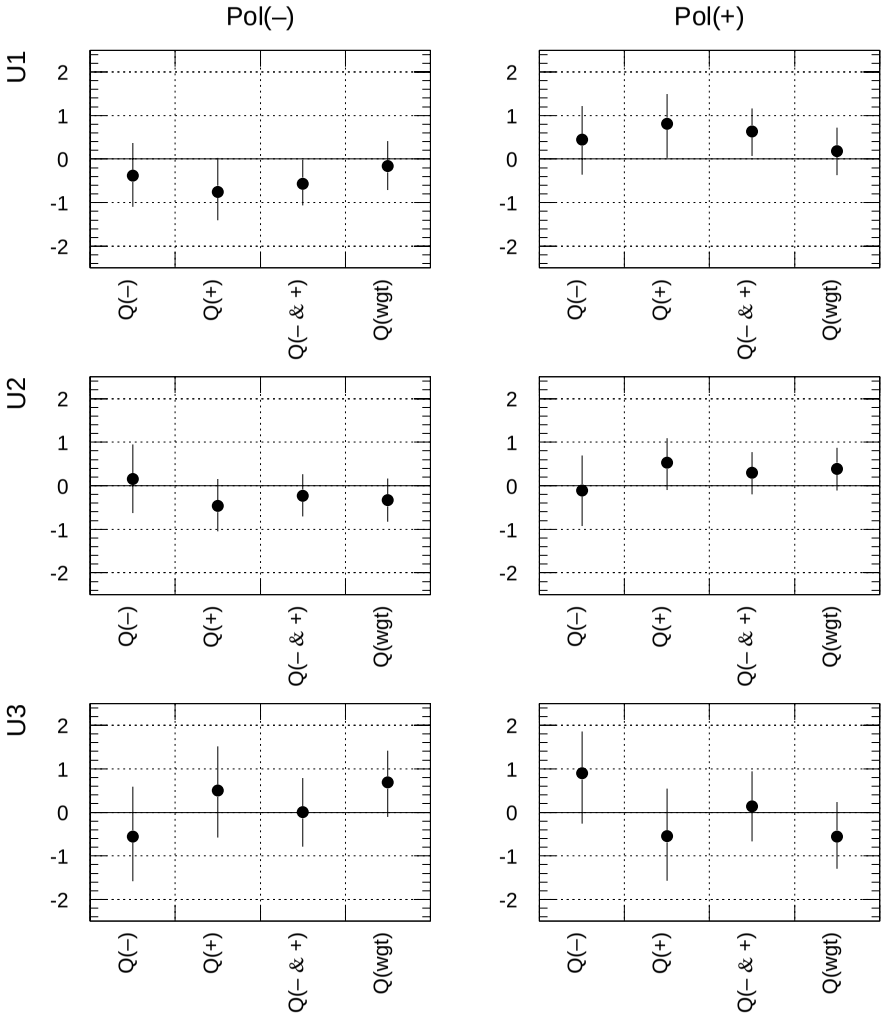


Figure 10.35: Graphical representation of the numbers shown in table 10.5.

HTCUT	MTCON	Q(−)	Q(+)	Q(− & +)
NO	NO	$0.622^{+0.512}_{-0.561}$	$0.707^{+0.429}_{-0.425}$	$0.675^{+0.332}_{-0.338}$
YES	NO	$0.596^{+0.510}_{-0.546}$	$0.967^{+0.403}_{-0.420}$	$0.823^{+0.313}_{-0.327}$
NO	YES	$0.816^{+0.674}_{-0.678}$	$0.701^{+0.423}_{-0.413}$	$0.732^{+0.361}_{-0.356}$
YES	YES	$1.007^{+0.595}_{-0.596}$	$1.091^{+0.396}_{-0.420}$	$1.064^{+0.334}_{-0.345}$

Table 10.7: The result of the soft transverse W helicity measurement in each optimization configuration.

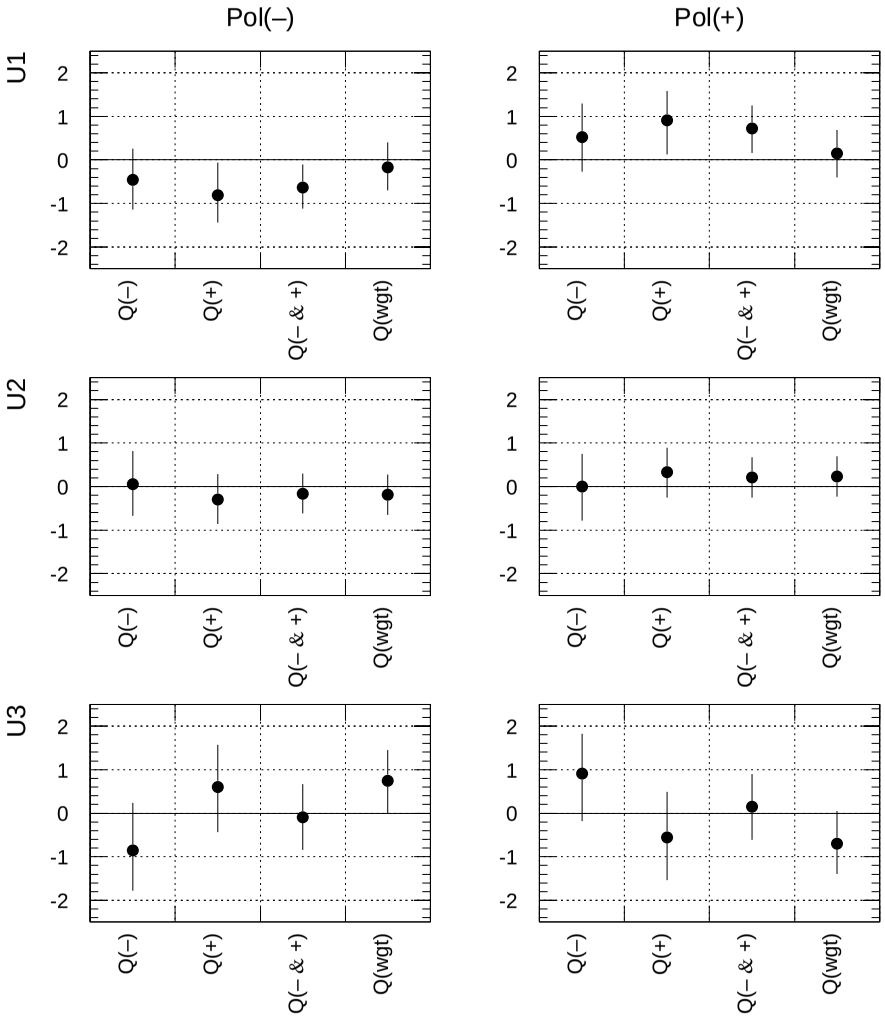


Figure 10.36: Graphical representation of the numbers shown in table 10.6.

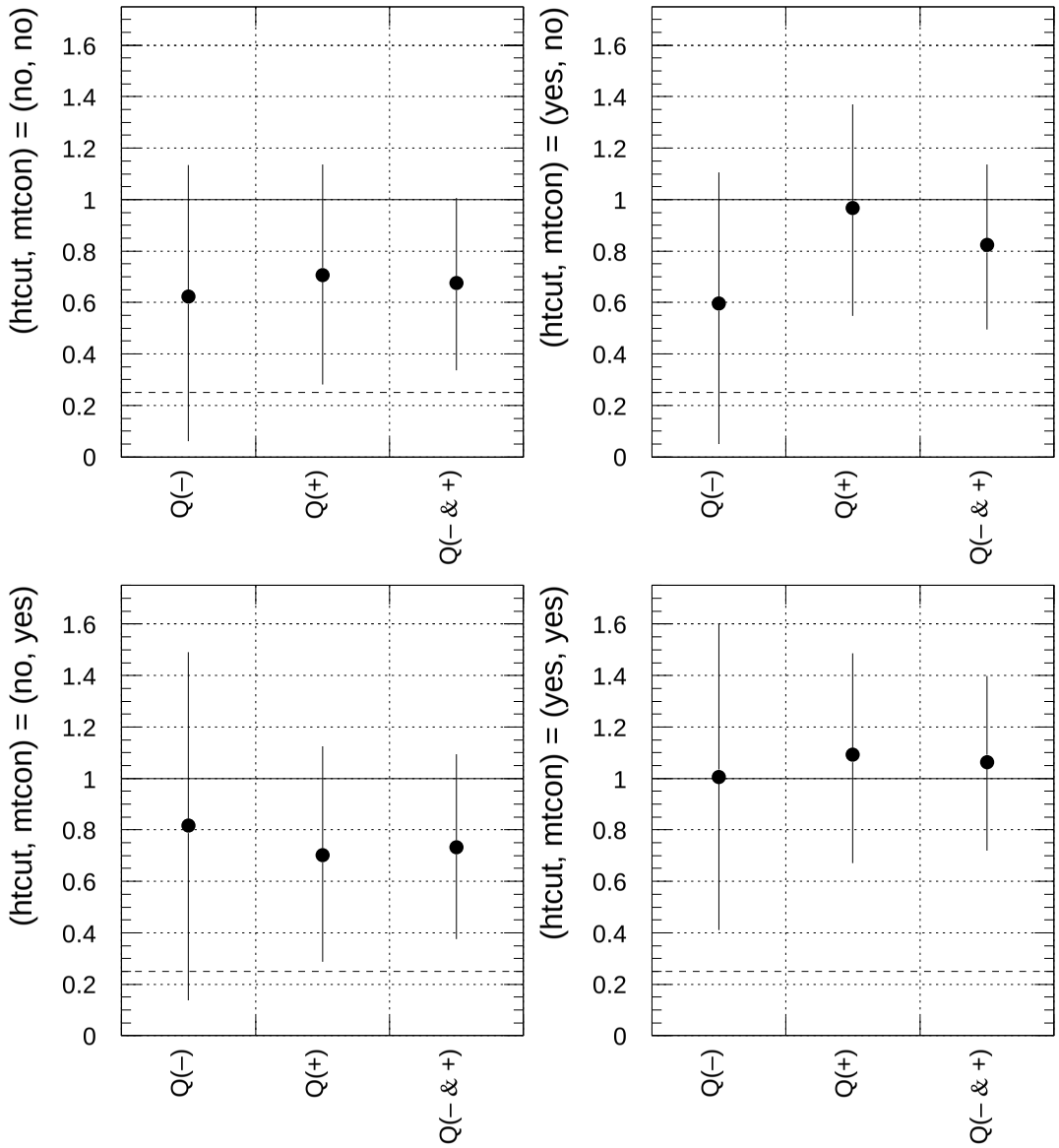


Figure 10.37: Graphical representation of the numbers shown in table 10.7. The dashed line shows the standard model prediction, $r_T = 0.25$.

OBSERVABLE	HTCUT	MTCON	Q(− & +)	Q(wgt)
U1	NO	NO	$1.198^{+0.549}_{-0.606}$	$0.172^{+0.584}_{-0.613}$
	YES	NO	$0.973^{+0.569}_{-0.611}$	$0.124^{+0.578}_{-0.602}$
	NO	YES	$0.637^{+0.531}_{-0.565}$	$0.187^{+0.533}_{-0.559}$
	YES	YES	$0.726^{+0.528}_{-0.567}$	$0.153^{+0.531}_{-0.553}$
U2	NO	NO	$0.047^{+0.548}_{-0.545}$	$-0.050^{+0.537}_{-0.524}$
	YES	NO	$-0.028^{+0.523}_{-0.517}$	$0.087^{+0.513}_{-0.508}$
	NO	YES	$0.293^{+0.475}_{-0.495}$	$0.388^{+0.479}_{-0.494}$
	YES	YES	$0.211^{+0.456}_{-0.466}$	$0.230^{+0.459}_{-0.462}$
U3	NO	NO	$0.368^{+0.739}_{-0.754}$	$-0.949^{+0.698}_{-0.609}$
	YES	NO	$0.418^{+0.706}_{-0.723}$	$-0.890^{+0.683}_{-0.608}$
	NO	YES	$0.138^{+0.800}_{-0.802}$	$-0.551^{+0.788}_{-0.742}$
	YES	YES	$0.147^{+0.756}_{-0.762}$	$-0.696^{+0.751}_{-0.703}$

Table 10.8: Summary of the top polarization measurement.

10.5 Summary of the Results

In this section, the results presented in the last section is summarized. For the top polarization measurement, the measurements under the configurations POL$(-)$ and POL$(+)$ are redundant. Both were shown for the sake of completeness, but so many redundant numbers tend to obscure the results. Thus, in this summary, the numbers obtained in the configuration POL$(+)$ will be chosen to represent the measured value of top polarization. Also, within this configuration, the charge configurations Q$(-$ & $+)$ and Q(wgt) are taken to represent the measurement. It should be noted that a positive value of top spin polarization in the configuration POL$(+)$ indicates that the experimental distribution favors positive values of $\cos U_i$ $(i = 1, 2, 3)$, while a negative value indicates that the experimental distribution favors negative values of $\cos U_i$. For the soft transverse W helicity measurement, the result for the charge configuration Q$(-$ & $+)$ is presented. Table 10.8 shows a summary of the top polarization measurement; figure 10.38 is a graphical representation of the results. Table 10.9 and figure 10.39 show the same for the W helicity measurement.

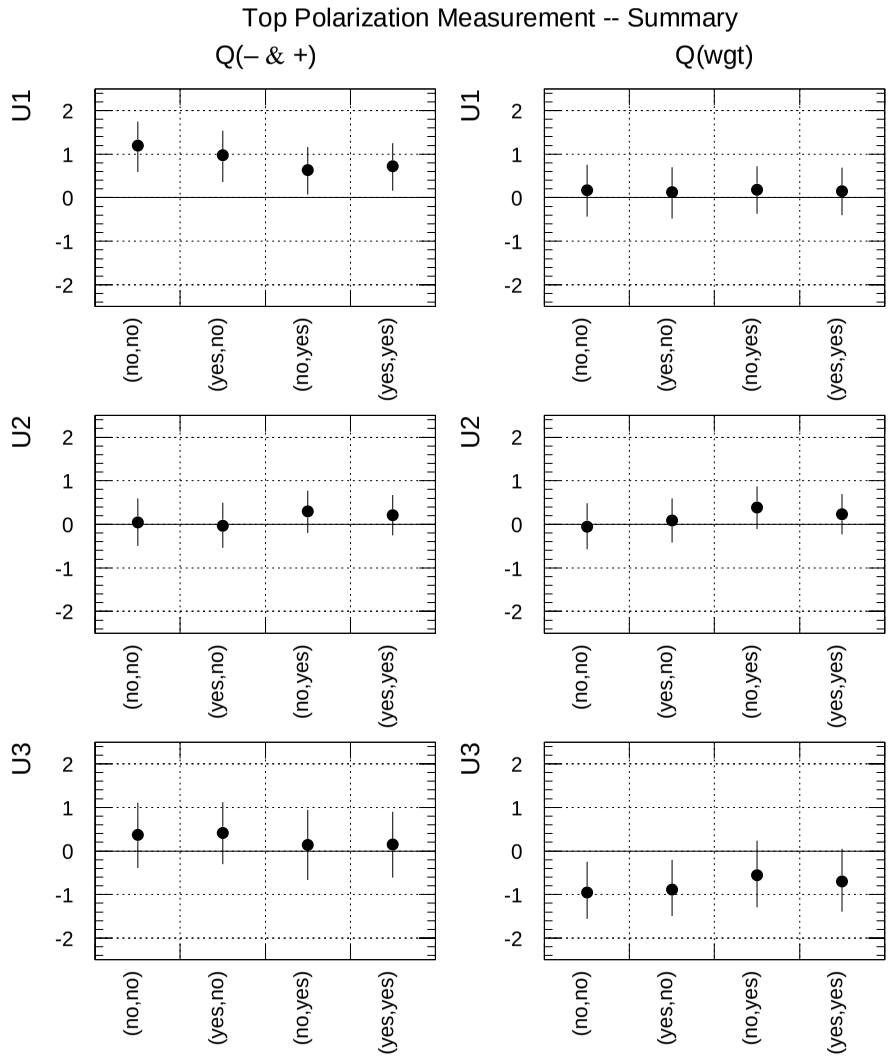


Figure 10.38: Graphical representation of the numbers shown in table 10.8.

HTCUT	MTCON	$Q(- \ \& \ +)$
NO	NO	$0.675^{+0.332}_{-0.338}$
YES	NO	$0.823^{+0.313}_{-0.327}$
NO	YES	$0.732^{+0.361}_{-0.356}$
YES	YES	$1.064^{+0.334}_{-0.345}$

Table 10.9: Summary of the W helicity measurement.

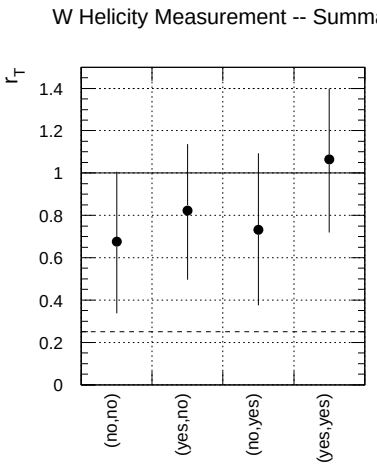


Figure 10.39: Graphical representation of the numbers shown in table 10.9. The dashed line indicates the standard model prediction.

10.6 Comparing the Measurements with the Standard Model Prediction

According to the standard model, the top quarks in $t\bar{t}$ production have negligible spin polarization. Thus, if the signal events in the experimental data originate from standard model $t\bar{t}$, the most likely measured value of spin polarization along the i-direction (i = 1, 2, 3) is zero, with errors given in table 8.16. Table 10.10 shows the number of standard deviations the measured values of the polarization are from the expected value — *i.e.* each entry in the table shows the following:

$$\frac{\alpha(meas) - \alpha(expect)}{\sigma} \tag{10.1}$$

where $\alpha(meas)$ is the measured top polarization, $\alpha(expect)$ is the expected polarization, and σ is the error shown in table 8.16. Note that for the '1'-direction, the configuration POL(+) was chosen, so the errors for U1+ are used. Figure 10.40 is a graphical representation of the numbers in table 10.10.

According to the standard model, the soft transverse W helicity fraction is $r_T = 0.30$, assuming $m_{top} = 175\ GeV$. After taking account of acceptance effects, the standard model prediction for r_T is 0.25. Thus, assuming that the signal events in the experimental data originate from standard model $t\bar{t}$, the most likely measured value of r_T is 0.25, with errors

OBSERVABLE	HTCUT	MTCON	Q(− & +)	Q(wgt)
U1	NO	NO	1.88	0.27
	YES	NO	1.86	0.20
	NO	YES	1.07	0.31
	YES	YES	1.25	0.26
U2	NO	NO	0.08	−0.09
	YES	NO	−0.05	0.17
	NO	YES	0.56	0.75
	YES	YES	0.43	0.47
U3	NO	NO	0.46	−1.19
	YES	NO	0.55	−1.16
	NO	YES	0.17	−0.70
	YES	YES	0.20	−0.94

Table 10.10: The difference between the measured polarization and the standard model value, in number of standard deviations.

HTCUT	MTCON	Q(− & +)
NO	NO	1.33
YES	NO	1.92
NO	YES	1.61
YES	YES	2.91

Table 10.11: The difference between the measured r_T and the standard model value, in number of standard deviations.

given in table 8.16. Table 10.11 and figure 10.41 show the number of standard deviations the measured value of r_T is from the expected value.

10.7 Interpreting the Results

10.7.1 Interpreting the Result of the Top Polarization Measurement

Let us first consider, in the abstract, the relationship between the sign of the degree of polarization in table 10.8 and the spin configuration for t or $\bar{t}$. Let us denote the degree of polarization by α, the spin polarization vector of t by $\hat{s}_+$, and that of $\bar{t}$ by $\hat{s}_-$. Then,

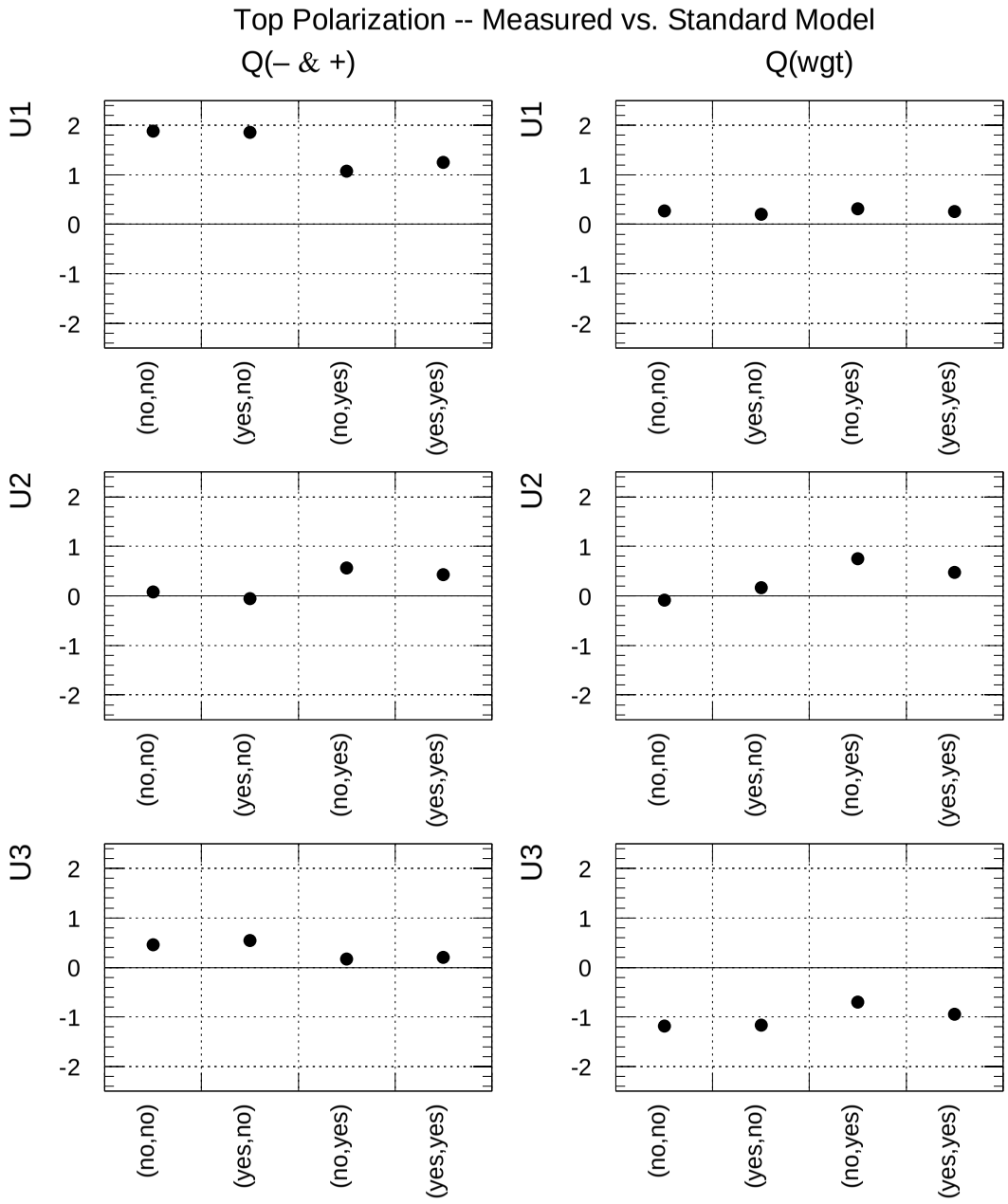


Figure 10.40: Graphical representation of the numbers shown in table 10.10.

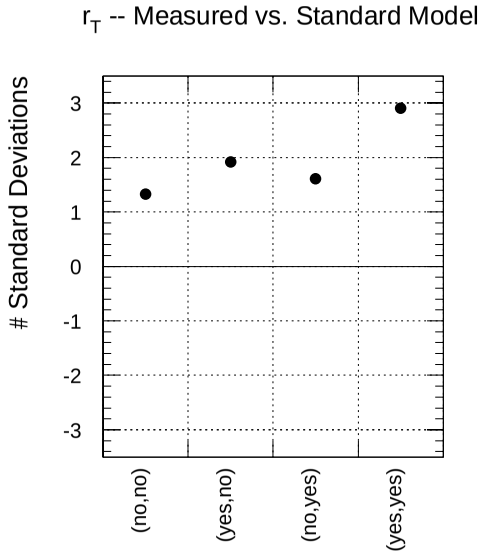


Figure 10.41: Graphical representation of the numbers shown in table 10.11.

for the charge configuration Q(− & +), $\alpha > 0$ implies $\hat{s}_+$ points *along* the i-direction ($i =$ 1, 2, 3), while $\hat{s}_-$ points *against* the i-direction; $\alpha < 0$ implies $\hat{s}_+$ points *against*, and $\hat{s}_-$ points *along*, the i-direction. For Q(wgt), $\alpha > 0$ implies that both $\hat{s}_+$ and $\hat{s}_-$ point *along* the i-direction, while $\alpha < 0$ implies that both $\hat{s}_+$ and $\hat{s}_-$ point *against* the i-direction. The possible spin configurations are illustrated in figure 10.42.

For the direction $i = 1$, the above spin configurations take on special significance because the reference axis is along the t or $\bar{t}$ direction. In this situation, Q(− & +) with $\alpha > 0$ implies that, t has, on average, a net right-handed helicity (spin pointing *along* momentum) while $\bar{t}$ has a net left-handed helicity (spin pointing *against* momentum). This situation will be denoted, suggestively, as (t_R, $\bar{t}_L$). For Q(− & +), $\alpha < 0$ implies (t_L, $\bar{t}_R$). For Q(wgt), $\alpha > 0$ implies (t_R, $\bar{t}_R$), while $\alpha < 0$ implies (t_L, $\bar{t}_L$). Figure 10.43 describes the possible spin configurations.

Now let us examine the degree of top spin polarization extracted from the experimental data (see table 10.8). According to table 10.10, all of the measured values are within two standard deviations of the unpolarized hypothesis. Therefore, within the precisions attainable in this analysis, one can conclude that the observable distributions in the experimental data are consistent with the standard model prediction.

The measurement with the largest deviation from the standard model is $\cos U_1$ with

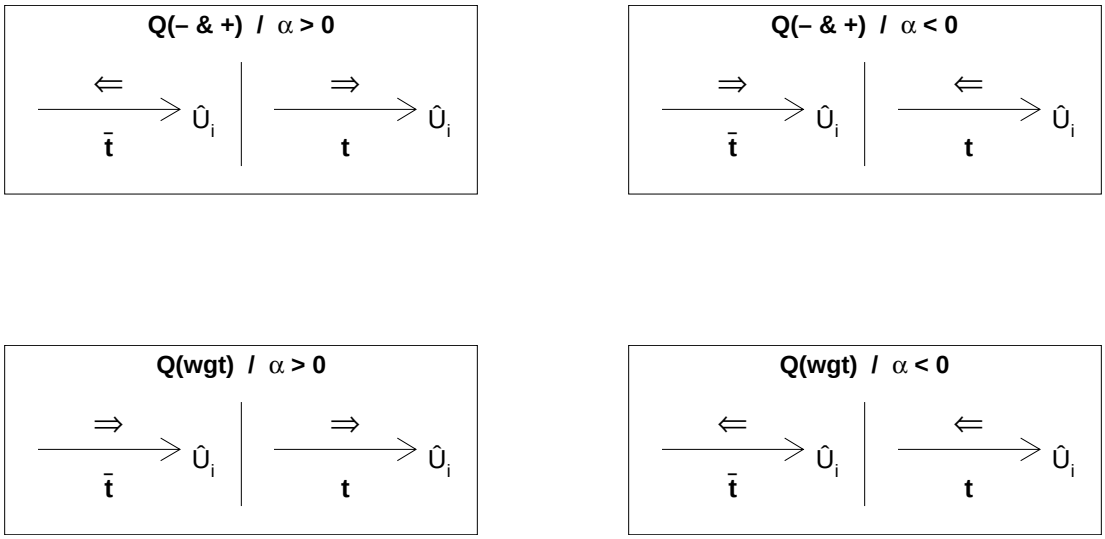


Figure 10.42: The spin configurations for $Q(-\ \&\ +)$ and $Q(\text{wgt})$. The large arrows in each diagram represent the unit vector $\hat{U}_i$ along the i-direction in the t and $\bar{t}$ rest frames. The small double-arrows represent the t and $\bar{t}$ spin polarization vector.

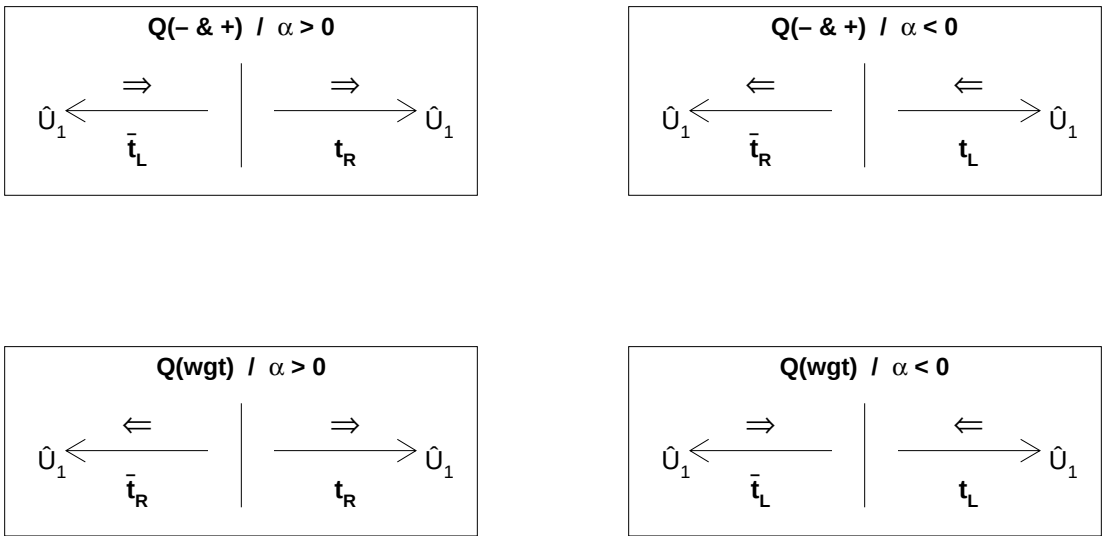


Figure 10.43: The same as figure 10.42, but for the specific case for the '1'-direction. In this case, the t and $\bar{t}$ unit vectors are suggestively made to point in opposite directions — this is done to indicate the fact that the t and $\bar{t}$ momenta are equal-and-opposite in the $t\bar{t}$ rest frame. The subscript under t and $\bar{t}$ indicate the net helicity of these quarks.

the optimization configuration (HTCUT, MTCON) = (NO, NO) and charge configuration Q(− & +). In this case, the top polarization is measured to be 1.198; this is 1.88 standard deviations from the unpolarized hypothesis. If this mildly large deviation were due to top spin polarization, then figure 10.43 indicates that the t and $\bar{t}$ have helicity configuration $(t_R, \bar{t}_L)$. This would indicate a CP-conserving chiral anomaly in the production mechanism of $t\bar{t}$.

Another measurement with mildly large deviation from the standard model is $\cos U_3$ with optimization configuration (HTCUT, MTCON) = (NO, NO) and charge configuration Q(wgt). In this case, the top polarization is measured to be -0.949, which is -1.19 standard deviations from the unpolarized hypothesis. If this deviation were due to top spin polarization, then figure 10.42 indicates that the spin polarization vectors for t and $\bar{t}$ both point *against* the 3-direction. There is no plausible scenario for a $t\bar{t}$ production mechanism that produces this spin configuration. Any such mechanism violates CP.

10.7.2 Interpreting the Result of the W Helicity Measurement

Let us first consider, in the abstract, the physical implications of different values of r_T. For a given value of r_T, the E_ℓ distribution from the semileptonic decay of t or $\bar{t}$ is given as follows:

$$g(E_\ell) = r_T \cdot g_s(E_\ell) + (1 - r_T) \cdot g_m(E_\ell) \tag{10.2}$$

The function g_s is the soft E_ℓ distribution, while g_m is the medium distribution. The extreme case $r_T = 0$ corresponds to t and $\bar{t}$ that couple only to the longitudinal helicity state of W. In the other extreme $r_T = 1$, t and $\bar{t}$ couple only to the transverse helicity states of W (the t ($\bar{t}$) quark couples to the left- (right-) handed helicity state of W).

Now, let us examine r_T extracted from the experimental data (see table 10.9). Whereas the expected value is $r_T = 0.25$, the measured values vary between $r_T = 0.675$ to 1.064, depending on the optimization configuration. The fact that the measured value of r_T is larger than the theoretically expected value implies that the E_ℓ distributions in the experimental data are *softer* than predicted by the monte carlo distributions — this is well-demonstrated by the E_ℓ distributions shown in section 10.3. Table 10.11 shows that, in

all configurations except (HTCUT,MTCON) = (YES,YES), the softness of the E_ℓ distributions in the experimental data are not unlikely to be a statistical fluctuation. The configuration (HTCUT,MTCON) = (YES,YES), however, has $r_T = 1.064$, which is 2.91 standard deviations from the standard model prediction. Thus, *if it is assumed that the modeling of the signal and background is correct*, this measurement would indicate strong experimental evidence for non-standard model physics. There are indications, however, that the modeling is, in fact, inadequate. This issue is discussed in appendix P.

Chapter 11

Prospects for Run II

11.1 Introduction

This chapter presents an overview of the expected top polarization and W helicity measurement resolution in Tevatron Run II.

11.2 Run II Experimental Parameters

Details of the Run II experimental parameters for the CDF detector can be found in [36]. Here, pieces of information relevant to $t\bar{t}$ events in the lepton + jets decay channel are summarized:

Integrated Luminosity

The target integrated luminosity for Run II is 2 fb^{-1}, compared to 0.109 fb^{-1} in Run I.

$p\overline{p}$ Center-of-mass Energy

The center-of-mass energy of the $p\overline{p}$ system will be 2.0 TeV, up from 1.8 TeV in Run I.

Acceptance for Lepton + Jets $t\bar{t}$ Events

One of the most important factors determining the acceptance of $t\bar{t}$ events in the lepton + jets decay channel is the acceptance for isolated high-P_T electrons and muons. The new plug calorimeter and the Intermediate/Forward Tracker will extend the coverage

of electrons to the region $1 \leq |\eta| \leq 2$, which is expected to increase the electron acceptance by 33%. The improvement in the η and ϕ coverage of the muon system is expected to increase the muon acceptance by 12%.

b-tag Efficiency

The SVX II improves upon the Run I silicon tracker by, among other things, (1) increasing the length of coverage from 52 cm to 96 cm; and (2) having five double-sided layers compared to four single-sided ones. With SVX II, the efficiency for tagging ≥ 1 b jet in $t\bar{t}$ events is expected to increase from 40% to 65%; the corresponding figure for ≥ 2 b jets is 20%, up from 7% in Run I. Furthermore, although no firm numbers are available, the mistag rate is expected to decrease significantly.

11.3 Implications of the Run II Experimental Parameters

The implication of increased luminosity is clear: all things being equal, an increase from 0.109 fb^{-1} to 2 fb^{-1} should increase the number of $t\bar{t}$ events by a factor of 2/0.109 = 18.3. As noted above, however, all things are not equal between Run I and Run II. Because of the increase in $p\overline{p}$ $\sqrt{s}$, $\sigma(t\bar{t})$ should increase by 40%. Thus, the number of $t\bar{t}$ events produced in Run II should be $18.3 \times 1.4 = 25.6$ times that produced in Run I. If one assumes a theoretical cross section of $\sigma(t\bar{t}) = 4.9$ pb for $m_{top} = 175$ GeV at $\sqrt{s} = 1.8$ TeV, then the number of $t\bar{t}$ events produced in Run I is expected to be about 523. Thus, the number of $t\bar{t}$ events produced in Run II should be about $523 \times 25.6 = 13,400$.

Increased electron and muon acceptance and improved b-tagging imply that a greater fraction of the produced $t\bar{t}$ events is accepted. According to [36], 5.8% of the 13,400 $t\bar{t}$ events is expected to have ≥ 1 SVX b-tag and pass the event selection cuts of this analysis. The corresponding figure for ≥ 2 SVX b-tags is 1.8%. The corresponding acceptances in Run I are 3.0% and 0.52%, respectively. Thus, in Run II, one expects 777 and 241 $\geq$ 1- and $\geq$ 2-SVX b-tagged events originating from $t\bar{t}$.

So far, no mention has been made about the background contributions. The greatest source of background is the QCD W + jets events. The following are some observations on this process[36]:

- The W cross section increases by about 12% going from $\sqrt{s} = 1.8\ TeV$ to $2.0\ TeV$.
- The $W + \geq n$ jets cross section scales as the W cross section.

These two observations imply that the number of W + jets events that fake $t\bar{t}$ events should increase by $18.3 \times 1.12 = 20.5$. This rate of increase is smaller than that for $t\bar{t}$ (25.6). Thus, one can expect a significant reduction in the signal to background ratio. There are, however, complicating issues which have not yet been examined closely:

- In Run II, typical instantaneous luminosities are expected to be considerably larger than those in Run I. An increase in instantaneous luminosity results in increased multiple interactions. This, in turn, will cause W + 0, 1, 2, and 3 jets events to be promoted to W + 4 jets events. This can cause the background fraction to increase.
- Because of improvements in the silicon tracker, the mistag rate is expected to decrease significantly. This will tend to decrease the background fraction.
- Non-W backgrounds (primarily $b\bar{b}$ events) are important secondary sources of background. No estimate has yet been made as to how this will change in Run II.

11.4 Measurement Resolution in Run II

The measurement resolution (*i.e.* statistical error) for the top polarization and W helicity measurements is given by the following formula:

$$\sigma = \frac{1}{(1-\beta)} \cdot \frac{K}{\sqrt{N_{ev}}} \tag{11.1}$$

β is the background fraction, N_{ev} is the number of signal + background events, and K is a factor that depends mostly on the shape of the component functions describing the signal distribution, and is given by equation 9.4. For the purpose of estimating the Run II measurement resolution, it will be more useful to express σ as follows:

$$\sigma = \frac{1}{\sqrt{1-\beta}} \cdot \frac{K}{\sqrt{S}} \tag{11.2}$$

	# SVX tags	S	β	$f_{b\ell}$ (%)
Run I	$= 1$	13.1	0.13	46.4
	≥ 2	2.7	0.05	66.2
Run II	$= 1$	536	0.11	46.4
	≥ 2	241	0.04	66.2

Table 11.1: Observable-independent factors that determine the top polarization and W helicity measurement resolution. S = the number of expected signal events, β = background fraction, and $f_{b\ell}$ = the fraction of events with the correct jet matched to b_ℓ.

$S = (1 - \beta) \cdot N_{ev}$ is the number of signal events.

According to equation 11.2, σ is determined by observable-independent factors β and S, and an observable-dependent factor K. Table 11.1 shows the observable-independent factors for number of SVX tagged jets $= 1$ and ≥ 2. Also shown is $f_{b\ell}$, the fraction of events with the correct jet matched to b_ℓ (see chapters 6 and 9 for details on $f_{b\ell}$). $f_{b\ell}$ is used later to estimate the K-factor for each observable.

A few comments on the numbers in table 11.1 are in order:

The Number of Signal Events

The number signal events in Run I is obtained taking: (1) $\sigma(t\bar{t}) = 4.9pb$; (2) the integrated luminosity $= 0.109\ fb^{-1}$; and (3) the acceptance for ≥ 1 and ≥ 2 SVX tags is 3.0% and 0.52%, respectively. For Run II, the corresponding figures are as follows: (1) $\sigma(t\bar{t})$ is 1.4 times that of Run I; (2) integrated luminosity $= 2\ fb^{-1}$; and (3) the acceptance for ≥ 1 and ≥ 2 SVX tags is 5.8% and 1.8%, respectively.

The Background Fraction

The background fractions for Run II is obtained from the Run I values using the following facts and simplifying assumptions. Let us assume that all of the background events come from QCD W + jets events. It was stated earlier that the rate of increase in such events from Run I to Run II is 20.5. Let us assume that the increase in acceptance in Run II for background events is the same as for the signal events — in making this assumption, one ignores, among other things: (1) the improvement in the silicon tracker (which would tend to decrease this factor); (2) the increase in instanta-

	# SVX tags	
OBSERVABLE	$= 1$	≥ 2
U1+	3.55	2.91
U1−	3.35	2.78
U2	3.27	2.46
U3	4.54	3.67
E_ℓ	1.82	1.54

Table 11.2: K-factor for all observables.

neous luminosity with consequent increase in underlying event (which would tend to increase this factor); and (3) the η-dependence of signal/background (which could be significant for events with electrons detected by the new plug calorimeter). With this assumption, if the rate of increase in accepted signal and background events is denoted by A_s and A_b, respectively, then their ratio should simply be $A_s/A_b = 25.6/20.5$ — *i.e.* the acceptance factor cancels. Given A_b and A_s, the Run II background fraction β' is given in terms of the Run I fraction β by the following relation:

$$\beta' = \frac{1}{1 + \left(\frac{1-\beta}{\beta}\right) \cdot \left(\frac{A_s}{A_b}\right)} \tag{11.3}$$

$f_{b\ell}$

$f_{b\ell}$ for Run II is assumed to be the same as for Run I. Factors in favor of improved $f_{b\ell}$ in Run II are: (1) improved jet resolution with the new plug calorimeter; (2) better jet energy reconstruction algorithm. At least one factor favors the worsening of $f_{b\ell}$: the increased instantaneous luminosity.

The observable-dependent factor K in equation 11.2 is obtained from the plots in figure 9.3(a). For each observable, one evaluates the value of K at the values of $f_{b\ell}$ given in table 11.1. The results are shown in table 11.2.

Tables 11.1 and 11.2 contain the numbers necessary to calculate the measurement resolutions. The results are given in table 11.3 Note that the combined error is given by the following formula:

		# SVX tags		
	OBSERVABLE	$= 1$	≥ 2	COMBINED
Run I	U1+	1.05	1.82	0.91
	U1−	0.99	1.74	0.86
	U2	0.97	1.54	0.82
	U3	1.34	2.29	1.16
	E_ℓ	0.54	0.96	0.47
Run II	U1+	0.16	0.19	0.12
	U1−	0.15	0.18	0.12
	U2	0.15	0.16	0.11
	U3	0.21	0.24	0.16
	E_ℓ	0.08	0.10	0.06

Table 11.3: Measurement resolutions for Run I and Run II.

$$\frac{1}{\sigma^2} = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} \tag{11.4}$$

The resolutions σ_1 and σ_2 are for # SVX tagged jets $= 1$ and ≥ 2, respectively.

The numbers in table 11.3 can be taken as base-line estimates of Run II resolutions. Let us now try to estimate how much the resolutions can be improved when the following optimizations are performed:

- Inclusion of SLT-only and no-tag events.
- Division of events into the eight subsamples, as per chapter 8.
- Application of H_T cuts.
- Applying a top mass constraint in the $t\bar{t}$ reconstruction process.

The method used to estimate the improvement in resolution is the following. Let us take, for example, the combined Run I measurement resolution for the observable U1+ in table 11.3, $\sigma = 0.91$. Table 8.14, on the other hand, gives the measurement resolution when the above optimizations are taken into account: $\sigma' = 0.57$. The error decreases by a factor of $0.57/0.91 = 0.63$. It will be assumed that the same factor applies in Run II. This procedure is only intended to give a very rough, but not unrealistic, estimate. Using this

OBSERVABLE	S.F.	σ
U1+	0.63	0.08
U1−	0.63	0.08
U2	0.60	0.07
U3	0.63	0.10
E_ℓ	0.57	0.03

Table 11.4: A rough estimate of the optimized Run II measurement resolution. The column under S.F. gives the scale factors giving the decrease in measurement resolution after optimization.

method, the optimized resolution for U1+ in Run II is $\sigma' = 0.63 \times 0.12 = 0.08$. Table 11.4 shows the scale factor and optimized error for all of the observables.

11.5 Implications for Run II

Let us first consider the implications of the Run II resolutions on the top spin polarization measurement. The statistical error σ is estimated to be about 0.07 to 0.10. These numbers are to be compared to the physical range of the degree of polarization α, which is 0 to 1. Thus, in Run II, it should be possible to rule out a large portion of the parameter space. On the other hand, the resolution for U2, $\sigma = 0.07$, is not good enough to measure the $\sim 1\%$ transverse polarization expected from one-loop QCD diagrams in $t\bar{t}$ production [30].

Next, let us consider the W helicity measurement. The standard model prediction for the soft transverse W helicity fraction is $r_T = 0.30$; after acceptance corrections, it is 0.25. Thus, in Run II, the r_T measurement would be $r_T = 0.25 \pm 0.03$. With this level of precision, a 3-component, 2-parameter fit of the sort discussed in appendix N would be practical. If one denotes the soft-transverse, hard-transverse, and longitudinal helicity fraction by r_T, r'_T, and r_L, respectively, then one can expect the following in Run II:

$$\sigma(r_T) = 0.25 \pm 0.04 \tag{11.5}$$

$$\sigma(r'_T) = 0.00 \pm 0.03 \tag{11.6}$$

$$\sigma(r_L) \quad = \quad 0.75 \pm 0.06 \tag{11.7}$$

These numbers show that, in Run II, it should be possible to place a rather stringent limit on the $V + A$ contribution to $t \rightarrow W + b$.

Chapter 12

Conclusion

Top quark physics is a subject in its infancy. Although there are many pieces of direct experimental evidence in support of the top quark's existence, there are many of its properties that have yet to be demonstrated. Until they are, there is some degree of doubt as to whether the signal portion of the $t\bar{t}$ candidate events all originate from the pair production of the top quark. Moreover, even if it is supposed that they are all due to $t\bar{t}$, there is not yet any direct experimental evidence that their production and decay properties are described correctly by the standard model.

This thesis presented a study of the use of the kinematics of fully-reconstructed $t\bar{t}$ events in the testing of the standard model $t\bar{t}$ hypothesis. The ultimate objective of this study is to compare the kinematics in the experimental data with the standard model prediction and quantify the degree of similarity/difference between the two. The result of such comparison would shed light on the validity of the $t\bar{t}$ hypothesis. Before this ultimate objective can be met, however, there are many theoretical and technical issues that must be clarified. Among the more important of these issues are:

- How reliable is the $t\bar{t}$ reconstruction algorithm? How does one quantify its reliability? Is the reconstruction algorithm reliable enough to be used to study the properties of the top quark?

- Should lab frame observables be used? Or should one use top rest frame observables? What are the advantages and disadvantages of each?

- In the top rest frame, how many independent variables are needed to completely

specify the top quark decay? Which variables provide important information about properties of the top quark?

- How can one quantify the degree of similarity/difference between the kinematical distributions in the experimental data and those predicted by the standard model?
- What factors determine the size of the measurement resolution?
- What is the expected measurement resolution? Given this, what can one expect to conclude in Run I? in Run II?

All of these questions have been examined and answered in this thesis. It is the author's hope that the issues dealt with here are used as a starting point in related analyses in Run II.

Due to the limited number of $t\bar{t}$ candidate events available in Run I, one cannot conclude very much from the result of the comparison between the kinematic distributions in the experimental data and those predicted by the standard model. To give a sense of the limitation, it is noted that the measured properties — degree of polarization and transverse helicity fraction — both have a physical range that span $0 \sim 1$. The expected measurement resolution on the degree of polarization, however, is $0.5 \sim 0.7$; for the helicity fraction, it is about 0.3. With this sort of resolution, one cannot place significant limits on the parameter space. In Run II, however, the resolutions are expected to go down to $0.11 \sim 0.16$ for the polarization measurement and 0.06 for the helicity measurement. Thus it should be possible to set meaningful limits then.

Although the measurement resolution is too poor to make any meaningful conclusions about the properties of top quark spin and weak coupling, the resolution for the helicity fraction measurement is good enough to expose an inconsistency in the modeling of the charged lepton energy (E_ℓ) distribution. Specifically, it was found that the change in the measured helicity fraction before and after optimizing the analysis is large; this degree of change is unlikely to be due to a statistical fluctuation. This inconsistency could be due to: (1) incorrect shape of the signal/and or background E_ℓ distribution; (2) incorrect background fraction estimate; or (3) the experimental data are not described correctly by

the standard model $t\bar{t}$ hypothesis. This observation suggests that the comparison of measurements before and after optimization could be a useful technique in Run II for testing the consistency of the signal and background models.

Appendix A

Why No χ^2 Cut Is Made in this Analysis

A.1 Introduction

The event selection criteria used in this analysis to select lepton + jets $t\bar{t}$ candidate events is very similar to those used in the CDF top mass analysis [44]. However, there are a couple of points in which the criteria differ; one of them is the fact that, in this analysis, no cut is made on the $t\bar{t}$ reconstruction χ^2 (see chapter 6 for a definition of χ^2). This choice was made because such a cut does not noticeably improve the measurement resolution of the observables considered in this analysis. The purpose of this appendix is to demonstrate this fact.

A.2 Factors that Determine the Measurement Resolution

In this analysis, a two-component minimum log-likelihood method is used to measure the degree of polarization and the transverse helicity fraction in the experimental data (see chapter 7 for a definition of observables used in this analysis, and see chapter 8 for a discussion of the measurement method and resolution). The measurement resolution of a given observable is given by the following expression:

$$\sigma = K \cdot \frac{\sqrt{S+B}}{S} \tag{A.1}$$

The quantities S and B are the estimated number of signal and background events in the data sample, respectively, and K is a geometric factor whose size is determined by the shape of the component functions:

$$\frac{1}{K^2} = \int dx \, \frac{[f_1(x) - f_0(x)]^2}{f_{exp}(x)} \tag{A.2}$$

The functions f_1 and f_0 are the two component functions the data are being fit to, and f_{exp} is the theoretically expected distribution of the observable x (all functions are normalized to 1). It should be noted that the factor K is small when f_1 and f_0 have small amount of overlap, and *vice versa* when they overlap a lot.

A.3 The Properties χ^2 Must Have

If χ^2 is to be useable as a means of improving the measurement resolution, it must have one or both of the following two properties:

1. The distribution of χ^2 in the signal must differ significantly from that in the background.

2. In signal events, the distribution of χ^2 in events with correct b-quark matching must be significantly different from that in events where the b-quark is mismatched (see section 6.3 and chapter 9 for a definition of "correct" b-quark matching).

Without the first property, the term $\sqrt{S+B}/S$ in equation A.1 would not vary much with χ^2. Without the second, the term K, which is a function of $f_{b\ell}$ (the fraction of signal events with correct b-quark matching), would not vary much. If neither of these terms vary significantly with χ^2, there would be no point in applying a χ^2 cut.

There are reasons to believe that χ^2 possesses these two properties. In chapter 6, it is shown that χ^2 is a measure of how well a given event satisfies the "$t\bar{t}$ hypothesis" — *i.e.* how $t\bar{t}$-like an event is. Thus it is natural to suppose that $t\bar{t}$ events will fit the $t\bar{t}$ hypothesis better than background events, so that χ^2 in $t\bar{t}$ events should be smaller, on average, than in background events. It is also shown in chapter 6 that, for a given event, there are a

range of possible jet-quark combinations, one of which has correct b-quark matching. By choosing the combination with the smallest χ^2, one has a significantly better chance at finding the correct combination than if one were to choose a combination at random. Given this property of χ^2, it is natural to suppose that the fraction of events with correct b-quark matching increases as one eliminates events with large χ^2.

In the following sections, it will be shown that χ^2 does indeed have these two properties. It will also be shown, however, that it does not satisfy the properties to a great enough extent, so that it cannot be used to improve the measurement resolution.

A.4 The Ability of χ^2 to Discriminate between Signal and Background

Figure A.1 shows the efficiency of the cut $\chi^2 < \chi^2_{cut}$ in signal and background events. It is seen that, in all eight subsets of data (see sections 4.4 and 8.2.4 for the reason why the data sample is divided into eight parts), as χ^2_{cut} is decreased from ∞, the efficiency for the background decreases more quickly than for the signal. This demonstrates that fact that a χ^2 cut can be used to purify the event samples.

The effectiveness of a cut in improving the measurement resolution is, however, not determined by the purity ($S/(S+B)$), but by the "significance", defined as $S/\sqrt{S+B}$ (this is the *reciprocal* of the term appearing in equation A.1). If one assumes for the moment that the geometric factor K is constant as a function of χ^2_{cut} (it is not, and this is discussed in the next section), then the value of χ^2_{cut} that minimizes the measurement error is that which maximizes the significance. Figure A.2 shows the normalized significance[1] as a function of χ^2_{cut} in all eight subsamples. In none of the subsamples is there a noticeable maximum for $\chi^2_{cut} < \infty$. This demonstrates the fact that, from the point of view of signal and background differentiation alone, χ^2 cannot be used to improve the measurement resolution.

[1] The normalized significance is defined as significance divided by $1/\sqrt{N_\infty}$, where N_∞ is the number of events in the data sample before applying a χ^2 cut (*i.e.* $\chi^2 < \infty$).

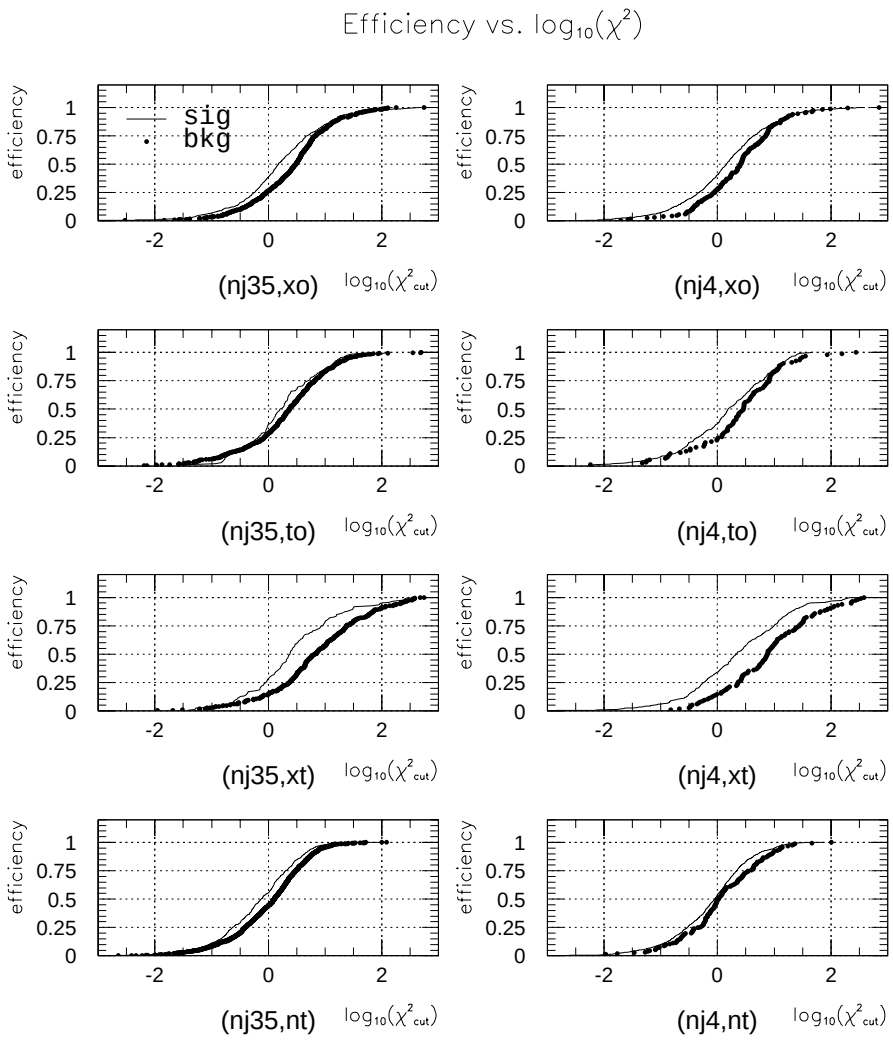


Figure A.1: The efficiency of the signal and background as a function of $\log_{10}(\chi^2_{cut})$. The signal sample is simulated with the HERWIG monte carlo with $m_{top} = 175\ GeV$. The background sample is simulated with the VECBOS monte carlo. See sections 4.4 and 8.2.4 for a description of the labels used to describe the eight subsamples (*i.e.* (NJ35, XO), *etc.*).

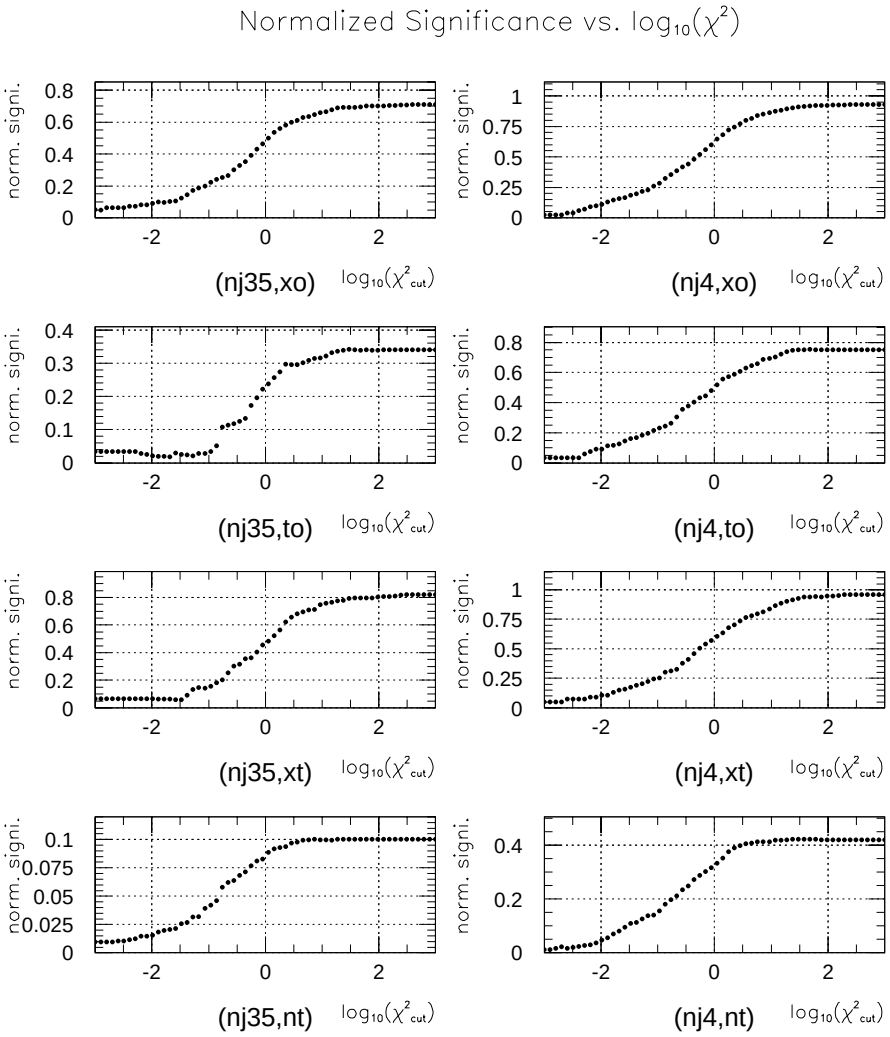


Figure A.2: The normalized significance as a function of $\log_{10}(\chi^2_{cut})$. The signal sample is simulated with the HERWIG monte carlo with $m_{top} = 175\ GeV$. The background sample is simulated with the VECBOS monte carlo. See sections 4.4 and 8.2.4 for a description of the labels used to describe the eight subsamples (*i.e.* (NJ35, XO), *etc.*).

A.5 The Ability of χ^2 to Discriminate between Events with Correct and Incorrect b-quark Matching

Figure A.3 shows the efficiency of signal events with correct ($\Delta R(b_\ell) < 0.4$) and incorrect ($\Delta R(b_\ell) \geq 0.4$) b-quark matching surviving the cut $\chi^2 < \chi^2_{cut}$. It is seen that, in all eight subsamples, the efficiency for the incorrect events decreases more rapidly than for the correct events. This implies that $f_{b\ell}$ increases with tighter χ^2 cut (figure A.4).[2] Since the geometric factor K for all observables is a monotonically decreasing function of $f_{b\ell}$ (figure 9.2 of chapter 9), this means that K achieves a minimum value for some finite value of χ^2_{cut}. Figure A.5 shows an example of this for the observable E_ℓ; the situation with the other observables is almost identical.

If one were to take the K-factor in isolation, then a smaller K-factor means better measurement resolution, since the error σ is proportional to K. However, in order to obtain a measure of the efficacy of the χ^2 cut in improving the measurement resolution, one must take into account the cost in statistics that is needed to allow K to decrease. A useful quantification of this efficacy is the following normalized significance:[3]

$$\frac{1}{K(\chi^2_{cut})} \cdot \sqrt{\frac{S_<(\chi^2_{cut}) + S_\geq(\chi^2_{cut})}{N_\infty}} \tag{A.3}$$

The quantity $S_<$ is the number of signal events with $\Delta R(b_\ell) < 0.4$ passing the cut $\chi^2 < \chi^2_{cut}$, $S_\geq$ is the corresponding number for events with $\Delta R(b_\ell) \geq 0.4$, and N_∞ is the total number of signal events before any cut is applied. Figure A.6 shows the χ^2_{cut}-dependence of the normalized significance for the observable E_ℓ for all eight subdivisions of data. A local maximum in the curve would indicate improved measurement resolution. None is seen, so a χ^2 cut is not efficacious. Almost identical results also hold for the other observables in this analysis.

[2] It increases up to a point. When the cut gets so tight that few events remain, statistical fluctuations make the true value of $f_{b\ell}$ unclear.

[3] This expression is the normalized significance taking the background fraction to be zero. The background fraction is set to zero here in order to isolate the effect of b-quark matching on the measurement resolution, as opposed to the effect of signal-to-background ratio, which was discussed in the last section. The significance that takes account both effects is discussed in the next section.

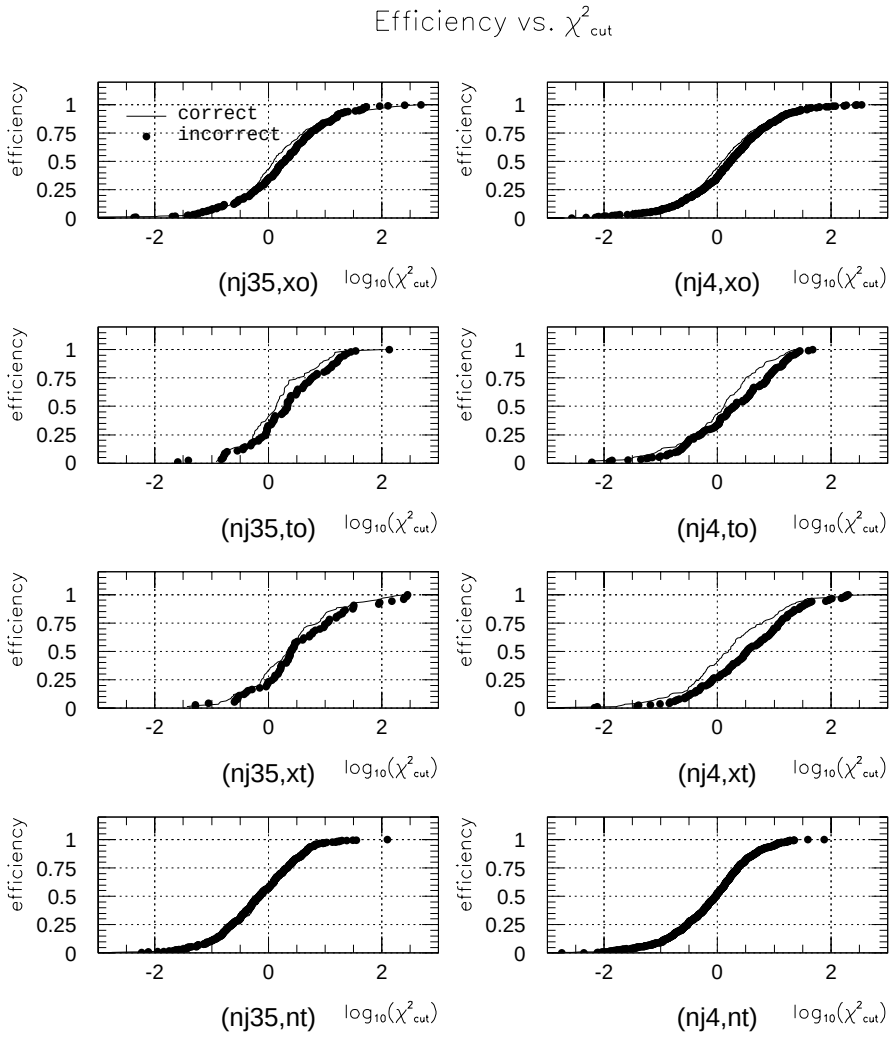


Figure A.3: The efficiency of signal events with correct and incorrect b-quark matching for the cut $\chi^2 < \chi^2_{cut}$. See sections 4.4 and 8.2.4 for a description of the labels used to describe the eight subsamples (*i.e.* (NJ35, XO), *etc.*).

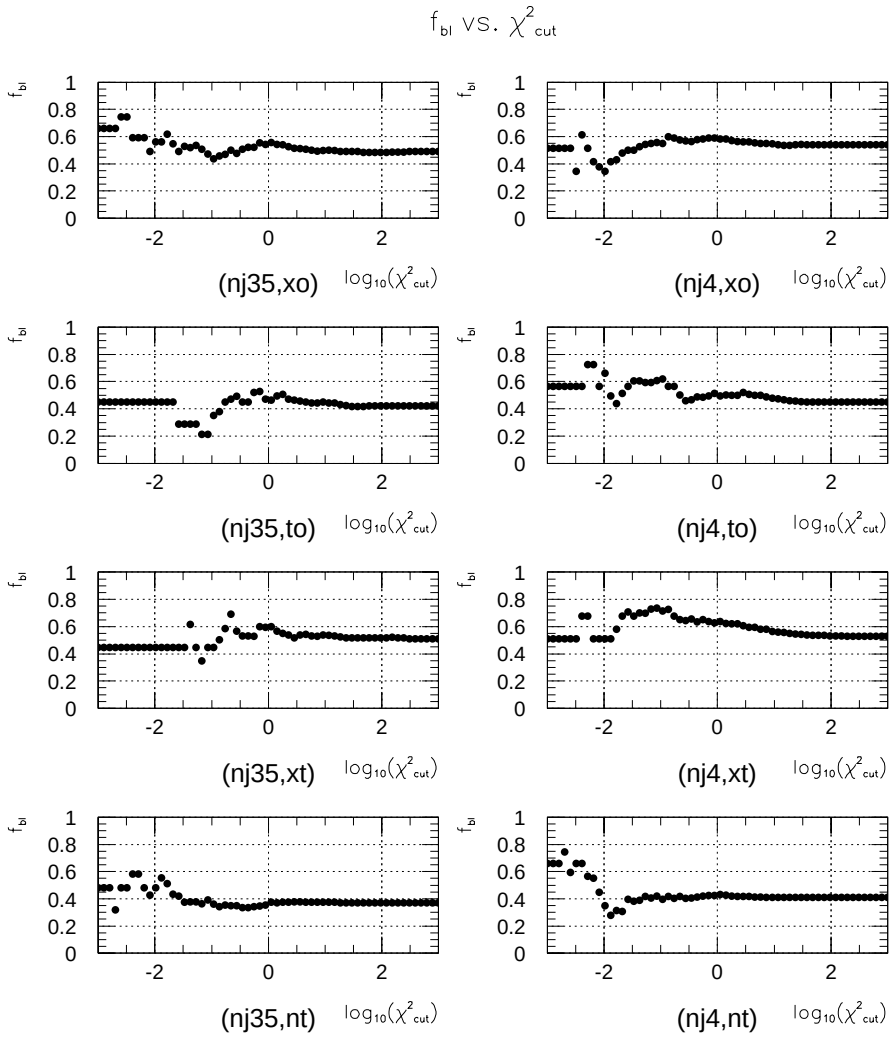


Figure A.4: The change in $f_{b\ell}$ with χ^2_{cut}. See sections 4.4 and 8.2.4 for a description of the labels used to describe the eight subsamples (*i.e.* (NJ35, XO), *etc.*).

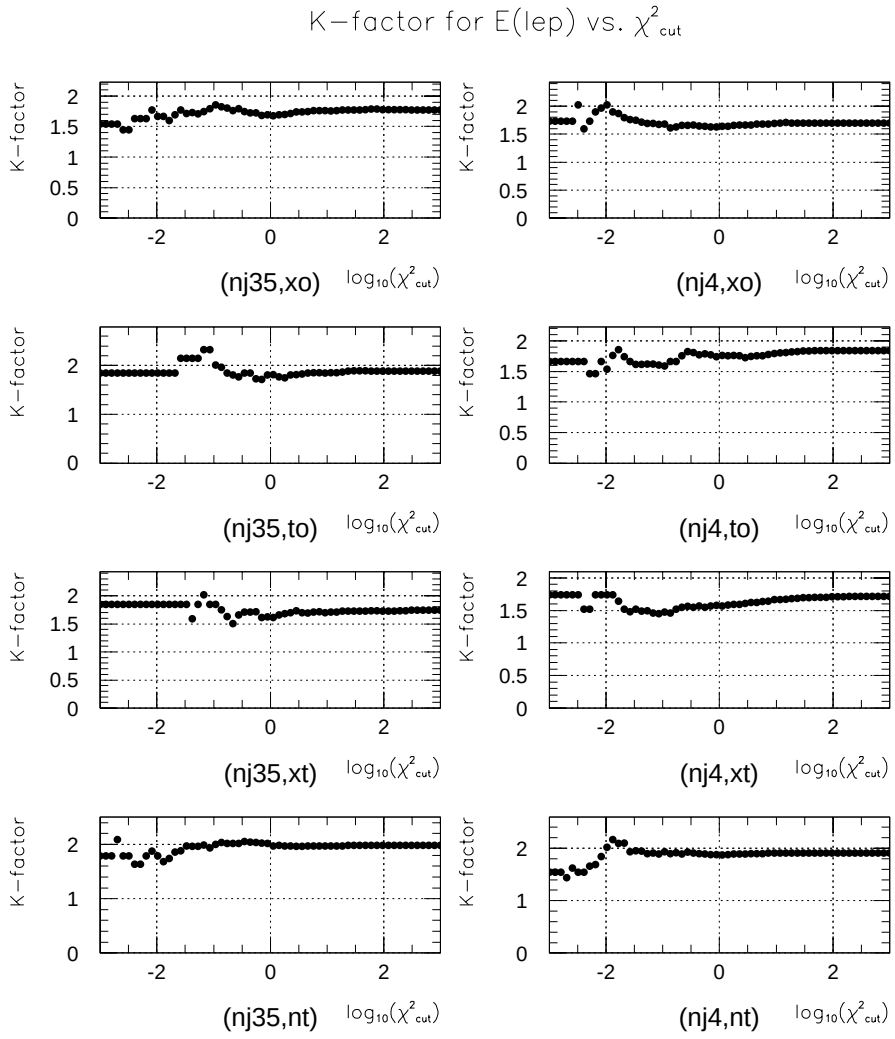


Figure A.5: The change in K-factor for the observable E_ℓ with χ^2_{cut}. See sections 4.4 and 8.2.4 for a description of the labels used to describe the eight subsamples (*i.e.* (NJ35, XO), *etc.*).

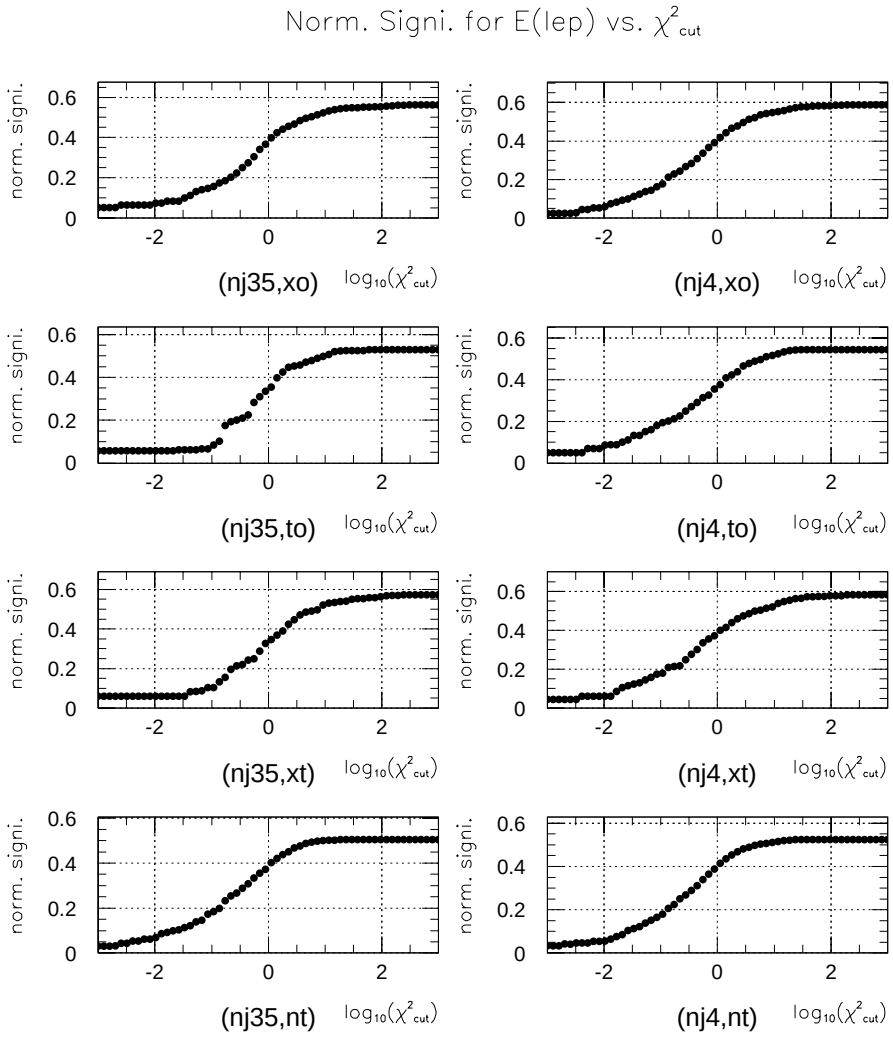


Figure A.6: The normalized significance for E_ℓ as a function of χ^2_{cut}. See sections 4.4 and 8.2.4 for a description of the labels used to describe the eight subsamples (*i.e.* (NJ35, XO), *etc.*).

A.6 Demonstrating the Fact that χ^2 Cannot Be Used to Improve the Measurement Resolution

In the last two sections, it was shown that a χ^2 cut does not improve the measurement resolution when viewing the problem from: (1) a purely signal-to-background issue alone; and (2) a purely geometric issue alone. It is possible, however, that if the two factors are combined, a net improvement in resolution might be seen. Figure A.7 shows that this is not the case for the observable E_ℓ. The figure is a plot of the normalized significance $(1/(\sigma \cdot \sqrt{N_\infty}))$ as a function of χ^2_{cut}, where σ is given in equation A.1 and N_∞ is the number of events in the data sample before any χ^2 cut. A local maximum in any of the plots would indicate improved measurement resolution: none is seen. Almost identical results also hold for the other observables in this analysis.

A.7 Conclusion

Let us summarize the results obtained in this appendix. First, it was shown that χ^2 has properties that tend to improve the measurement resolution. Specifically, χ^2 can be used to separate the signal from the background to a certain extent. Also, χ^2 can be used to improve, to a certain extent, the fraction of signal events with correct b-quark matching. Both of these properties tend to improve the measurement resolution. However, these tendencies must be weighed against the cost in lost statistics when applying χ^2 cuts. Cost-benefit analyses show that the cost in statistics outweigh the benefits in improved signal purity and b-quark matching fraction. This is the reason why no χ^2 cut is applied in this analysis.

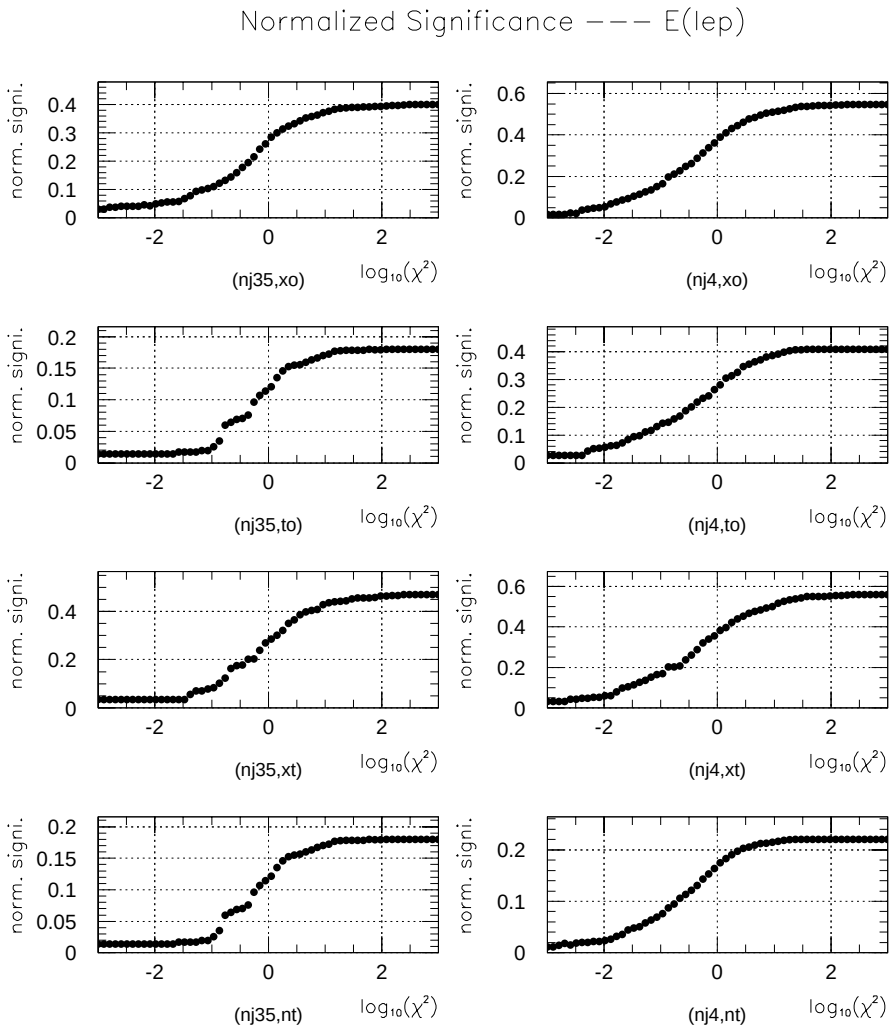


Figure A.7: The *complete* normalized significance (*i.e.* the significance taking into account both the signal-to-background and the geometric issues) for E_ℓ as a function of χ^2_{cut}. See sections 4.4 and 8.2.4 for a description of the labels used to describe the eight subsamples (*i.e.* (NJ35, XO), *etc.*).

Appendix B

Complex Solutions of $P_z(\nu)$

B.1 Introduction

In section 6.2.5, it was stated that the initial estimate of the neutrino's longitudinal momentum $P_z(\nu)$ can, under certain circumstances, be complex. These situations are dealt with by taking the real part of $P_z(\nu)$ and adding $\pm 20\ GeV$ to it:

$$P_z(\nu) \ = A \pm 20\, GeV \tag{B.1}$$

A is the real part of the complex $P_z(\nu)$, and is defined in equation 6.3. This prescription for dealing with complex solutions of $P_z(\nu)$ may seem *ad hoc*. In fact, however, the details of how real numbers are assigned in place of complex solutions are unimportant. The only requirement that any such procedure must satisfy is this: the two real replacement values must be chosen so that when χ^2 has two minima, the χ^2 minimization process will find them both. In this appendix, complex solutions of $P_z(\nu)$ and how to deal with them is discussed.

B.2 A Graphical Description of Real and Complex $P_z(\nu)$

The prescription for obtaining the initial estimate of $P_z(\nu)$ — *i.e.* choose $P_z(\nu)$ such that the invariant mass of the ℓ-ν_ℓ system $m_{\ell\nu} = M_W$ — can be described graphically as

follows. First, consider the set of all pairs $(P_z(\nu), P_T(\nu))$ such that $m_{\ell\nu} = M_W$. These points describe a tilted ellipse in the $P_z(\nu)$-$P_T(\nu)$ plane. The size and tilt of the ellipse are fixed by the following parameters:

1. E_ℓ, the charged lepton energy.

2. θ_ℓ, the charged lepton polar angle (*i.e.* angle relative to the beam).

3. $\Delta\phi_{\ell\text{-}\nu}$, the azimuthal angular separation between ℓ and ν_ℓ.

All of the above are defined in the lab frame. The following equation describes the ellipse:

$$\begin{aligned} y - y_c &= \left(\frac{T\cos\theta_\ell}{1-T^2}\right)(x - x_c) \\ &\pm \left(\frac{1}{1-T^2}\right)\sqrt{\left(\frac{1-T^2}{\sin^2\Delta\phi_{\ell\text{-}\nu}\sin^2\theta_\ell}\right)\varepsilon^2 - \sin^2\Delta\phi_{\ell\text{-}\nu}\sin^2\theta_\ell(x - x_c)^2} \end{aligned} \tag{B.2}$$

$$T = \cos\Delta\phi_{\ell\text{-}\nu}\sin\theta_\ell \tag{B.3}$$

$$x_c = \left(\frac{1}{\sin^2\Delta\phi_{\ell\text{-}\nu}}\right) m_x\,\varepsilon \tag{B.4}$$

$$y_c = \left(\frac{\cos\Delta\phi_{\ell\text{-}\nu}}{\sin^2\Delta\phi_{\ell\text{-}\nu}}\right) m_y\,\varepsilon \tag{B.5}$$

$$\varepsilon = \frac{M_W^2}{2E_\ell} \tag{B.6}$$

$$m_x = \frac{\cos\theta_\ell}{\sin^2\theta_\ell} \tag{B.7}$$

$$m_y = \frac{1}{\sin\theta_\ell} \tag{B.8}$$

In the above equations, y and x stand for $P_T(\nu)$ and $P_z(\nu)$, respectively. The ordered pair (x_c, y_c) gives the coordinate of the center of the ellipse in the $P_z(\nu)$-$P_T(\nu)$ plane. The parameters m_x and m_y determine the *tilt* of the ellipse's major and minor axes. The

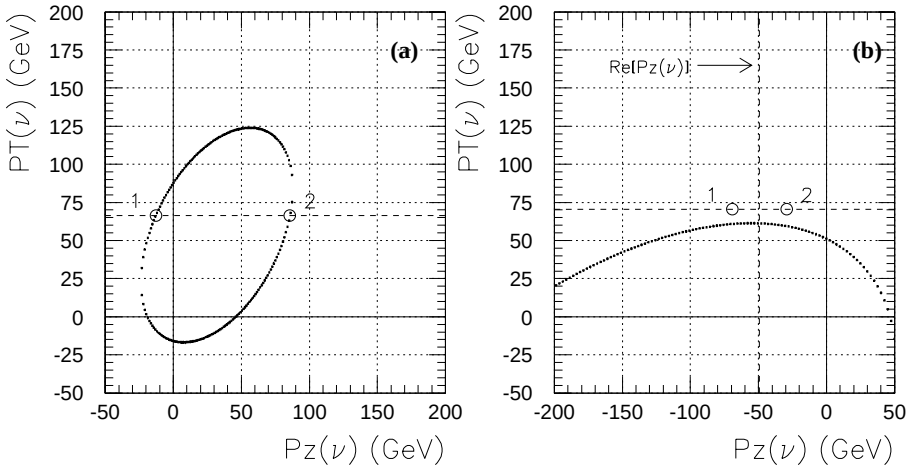


Figure B.1: The $P_z(\nu)$-$P_T(\nu)$ ellipse from: (a) Run/Event = 40758/044414; and (b) Run/Event = 43096/047223. In both figures, the dashed horizontal line corresponds to the initial estimate of $\left|\vec{P}_T(\nu)\right|$. In (a), the values of $P_z(\nu)$ at the intersection points of the line with the ellipse, labeled "1" and "2", are the initial estimates of the neutrino longitudinal momentum. In (b), the horizontal line does not intersect the ellipse, so the $P_z(\nu)$ solutions are complex. The value of $P_z(\nu)$ at points labeled "1" and "2" are obtained by adding $\pm 20\ GeV$ to the real part of the complex solution (dashed vertical line). These values are chosen as the initial estimate of the neutrino longitudinal momentum.

parameter ε has the unit of energy, and it determines the size of the ellipse; when ε is large, so is the ellipse, and *vice versa* when ε is small. It should be noted that ε is inversely proportional to E_ℓ. Thus, for example, when E_ℓ is large, ε is small, and therefore $P_T(\nu)$ and $P_z(\nu)$ tend to be small. This inverse relation between the energy of ℓ and ν_ℓ is necessary in order for the invariant mass of the ℓ-ν_ℓ system to be fixed at M_W.

Figure B.1 shows ellipses from two events collected at CDF during Tevatron Collider Run I. The horizontal dashed line in each diagram represents the initial estimate of $\left|\vec{P}_T(\nu)\right|$. In figure B.1(a), the horizontal line intersects the ellipse at two values of $P_z(\nu)$, labeled "1" and "2". These values are chosen as the initial estimates of $P_z(\nu)$ by the $t\bar{t}$ reconstruction algorithm. In figure B.1(b), the horizontal line does not intersect the ellipse. In such cases, the solutions of $P_z(\nu)$ are complex. The $P_z(\nu)$ at points labeled "1" and "2" in figure B.1(b) are obtained by adding $\pm 20\ GeV$ to the real part of the complex solution, indicated by the vertical dashed line.

B.3 The $P_z(\nu)$-$P_T(\nu)$ Ellipse during and after the χ^2 Minimization Procedure

In order to determine whether the seemingly *ad hoc* rule in equation B.1 for dealing with complex $P_z(\nu)$ solutions is sufficient or not, one needs to understand what happens to the $P_z(\nu)$-$P_T(\nu)$ ellipse during and after the χ^2 minimization procedure. The $t\bar{t}$ reconstruction algorithm stretches/shrinks the energy of jets, charged lepton, and unclustered energy in order to minimize χ^2. Since the charged lepton momentum is usually well measured, its estimated error is small, and thus its energy is not allowed to stretch/shrink very much. Thus, to a good approximation, one can take the charged lepton 4-momentum to be constant during the χ^2 minimization procedure. The jet and unclustered energy errors, however, are sizable. Their energies, therefore, can be stretched/shrunk to a considerable degree. This stretching/shrinking propagates directly to the direction and magnitude of $\vec{P}_T(\nu)$. Therefore, during the χ^2 minimization procedure, the following take place:

- The horizontal line corresponding to $\left|\vec{P}_T(\nu)\right|$ can be adjusted up and down.
- Of the three parameters that describe the $P_z(\nu)$-$P_T(\nu)$ ellipse, only $\Delta\phi_{\ell\text{-}\nu}$ can be changed significantly; E_ℓ and θ_ℓ essentially stay fixed. The change in $\Delta\phi_{\ell\text{-}\nu}$ changes the size, position, and tilt of the ellipse. See section B.4 for details on how the ellipse changes with $\Delta\phi_{\ell\text{-}\nu}$.

After the χ^2 minimization process is complete, the location of the horizontal line representing $\left|\vec{P}_T(\nu)\right|$ and the location, size, and tilt of the $P_z(\nu)$-$P_T(\nu)$ ellipse are, to varying degrees, different from those at the beginning of the process. See figure B.2.

The final result of the χ^2 minimization process can be classified as follows:

Non-degenerate — χ^2 has Two Minima

An example of this is shown in figure B.2(a). The pair $(P_z(\nu), P_T(\nu))$ for the two minima are usually close to the initial positions in the $P_z(\nu)$-$P_T(\nu)$ plane.

Degenerate — χ^2 has One Minimum

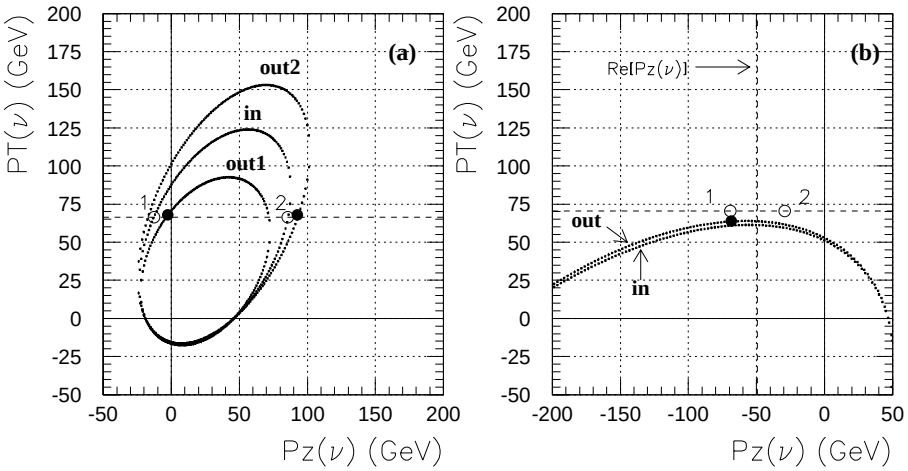


Figure B.2: (a) and (b) are from the same events as in figure B.1. In (a), the ellipse labeled "in" is that at the beginning of the χ^2 minimization process, while those labeled "out1" and "out2" correspond to the two minima of χ^2. In (b), the ellipse labeled "in" is that at the beginning of the χ^2 minimization, and the one labeled "out" correspond to the single, degenerate minimum of χ^2. The open circles labeled "1" and "2" in both diagrams indicate the two pairs $(P_z(\nu), P_z(\nu))$ at the beginning of χ^2 minimization. The filled circles correspond to $(P_z(\nu), P_z(\nu))$ at the end of minimization.

An example of this is shown in figure B.2(b). Both of the initial estimates of $(P_z(\nu), P_T(\nu))$ converge to a single solution.

The relevance of the existence of these two categories of solutions on the issue of complex $P_z(\nu)$ is the following: the prescription for assigning real numbers in place of complex $P_z(\nu)$ solutions matters only in events with non-degenerate χ^2. This is because, in degenerate events, no matter what real numbers one assigns in place of the complex solutions of $P_z(\nu)$, the final output of the $t\bar{t}$ reconstruction algorithm converges to a single configuration corresponding to the unique minimum χ^2. Of the events with complex $P_z(\nu)$, 80% of events have degenerate χ^2. Thus the problem of complex $P_z(\nu)$ matters in 20% of such events.

When χ^2 has non-degenerate solutions, the method for dealing with complex $P_z(\nu)$ must be chosen with some care. Figure B.3 is helpful in making this point. This figure shows the $P_z(\nu)$-$P_T(\nu)$ ellipses before and after minimizing χ^2: the lower one is before, and the upper ones are after (there are two closely spaced ellipses, one for each minimum). The dashed horizontal line corresponds to the input $\left|\vec{P}_T(\nu)\right|$. Since this line does not intersect the lower ellipse, the initial estimate of $P_z(\nu)$ is complex. The open circles labeled "1" and "2" are the real replacement values for the complex solutions, while the vertical dashed line indicates the real part of the complex $P_z(\nu)$ solutions. The closed circles labeled "out1" and "out2" are the location of $(P_z(\nu), P_T(\nu))$ at the minima of χ^2.

Figure B.3 gives an intuitive feel for why the rule for dealing with complex $P_z(\nu)$ solutions must be chosen with some care. For instance, if the points "1" and "2" were placed too close together, it is possible that the two initial configurations will converge to only one of the two minima. Also, keeping the $P_z(\nu)$ of "1" and "2" separated does not guarantee that the two initial configurations will converge to the two minima. For instance, if $P_z(\nu)$ for "1" and "2" are assigned $-120\ GeV$ and $-80\ GeV$, these initial configurations will probably both converge to "out1", even though the two $P_z(\nu)$ are separated by $40\ GeV$, which is the same as for the rule $P_z(\nu) = A \pm 20\ GeV$.

The author is not aware of any method of dealing with complex $P_z(\nu)$ solutions that will guarantee that both minima in non-degenerate χ^2 will be found. The method $P_z(\nu) = A \pm 20\ GeV$, however, has been shown in monte carlo studies to always find both minima. Thus, if this method does fail at all, its rate of failure must be very small.

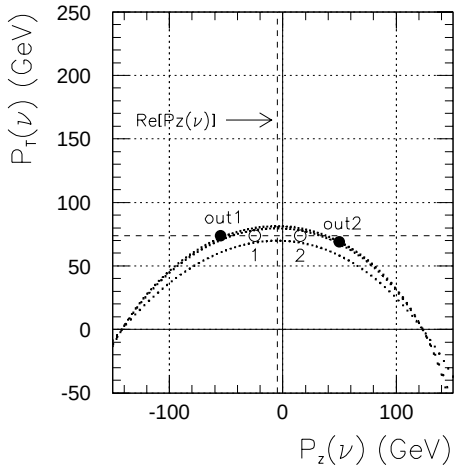


Figure B.3: From Run/Event = 43351/266423. The lower ellipse is for the initial configuration. Two closely spaced ellipses lie above the lower one, one ellipse for each minimum in χ^2. The horizontal dashed line indicates the initial $\left|\vec{P}_T(\nu)\right|$. The vertical dashed line indicates the real part of the complex $P_z(\nu)$ solutions. The open circles indicate the initial estimates of $(P_z(\nu), P_T(\nu))$, while the solid circles indicate $(P_z(\nu), P_T(\nu))$ at the χ^2 minima.

Let us summarize what has been learned: (1) the choice of real replacement values for complex $P_z(\nu)$ matters in only 20% of events with complex $P_z(\nu)$; (2) monte carlo studies show that the prescription $P_z(\nu) = A \pm 20\ GeV$ allows the $t\bar{t}$ reconstruction algorithm to find both χ^2 minima with an efficiency very close to 100%; (3) this prescription is, admittedly, *ad hoc*, but it works.

B.4 The Change in the $P_z(\nu)$-$P_T(\nu)$ Ellipse with $\Delta\phi_{\ell\text{-}\nu}$

This section is an addendum to this appendix. It describes how the $P_z(\nu)$-$P_T(\nu)$ ellipse changes with $\Delta\phi_{\ell\text{-}\nu}$.

In section B.2, it was shown that, during the χ^2 minimization process, the jet and unclustered energies have considerable freedom to stretch/shrink, while the charged lepton remains essentially fixed. Because of this, even though many physics objects can be stretched/shrunk in many different ways, the effect of such stretching/shrinking on the $P_z(\nu)$-$P_T(\nu)$ ellipse is described by the single parameter $\Delta\phi_{\ell\text{-}\nu}$. In other words, even though

the stretching/shrinking of physics object energies takes place in a space of many dimensions, the resulting change in location, size, and tilt of the ellipse is described by a variable in a 1-dimensional space.

Equation B.2 describes an ellipse whose major and minor axes are, in general, tilted relative to the $P_z(\nu)$-$P_T(\nu)$ axes. Although these axes can be used to describe the ellipse, it is more useful in this context to describe it by the coordinates where the value of $P_T(\nu)$ is maximum and minimum — *i.e.* where the slope of the ellipse is zero. The point where $P_T(\nu)$ is maximum will be referred to by (x_{max}, y_{max}), while (x_{min}, y_{min}) refers to the point where $P_T(\nu)$ is minimum. The equations for these points are given as follows:

$$x_{max} = (1 + \cos\Delta\phi_{\ell\text{-}\nu})x_c \tag{B.9}$$

$$y_{max} = \left(\frac{1 + \cos\Delta\phi_{\ell\text{-}\nu}}{\cos\Delta\phi_{\ell\text{-}\nu}}\right) y_c \tag{B.10}$$

$$x_{min} = (1 - \cos\Delta\phi_{\ell\text{-}\nu})x_c \tag{B.11}$$

$$y_{min} = -\left(\frac{1 - \cos\Delta\phi_{\ell\text{-}\nu}}{\cos\Delta\phi_{\ell\text{-}\nu}}\right) y_c \tag{B.12}$$

These equations are useful in deducing how the ellipse moves in the $P_z(\nu)$-$P_T(\nu)$ plane as $\Delta\phi_{\ell\text{-}\nu}$ changes. The following are some observations:

- Assume that $\cos\theta_\ell > 0$. Then, as $\Delta\phi_{\ell\text{-}\nu}$ changes, the center of the ellipse (x_c, y_c) moves along an hyperbola defined by the following equation:

  $$y_c = \pm\frac{m_y}{m_x}\sqrt{x_c\,[x_c - m_x\,\varepsilon]} \tag{B.13}$$

  The positive branch of the hyperbola is chosen when $\Delta\phi_{\ell\text{-}\nu} < 90^o$, while the negative one is chosen when $\Delta\phi_{\ell\text{-}\nu} > 90^o$. See figure B.4(a).

- Assume that $\cos\theta_\ell > 0$. Then the maximum point (x_{max}, y_{max}) moves along the following line:

  $$y_{max} = \left(\frac{m_y}{m_x}\right) x_{max} \tag{B.14}$$

Similarly, the minimum point (x_{min}, y_{min}) moves along the following line:

$$y_{min} = -\left(\frac{m_y}{m_x}\right) x_{min} \tag{B.15}$$

See figure B.4(a).

- When $\cos\theta_\ell < 0$, the lines and curves described above are reflected about the vertical line $P_z(\nu) = 0$. See figure B.4(a).

- The curves described above give the motion of (x_c, y_c), (x_{max}, y_{max}), and (x_{min}, y_{min}) as $\Delta\phi_{\ell\text{-}\nu}$ changes, but it does not give any information about how the points move relative to each other. The following equations fills in this gap:

$$\frac{x_{max}}{x_c} = 1 + \cos\Delta\phi_{\ell\text{-}\nu} \tag{B.16}$$

$$\frac{y_{max}}{y_c} = \frac{1 + \cos\Delta\phi_{\ell\text{-}\nu}}{\cos\Delta\phi_{\ell\text{-}\nu}} \tag{B.17}$$

$$\frac{x_{min}}{x_c} = 1 - \cos\Delta\phi_{\ell\text{-}\nu} \tag{B.18}$$

$$\frac{y_{min}}{y_c} = -\frac{1 - \cos\Delta\phi_{\ell\text{-}\nu}}{\cos\Delta\phi_{\ell\text{-}\nu}} \tag{B.19}$$

- In the limit $\Delta\phi_{\ell\text{-}\nu} \to 0$, $(x_{max}, y_{max}) \to (2x_c, 2y_c)$.

- In the limit $\Delta\phi_{\ell\text{-}\nu} \to 0$, the minimum point converges to $(x_{min}, y_{min}) = (\frac{1}{2}m_x\,\varepsilon, -\frac{1}{2}m_y\,\varepsilon)$.

- In the limit $\Delta\phi_{\ell\text{-}\nu} \to 180^o$, $(x_{max}, y_{max}) \to (2x_c, -2y_c)$.

- In the limit $\Delta\phi_{\ell\text{-}\nu} \to 180^o$, the minimum point converges to $(x_{min}, y_{min}) = (\frac{1}{2}m_x\,\varepsilon, \frac{1}{2}m_y\,\varepsilon)$.

- Figure B.4(b) shows the $P_z(\nu)$-$P_T(\nu)$ ellipse for $\Delta\phi_{\ell\text{-}\nu} = 170^o$, 135^o, 90^o, 45^o, and 10^o. E_ℓ and θ_ℓ are fixed at 120.2 GeV and 65.5^o, respectively (these values are taken from Run/Event = 40758/044414 — see figure B.1(a)). As $\Delta\phi_{\ell\text{-}\nu}$ changes from $180^o \to 90^o$, the ellipse moves toward the upper left, while as $\Delta\phi_{\ell\text{-}\nu}$ goes from $90^o \to 0^o$, the ellipse moves toward the upper right. At $\Delta\phi_{\ell\text{-}\nu} = 90^o$, the ellipse is circular and is divided into equal parts by the line $P_T(\nu) = 0$.

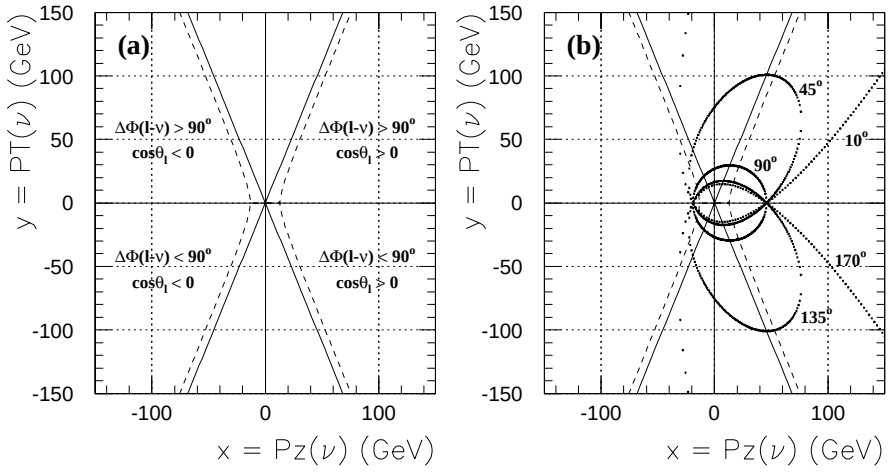


Figure B.4: (a) The dashed curve shows the location of the center of the ellipse (x_c, y_c) as $\Delta\phi_{\ell\text{-}\nu}$ is varied. The straight lines in the upper half-plane ($y > 0$) shows the location of the maximum point (x_{max}, y_{max}) of the ellipse as $\Delta\phi_{\ell\text{-}\nu}$ is varied. The straight lines in the lower half-plane is for the minimum point (x_{min}, y_{min}). (b) The ellipse with $\Delta\phi_{\ell\text{-}\nu} = 170^o$, 135^o, 90^o, 45^o, and 10^o. For both (a) and (b), $E_\ell = 120.2\ GeV$ and $\theta_\ell = 65.5^o$ (from Run 40758 / Event 44414).

Appendix C

The Unclustered Energy

C.1 Introduction

In section 6.2.6, the unclustered energy was described as being that part of the transverse energy deposited in the calorimeter that cannot be accounted for by jets and the charged lepton. This appendix describes how the unclustered energy is estimated using the physics object momenta.

C.2 The Unclustered Energy at Three Stages in the $t\bar{t}$ Reconstruction Algorithm

The definition of unclustered energy depends on the different stages of the $t\bar{t}$ reconstruction algorithm. These stages are:

Raw

This is the stage where the physics object momenta have minimal corrections applied. The unclustered energy in this stage is defined mostly in terms of the raw physics object momenta.

Input

This is the stage at the beginning of the χ^2 minimization process (see appendix D for

information on χ^2). All of the physics object momenta have full corrections applied. The momenta at this stage can be thought of as the initial guess of the parton-level momenta. The unclustered energy in this stage is defined in terms of the raw unclustered energy, with some modifications from the underlying event and out-of-cone corrections for each jet.

Output

This is the final stage of the $t\bar{t}$ reconstruction algorithm, where the χ^2 for all possible configurations have been minimized by stretching/shrinking physics object momenta. The two transverse components of the unclustered energy are independently stretched/shrunk during the χ^2 minimization process. The amount of stretching/shrinking depends on the estimated error on the unclustered energy.

In this appendix, the definition of the unclustered energy at the raw and input stages is discussed. There is no need to define the unclustered energy at the output level because it is simply a stretched/shrunk version of that at the input level.

C.3 The Definition of the Raw Unclustered Energy

At the raw stage, the unclustered energy is defined to be a two-component vector that balances the sum of the transverse momentum vectors of all raw-level physics objects:

$$\vec{P}_T(uce, raw) = -\left[\vec{P}_T(\ell) + \vec{\not\!\!E}_T + \sum_{i=1}^{Njets} \vec{P}_T(i)\right] \qquad \text{(C.1)}$$

The momentum vectors for the jets have no corrections applied[1], while that for ℓ is fully corrected.[2] The missing transverse energy vector $\vec{\not\!\!E}_T$ is given by the raw value[3] with primary lepton corrections applied. See chapter 3 for details on the the identification and momentum reconstruction of the physics objects.

[1] In CDF jargon, the momentum of these uncorrected jet momenta are obtained from the JETS bank.

[2] There is no significance in the fact that the fully corrected charged lepton momentum is used here instead of the raw one. The definition of the unclustered energy does not change by much if the raw momentum were used instead. The particular choice given here is just one of many somewhat arbitrary choices involved in defining the unclustered energy.

[3] In CDF jargon, the raw missing transverse energy vector is obtained from the DETS bank.

Physically, what does the raw unclustered energy represent? One way to answer this is to consider an hypothetical situation where the raw-level momentum of each object has no measurement error, and all parton-level objects are identified with 100% efficiency. Then the raw $\vec{P}_T(uce, raw)$ would be equal to the small transverse kick in the $t\bar{t}$ system due to non-perturbative processes. In practice, $\left|\vec{P}_T(uce, raw)\right|$ is considerably larger than the predicted transverse kick because of the finite resolution of physics object momenta the possible presence of soft jets that fail jet acceptance cuts or are not identified by the jet reconstruction algorithm.

To summarize, the sources of the unclustered energy are:

1. Particles responsible for the non-perturbative kick in the $t\bar{t}$ system
2. Soft jets
3. Finite resolution of physics object momenta

The first two sources are referred to as *physical* because the energy deposited in the calorimeters by actual particles gives rise to $\vec{P}_T(uce)$. The last source is referred to as *resolution* because it contributes to $\vec{P}_T(uce)$ by virtue of mismeasurement of the physics object momenta.

C.4 The Definition of the Input Unclustered Energy

The definition of the unclustered energy at the beginning of the χ^2 minimization process is this:

$$\begin{aligned} \vec{P}_T(uce, in) &= 1.6 \cdot \vec{P}_T(uce, raw) \\ &\quad + \vec{P}_T(\text{underlying event}) \\ &\quad - \vec{P}_T(\text{out-of-cone}) \end{aligned} \tag{C.2}$$

The factor 1.6 in front of the raw unclustered energy is the average correction factor applied to low-energy jets ($E_T \sim 8\ GeV$)[44]. The assumption underlying the choice of this

factor is that the unclustered energy is vaguely similar to a low-energy jet. The quantities $\vec{P}_T$(underlying event) and $\vec{P}_T$(out-of-cone) are defined as follows:

$\vec{P}_T$(underlying event)

The underlying event can be thought of, basically, as the debris from the spectator particles in p and $\overline{p}$. Some of the particles that form this debris fly into the cones that define jets, and thus make spurious contributions to the jet energies. The standard jet correction routine described in chapter 3 estimates the average underlying event contribution to each jet E_T, and subtracts this from each jet. The E_T subtracted from each jet is then added to the unclustered energy because the underlying event is a part of the physical sources that comprise the unclustered energy. The quantity $\vec{P}_T$(underlying event) is the vector sum of the underlying event E_T contribution to each jet.

$\vec{P}_T$(out-of-cone)

It was mentioned in section 3.3.1 that jets are identified with the fixed-cone jet clustering algorithm, with $\Delta R = 0.4$ in this analysis. The use of a fixed cone size entails the loss of some of the original parton-level energy outside of the cone. The escaped energy ends up being counted as part of the unclustered energy. The standard jet correction routine estimates an average amount of escaped energy as a function of the jet E_T, and adds this to the raw jet energy. To conserve energy, the amount added to each jet is subtracted from the unclustered energy. The quantity $\vec{P}_T$(out-of-cone) is the vector sum of the out-of-cone E_T contribution to each jet.

Appendix D

The Definition of χ^2

The quantity χ^2 consists of many terms which can be placed into two categories: the *mass terms* and the *kinematic terms*:

$$\chi^2 = \chi^2_{mass} + \chi^2_{kinematics} \tag{D.1}$$

The mass terms are defined so that configurations with event topology consistent with that of $t\bar{t}$ events in the lepton + jets decay channel have small χ^2, and *vice versa* for configurations whose event topology is inconsistent with that of $t\bar{t}$. The kinematic terms are defined so that they are all zero when the physics object energies are all at their initial estimated value, while they increase as the energies are changed from the initial value.

The quantity χ^2_{mass} is composed of four terms:

$$\chi^2_{mass} = \chi^2(t_\ell) + \chi^2(t_h) + \chi^2(W_\ell) + \chi^2(W_h) \tag{D.2}$$

$$\chi^2(t_\ell) = \left(\frac{m_{\ell\nu b} - MTOP}{\Gamma_{top}}\right)^2 \tag{D.3}$$

$$\chi^2(t_h) = \left(\frac{m_{jjb} - MTOP}{\Gamma_{top}}\right)^2 \tag{D.4}$$

$$\chi^2(W_\ell) = \left(\frac{m_{\ell\nu} - M_W}{\Gamma_W}\right)^2 \tag{D.5}$$

$$\chi^2(W_h) = \left(\frac{m_{jj} - M_W}{\Gamma_W}\right)^2 \tag{D.6}$$

The symbols appearing in the above equations are described below:

$m_{\ell\nu b}$	The invariant mass of the ℓ-ν_ℓ-b_ℓ system — *i.e.* the three particle system presumed to originate from the decay of t_ℓ.
m_{jjb}	The invariant mass of the 3-jet system presumed to originate from the decay of t_h.
$m_{\ell\nu}$	The invariant mass of the ℓ-ν_ℓ system.
m_{jj}	The invariant mass of the 2-jet system presumed to originate from the hadronically decaying W.
$MTOP$	One of the fit parameters. The value of $MTOP$ at the χ^2 absolute minimum is taken as the output top mass.
M_W	A constant equal 80.41 GeV, the world-average W mass [60].
Γ_t	A constant equal to 2.5 GeV. This is comparable to the physical top quark width, which is 1.4 GeV (see section 7.3). The specific value is based on monte carlo studies of $t\bar{t}$ reconstruction.
Γ_W	A constant equal to 2.12 GeV. This is the world-average W width [60].

The following are some comments on χ^2_{mass}:

- The top mass terms $\chi^2(t_\ell)$ and $\chi^2(t_h)$ are related to each other through the fit parameter $MTOP$. Because of this, these terms favor configurations where $m_{\ell\nu b}$ and m_{jjb} are close together. Thus the combination $\chi^2(t_\ell)+\chi^2(t_h)$ could have been replaced by the following:

$$\frac{1}{2}\left(\frac{m_{\ell\nu b}-m_{jjb}}{\Gamma_{top}}\right)^2 \qquad \text{(D.7)}$$

This equation is obtained from $\chi^2(t_\ell)+\chi^2(t_h)$ by replacing $MTOP$ with $0.5\cdot(m_{\ell\nu b}+m_{jjb})$. The use of this alternative top mass condition does not alter the result of the algorithm by very much. The top mass condition that uses $MTOP$ is chosen for historical reasons — if $MTOP$ is a fit parameter, then a minimization routine such as CERN's MINUIT [59] can give properties (*e.g.* the error matrix) of the region

surrounding the absolute minimum of χ^2. This information was useful in the early stages of the development of the $t\bar{t}$ reconstruction algorithm.

- The term $\chi^2(W_\ell)$ is determined by the 4-momenta of ℓ and ν_ℓ. The 4-momentum of ℓ is usually well measured, so it can be considered as being fixed throughout the χ^2 minimization procedure. Thus the value of $\chi^2(W_\ell)$ is almost completely determined by the 4-momentum of ν_ℓ. The transverse component of the momentum of ν_ℓ is defined to balance the sum of the transverse momentum of all of the other physics objects in an event, while the longitudinal component is allowed to take on any value (see appendix B). In practice, because of $\chi^2(W_\ell)$, $P_z(\nu)$ will be chosen by the $t\bar{t}$ reconstruction algorithm to be such as to make $m_{\ell\nu} \approx M_W$. In some cases, however, this is not possible because the values of $P_z(\nu)$ that satisfy the equation $m_{\ell\nu} = M_W$ are complex. In these situations, the algorithm stretches the physics object energies in a way such that $P_z(\nu)$ has real solutions. See appendix B for further details.

- The choice of the denominators in the mass terms — *i.e.* Γ_{top} and Γ_W — is somewhat arbitrary. In particular, there is no special significance in the fact that Γ_{top} is chosen as a value comparable to the physical top width or that Γ_W is chosen as the world-average W width. In order to see what issues are involved in choosing specific values for Γ_{top} and Γ_W, it is useful to consider what happens in extreme cases. For instance, if the Γ's are made very large, then χ^2_{mass} is very small before the χ^2 minimization process starts. Since χ^2_{mass} is small, $\chi^2_{kinematics}$ becomes comparable to χ^2_{mass} before the physics object energies are stretched an appreciable amount. As a consequence, the total χ^2 would reach a minimum before the physics object momenta are changed appreciably. This situation, therefore, is almost equivalent to not performing the χ^2 minimization procedure at all. This, clearly, is not desirable, so the Γ's must be made sufficiently small. Making the Γ's very small, however, poses no special problems since minor adjustments in the physics object energies can almost always make χ^2_{mass} to be arbitrarily close to zero. Finally, one has to consider the relative size of Γ_W and Γ_{top}. This choice determines the relative importance of the top and W mass conditions. It is claimed in [44] that the exact values chosen for the Γ's do

not significantly change the performance of the $t\bar{t}$ reconstruction algorithm. This is probably an oversimplification; the reconstruction algorithm's performance probably does improve or worsen depending on the relative size of Γ_W and Γ_{top}. This issue, however, is beyond the scope of this study; in this analysis, the choice in [44] is adopted.

The quantity $\chi^2_{kinematic}$ is composed of the following terms:

$$\begin{aligned} \chi^2_{kinematics} &= \sum_i \left(\frac{E_i - E_i^0}{\sigma_i} \right)^2 \\ &+ \left(\frac{P_x(uce) - P_x^0(uce)}{1 + \sigma(uce)} \right)^2 + \left(\frac{P_y(uce) - P_y^0(uce)}{1 + \sigma(uce)} \right)^2 \end{aligned} \tag{D.8}$$

The following are some comments on this equation:

- The index i in the first line spans all jets and the primary lepton.
- The last two terms correspond to the x and y components of the unclustered energy.
- The jet and charged lepton momenta are allowed to stretch/shrink, but they are not allowed to change direction. On the other hand, the unclustered energy momentum is allowed to change both its magnitude and direction.
- The superscript "0" indicates that a quantity's value is that of the original estimate. Variables without the superscript are the altered values.
- The error σ_i of the jets and the charged lepton are obtained from monte carlo studies that compare the parton-level energies with those reconstructed from the detector. The size of the error for a given object (electron, muon, generic jets, SVX b-tagged jets, SLT b-tagged jets) depends, in general, on the transverse energy and the detector position.
- The "error" of the unclustered energy, $\sigma(uce)$, is not an error in the strict, statistical sense of the word. It is a quantity that is chosen by tuning the $t\bar{t}$ reconstruction algorithm. The CDF lepton + jets top mass measurement in [44] chose $\sigma(uce)$ to be

100% of $\left|\vec{P}_T(uce)\right|$ based on the fact that: (1) the unclustered energy is poorly measured; and (2) this choice gives adequate results for the performance of the algorithm. This particular choice, however, is not crucial in the performance of the algorithm — a considerable degree of arbitrariness exists. Also, the fact that 1 GeV is added to $\sigma(uce)$ is a technicality that deals with rare situations where $\sigma(uce)$ is very close to zero.

Appendix E

The Standard Model 3-body $(V-A)\times(V-A)$ Decay of the Top Quark and Other Fermions

E.1 Introduction

According to the standard model, the top quark decays via the following charged weak process: $t \to b + W$ followed by $W \to$ lepton or quark pair. The decay is illustrated schematically below:

$$t \to b + W \longrightarrow \begin{cases} \ell + \nu_\ell \\ \text{or} \\ W_d + W_u \end{cases}$$

W_d and W_u stand for the down- and up-type quark from the hadronic decay of W. The decay vertices t-W-b and W-ℓ-ν_ℓ (or W-W_d-W_u) each have $V-A$ coupling. For this reason, the top quark decay can be characterized as a 3-body $(V-A)\times(V-A)$ decay. For the experimentally measured top quark mass of $m_t \approx 175\ GeV$, the intermediate W is on mass shell because the following inequality holds: $m_t - m_b > M_W$.

The top quark is just one fermion that decays via a 3-body $(V-A)\times(V-A)$ process. Others are the leptons μ and τ and the quarks c and b.[1] The decay of these other fermions are, in many ways, similar to the decay of the top quark. In one respect, however, the top quark decay is unique: whereas the intermediate W in the decay of μ, τ, c, and b is

[1] The s- and d-quark decays are not included in this list of fermions because they are too light for their decays to be usefully described by the spectator approximation.

off mass shell, that in the decay of t is on mass shell. Because of this, the intermediate W from the decay of t has invariant mass in a narrow region centered about M_W, whereas that from the decay of the other fermions is spread out over the kinematically allowed region. One consequence of this difference is the following: for top quark decay, four parameters are necessary and sufficient to describe the decay product kinematics in the top rest frame[2]; for all other fermions, an extra parameter is necessary, this parameter being the virtual W mass.

The objective of this appendix is to compare and contrast the standard model 3-body $(V-A)\times(V-A)$ decay of the top quark on the one hand, and that of the μ and τ leptons and the c and b-quarks on the other.

E.2 General Treatment of the 3-body $(V-A)\times(V-A)$ Decay

Let A be a fermion that decays to B, $\overline{X}$, and Y via the 3-body $(V-A)\times(V-A)$ process. In this discussion, it is important to note that A is assumed to be a *particle*. Thus B and Y are also particles, while $\overline{X}$ is an anti-particle. Discussions on the charge-conjugate process will be given later on. The decay process is illustrated below:

$$A \to B + W \hookrightarrow \overline{X} + Y$$

For the sake of concreteness, the decay products of W will be taken as leptons: $\overline{X}$, $Y = \ell$, ν_ℓ. The following chart shows B, $\overline{X}$, and Y for $A = \mu$, τ, c, b, and t:

A	B	$\overline{X}$	Y
μ^-	ν_μ	$\overline{\nu}_e$	e^-
τ^-	ν_τ	$\overline{\nu}_\ell$	ℓ^-
c	s	ℓ^+	ν_ℓ
b	c	$\overline{\nu}_\ell$	ℓ^-
t	b	ℓ^+	ν_ℓ

[2]This statement is true in the limit that the top and W widths are negligible. The zero-width approximation is adequate for the purposes of this study.

Whether the lepton under $\overline{X}$ is charged or neutral depends on the weak isospin of the parent particle. If A has weak isospin $T_3 = -1/2$ (μ^-, τ^-, b), then $\overline{X}$ is a neutral lepton; if A has $T_3 = +1/2$ (c, t), then $\overline{X}$ is a charged lepton. Another way of looking at the nature of $\overline{X}$ is this: if A has $T_3 = \mp 1/2$, then $\overline{X}$ has $T_3 = \pm 1/2$.

The matrix element for the process $A \to B + W$ followed by $W \to \overline{X} + Y$ is given by the following:

$$\begin{aligned} \mathcal{M}_{SA,SB,SX,SY} &= i(\frac{-ig}{2\sqrt{2}})\overline{u}(B,s_B)\gamma^\mu(1-\gamma^5)u(A,s_A)\frac{i(g_{\mu\nu} - W_\mu W_\nu/M_W^2)}{W^2 - M_W^2 + i\Gamma_W M_W} \\ &\quad \times(\frac{-ig}{2\sqrt{2}})\overline{u}(Y,s_Y)\gamma^\nu(1-\gamma^5)v(X,s_X) \end{aligned} \tag{E.1}$$

In the formula above, A, B, X, Y, and W stand for the 4-momentum of the respective particles, and s_A, s_B, s_X, and s_Y are the spin 4-vectors for each of the fermions. The spin of W has been summed over already in the above expression. The symbol g stands for the weak coupling constant, and is given by $g^2 = 8{M_W}^2 G_F/\sqrt{2}$.

The part of $\mathcal{M}_{SA,SB,SX,SY}$ corresponding to the W propagator is $\frac{i(g_{\mu\nu} - W_\mu W_\nu/M_W^2)}{W^2 - M_W^2 + i\Gamma_W M_W}$. It consists of two terms: one proportional to $g_{\mu\nu}$, and the other proportional to $W_\mu W_\nu/M_W^2$. The second term is comparable to the first term only if the mass of the decay products is comparable to M_W. This is clearly not the case for the particles dealt with here. Therefore the matrix element can be approximated as follows:

$$\begin{aligned} \mathcal{M}_{SA,SB,SX,SY} &= \frac{g^2}{8}\left(\frac{1}{W^2 - {M_W}^2 + i\Gamma_W M_W}\right) \times \\ &\quad \overline{u}(B,s_B)\gamma^\mu(1-\gamma^5)u(A,s_A)\,\overline{u}(Y,s_Y)\gamma_\mu(1-\gamma^5)v(X,s_X) \end{aligned} \tag{E.2}$$

After summing over the decay product spins and squaring the matrix elements, the result is the following:

$$|\mathcal{M}_{SA}|^2 = 64\,G_F^2\,\frac{1}{(1 - W^2/{M_W}^2)^2 + (\Gamma_W/M_W)^2}\;[(A - m_A s_A)\cdot X]\;[B\cdot Y] \tag{E.3}$$

The matrix element was not summed over the parent particle spin in order to allow for the possibility of spin polarization. This equation can also be written as follows:

$$|\mathcal{M}_{SA}|^2 = 64\, G_F^2\, \frac{1}{(1-\lambda^2)^2 + \varepsilon_W^2}\, [(A - m_A s_A)\cdot X]\, [B\cdot Y] \tag{E.4}$$

$$\lambda = \frac{\sqrt{W^2}}{M_W} \tag{E.5}$$

$$\varepsilon_W = \frac{\Gamma_W}{M_W} \tag{E.6}$$

The factor $((1-\lambda^2)^2 + \varepsilon_W^2)^{-1}$ in the above equation will be referred to as the W *resonance term.* The behavior of this term is crucial in determining the nature of the decay product kinematics. The factor that determines this behavior is the range of the W boson invariant mass $\sqrt{W^2}$:

$$m_X + m_Y \leq \sqrt{W^2} \leq m_A - m_B \tag{E.7}$$

If $m_A - m_B \ll M_W$, then the W propagator term is, to a good approximation, constant. If, on the other hand, $m_A - m_B > M_W$, then the W propagator term can be approximated as a delta function centered at $\sqrt{W^2} = M_W$. For $A = \mu$, τ, c, and b, the W resonance term can be taken as constant; for $A = t$, it can be taken as a delta function. With these approximations for the W resonance terms, the matrix element squared for both cases is given as follows:

Case 1: $m_A - m_B \ll M_W$

$$|\mathcal{M}_{SA}|^2 = 64\, G_F^2\, [(A - m_A s_A)\cdot X]\, [B\cdot Y] \tag{E.8}$$

Case 2: $m_A - m_B > M_W$

$$|\mathcal{M}_{SA}|^2 = 64\, G_F^2\, \frac{\pi}{2\varepsilon_W}\, \delta(1-\lambda)\, [(A - m_A s_A)\cdot X]\, [B\cdot Y] \tag{E.9}$$

E.3 Decay Product Kinematics for Case 2

The decay product kinematics for case 2 will be described first. An example of case 2 is A = top quark. The decay product kinematics for this situation was discussed in detail in chapter 7. Here the results will be summarized:

- Four parameters are necessary and sufficient to describe the decay product kinematics in the parent particle rest frame.
- Three are needed to orient the plane containing the decay product momenta $\vec{p}_B$, $\vec{p}_X$, and $\vec{p}_Y$. These three parameters can be chosen as θ_X, ϕ_X, and $\phi_{B\text{-}Y}$. See figure 7.1, taking $B = b$, $X = \ell$, and $Y = \nu_\ell$.
- One is needed to fix the direction and magnitude of $\vec{p}_B$, $\vec{p}_X$, and $\vec{p}_Y$ in the decay plane. This parameter can be chosen as E_X, the energy of X in the rest frame of A.

The parameters $(\theta_X, \phi_X, \phi_{B\text{-}Y}, E_X)$ are distributed according to the following:

$$F(\cos\theta_X, E_X) = f(\cos\theta_X)\, g(E_X) \tag{E.10}$$

The function f probes possible spin polarization of the parent A, while the function g probes the nature of the A-B-W and W-X-Y vertices. The fact that the function F depends on only two parameters $\cos\theta_X$ and E_X implies that the other two parameters, ϕ_X and $\phi_{B\text{-}Y}$, are randomly distributed.

The function f has the following form:

$$f(\cos\theta_X) = \frac{1}{2}\left(1 + K_A\, \alpha \cdot \cos\theta_X\right) \tag{E.11}$$

The quantity K_A is $+1$ if A is a particle and -1 if it is an anti-particle, while α is the degree of polarization.

The function g has the following form:

$$\begin{aligned} g(E_X) &= (h_-,\, h_0\, h_+) \cdot (g_s,\, g_m,\, g_h) \\ &= h_-(\xi_0) \cdot g_s(E_X) + h_0(\xi_0) \cdot g_m(E_X) + h_+(\xi_0) \cdot g_h(E_X) \end{aligned} \tag{E.12}$$

This equation is a generalization of those given in section 7.3.2 for $g(\cos\psi_\ell^*)$ and $g(E_\ell)$. Most of what was stated there concerning $g(\cos\psi_\ell^*)$ and $g(E_\ell)$ is applicable to $g(E_X)$, so that information will not be repeated here. Instead, aspects of equation E.12 that generalize those in section 7.3.2 will be commented on. The quantities h_-, h_0, and h_+ are the helicity fractions. Assuming $m_B \ll m_A$ (which is applicable to the only case of interest, $A = t$), these quantities can be described as a function of the *scaled, on-shell W mass* $\xi_0 = M_W/m_A$:

$$h_-(\xi_0) = \frac{2\xi_0^2}{1+2\xi_0^2} \tag{E.13}$$

$$h_0(\xi_0) = \frac{1}{1+2\xi_0^2} \tag{E.14}$$

$$h_+(\xi_0) = 0 \tag{E.15}$$

The functions g_s, g_m, and g_h are the *soft*, *medium*, and *hard* E_X distributions:

$$g_s(E_X) = \frac{3}{P_W}\left(\frac{E_X^{max}-E_X}{E_X^{max}-E_X^{min}}\right)^2 \tag{E.16}$$

$$g_m(E_X) = \frac{6}{P_W} \times \frac{(E_X^{max}-E_X)(E_X-E_X^{min})}{(E_X^{max}-E_X^{min})^2} \tag{E.17}$$

$$g_h(E_X) = \frac{3}{P_W}\left(\frac{E_X-E_X^{min}}{E_X^{max}-E_X^{min}}\right)^2 \tag{E.18}$$

$$E_X^{max} = \frac{1}{2}m_A \tag{E.19}$$

$$E_X^{min} = \frac{1}{2}\left(\frac{{M_W}^2}{m_A}\right) \tag{E.20}$$

$$P_W = E_X^{max} - E_X^{min} \tag{E.21}$$

The functions g_s, g_m, and g_h are assumed to be zero outside of the range (E_X^{min}, E_X^{max}). See figure E.1.

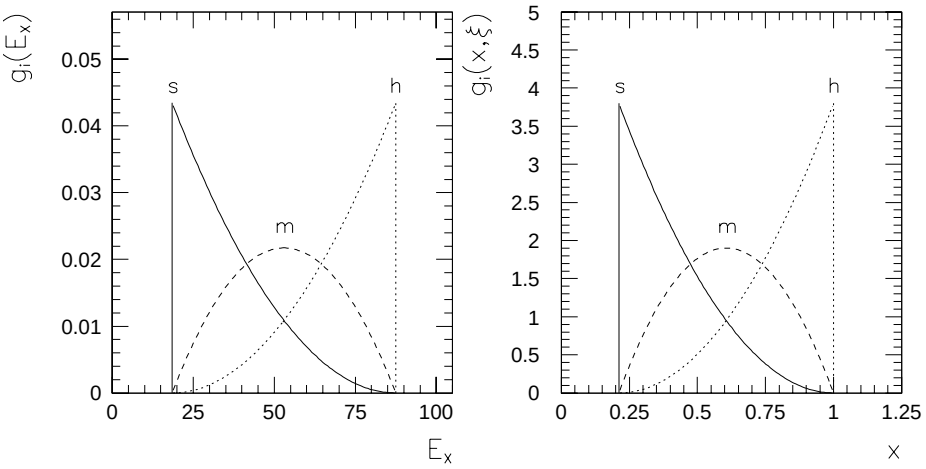


Figure E.1: Left: The energy distribution of X from the decay of the intermediate W in different helicity states. The labels 's', 'm', and 'h' stand for *soft*, *medium*, and *hard*. Right: The scaled energy distributions, with $x \equiv E_X/E_X^{max}$ and $\xi_0 \equiv M_W/m_A$. These plots use parameter values for the case A = top quark with $m_t = 175\ GeV$.

The function g can be described more usefully using the scaled energy x, defined as $x = E_X/E_X^{max}$. Substituting x for E_X, one obtains the following:

$$g(x,\xi_0) = h_-(\xi_0)\cdot g_s(x,\xi_0) + h_0(\xi_0)\cdot g_m(x,\xi_0) + h_+(\xi_0)\cdot g_h(x,\xi_0) \tag{E.22}$$

$$g_s(x,\xi_0) = \frac{3}{(1-\xi_0^2)^3}(1-x)^2 \tag{E.23}$$

$$g_m(x,\xi_0) = \frac{6}{(1-\xi_0^2)^3}(1-x)(x-\xi_0^2) \tag{E.24}$$

$$g_h(x,\xi_0) = \frac{3}{(1-\xi_0^2)^3}(x-\xi_0^2)^2 \tag{E.25}$$

The functions g_s, g_m, and $g_h$vanish outside the range $(\xi_0^2, 1)$. See figure E.1.

Using this new parameterization, the function F describing the decay kinematics of A in the A rest frame can be expressed as follows:

$$F(\cos\theta_X, x, \xi) = f(\cos\theta_X)\, g(x,\xi)\, \delta(\xi-\xi_0) \tag{E.26}$$

Although F is described as being a function of three parameters, the dependence on ξ drops out because the delta function $\delta(\xi - \xi_0)$ fixes ξ to the constant value ξ_0.

E.4 Decay Product Kinematics for Case 1

It was stated earlier that the difference between cases 1 and 2 is that, whereas in case 2 the invariant mass of the intermediate W is approximately constant, in case 1, it is spread out over the kinematically allowed region $m_X + m_Y \leq \sqrt{W^2} \leq m_A - m_B$. In other words, in case 2, the parameter ξ is constant, whereas in case 1, it is distributed according to some function, which will be derived presently.

Although the distribution of ξ differ between the two cases, the distribution of $\cos\theta_X$ and x *for a fixed value of* ξ is exactly the same between cases 1 and 2.[3] In other words, $F(\cos\theta_X, x, \xi)$ can be expressed as follows:

$$F(\cos\theta_X, x, \xi) = f(\cos\theta_X)\, g(x, \xi)\, Q(\xi) \tag{E.27}$$

The functional form of f and g is exactly the same as it is in case 2. The only difference is the distribution of ξ — the *scaled virtual W mass* — given by $Q(\xi)$.

The virtual W mass distribution $Q(\xi)$ can be obtained in the following manner. Let us assume that a particle A decays to particles 1, 2, and 3, and that case 1 is applicable (*i.e.* that the intermediate W is far below the mass shell). Let us define the *invariant distribution* of decay product i to be $E_i \cdot d\Gamma/d^3p_i$. Then, the following is true:

> The invariant distribution of decay product i is a function of the invariant mass of the other two decay products.

See chapter 3 of reference [1] for a proof. Thus, by evaluating the invariant distribution of the decay product B, one can obtain a function of the invariant mass distribution of the $\overline{X}$-Y system — *i.e.* the off-shell W:

$$E_B \frac{d\Gamma}{d^3p_B} = \frac{G_F^2}{2m_A\pi^5} \int (A \cdot X)(Y \cdot B)\, d_2(PS) \equiv G(\xi) \tag{E.28}$$

[3] This statement needs some qualification for $A = c$ and b. The functional form for $h_i(\xi)$, $i = -, 0, +$ given in the text assumes $m_B \ll m_A$. For the c- and b-quarks, m_B/m_A is about 0.2 to 0.3. Thus $h_i(\xi)$ must be modified significantly. Besides this, however, everything else in this discussion remains valid.

The symbol $d_2(PS)$ indicates integration over the 2-body phase space of the $\overline{X}$-Y system. The result of this integral is the following: (it will be assumed that $m_B \ll m_A$):

$$G(\xi) = \frac{G_F^2 m_A^3}{96\pi^4}\left[1 + \xi^2 - 2\xi^4\right] \tag{E.29}$$

The function $G(\xi)$ does not directly give the virtual W mass distribution $Q(\xi)$. The two functions, however, are related. Rearrangement of equation E.28 gives the following (this equation assumes $E_B \approx p_B$, which is valid for $m_B \ll m_A$):

$$\frac{d\Gamma}{dE_B} = 4\pi E_B \cdot G(\xi) \tag{E.30}$$

On the other hand, $Q(\xi)$, by definition, is the following:

$$Q(\xi) \equiv \frac{1}{\Gamma}\frac{d\Gamma}{d\xi} \tag{E.31}$$

The derivatives $d\Gamma/dE_b$ and $d\Gamma/d\xi$ are related by a Jacobian transformation:

$$\frac{d\Gamma}{d\xi} = \left|\frac{dE_B}{d\xi}\right| \frac{d\Gamma}{dE_B} \tag{E.32}$$

Using this relation, one obtains the following:

$$\begin{aligned} Q(\xi) &\equiv \frac{1}{\Gamma}\frac{d\Gamma}{d\xi} \\ &= \frac{4\pi}{\Gamma} E_B \left|\frac{dE_B}{d\xi}\right| G(\xi) \end{aligned} \tag{E.33}$$

The Jacobian can be determined using the formula for the energy of B in the rest frame of A:

$$\begin{aligned} E_B &= \frac{m_A^2 - W^2}{2m_A} \\ &= \frac{m_A}{2}(1 - \xi^2) \end{aligned} \tag{E.34}$$

Substituting this into the expression for $Q(\xi)$, one obtains the following:

$$Q(\xi) = 4\xi(1-\xi)^2(1+\xi)^2(1+2\xi^2) \tag{E.35}$$

The distribution of ξ is shown in figure E.2.

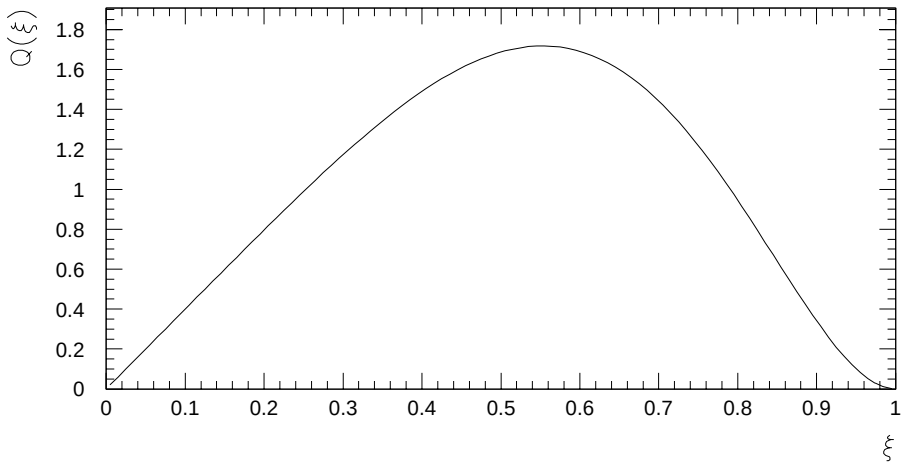


Figure E.2: The scaled virtual W invariant mass distribution $Q(\xi)$. This distribution assumes that m_A is much larger than m_B, m_X, and m_Y.

E.5 Verifying the Result of Case 1

The main result of the last section was the derivation of the virtual W invariant mass distribution, $Q(\xi)$. The derivation was, however, quite lengthy, and it involved rather abstract ideas such as Jacobian transformations and phase-space integrals. In the end of all of the steps leading to $Q(\xi)$ is a peculiar-looking 7^{th} degree polynomial. Does this really describe the virtual W mass distribution? Is there an intuitive way to appreciate this fact?

One way to do this may be to derive, using $Q(\xi)$, the energy distribution of e^- in the rest frame of μ^- in the decay $\mu^- \rightarrow \nu_\mu + \overline{\nu}_e + e^-$. This process is a famous one, and is discussed in detail in most introductory textbooks on high energy physics. The scaled energy distribution for e^- is derived from a direct matrix element calculation in chapter 3 of [1]:

$$G_e(x) = 2x^2(3 - 2x) \tag{E.36}$$

The quantity x is defined as $x = 2E/m_\mu$. The range of x is $0 \leq x \leq 1$ (ignoring the electron mass in comparison to m_μ). Let us see if this equation can be reproduced using the following facts about case 1 obtained in the last section:

1. The virtual W invariant mass distribution $Q(\xi)$ (equation E.35).

2. The W helicity-dependent scaled energy distributions $g_i(x,\xi)$, $i = s$, m, and h (equations E.16 to E.18).

3. The helicity fractions $h_j(\xi)$, $j = -, 0, +$ (equations E.13 to E.15).

The energy distribution of e^- *for a fixed value of* ξ is given by the following formula:

$$\begin{aligned} g(x,\xi) &= (h_-,\, h_0\, h_+) \cdot (g_h,\, g_m,\, g_s) \\ &= h_-(\xi) \cdot g_h(x,\xi) + h_0(\xi) \cdot g_m(x,\xi) + h_+(\xi) \cdot g_s(x,\xi) \end{aligned} \tag{E.37}$$

This is the same as equation E.22 for $\overline{X}$ in case 2 except for the fact that: (1) in general, $\xi \neq \xi_0$; and (2) g_s and g_h are swapped.[4] The distribution $G_e(x)$ is obtained by convoluting $g(x,\xi)$ with $Q(\xi)$:

$$\begin{aligned} G_e(x) &= \int_0^{\sqrt{x}} d\xi Q(\xi) g(x,\xi) \\ &= \int_0^{\sqrt{x}} d\xi Q(\xi) \left[h_-(\xi) \cdot g_h(x,\xi) + h_0(\xi) \cdot g_m(x,\xi) + h_+(\xi) \cdot g_s(x,\xi)\right] \end{aligned} \tag{E.38}$$

The range of integration is taken as $0 < \xi < \sqrt{x}$ because $g(x,\xi) \equiv 0$ for $\sqrt{x} < \xi < 1$ (see equations E.23 to E.25). Substituting equations E.13 to E.15 and E.23 to E.25 for the terms in square brackets, one obtains the following:

$$\begin{aligned} G_e(x) = \int_0^{\sqrt{x}} d\xi Q(\xi) \quad [\quad & \left(\frac{2\xi^2}{1+2\xi^2}\right)\left(\frac{3}{(1-\xi^2)^3}\right)(x-\xi^2)^2 \\ & + \left(\frac{1}{1+2\xi^2}\right)\left(\frac{6}{(1-\xi^2)^3}\right)(1-x)(x-\xi^2) \\ ] & \end{aligned} \tag{E.39}$$

An important point to note about this equation is this: the denominator of the integrand is an 8^{th} degree polynomial in the variable ξ. This implies that unless the numerator

[4] Item (2) is due to the fact that e^- is a Y-type particle in the decay $A \to B + \overline{X} + Y$. See section E.7 for a discussion of this issue.

cancels out the ξ-dependence in the denominator, the result of the integral cannot be a polynomial in x, which contradicts equation E.36. The above equation can be rearranged as follows:

$$G_e(x) = \int_0^{\sqrt{x}} d\xi \frac{6Q(\xi)}{(1+2\xi^2)(1-\xi^2)^3}(1-\xi^2)(x-\xi^2)\left[(1+\xi^2)-x\right] \qquad \text{(E.40)}$$

From equation E.35, it is seen that $Q(\xi)$ cancels out all but a factor of $1-\xi^2$ from the denominator. There is, however, an extra factor of $1-\xi^2$ coming from the other terms. Thus all the ξ-dependence in the denominator is, indeed, canceled out by the numerator. The resulting expression for $G_e(x)$ is:

$$G_e(x) = \int_0^{\sqrt{x}} d\xi\, 24\xi(x-\xi^2)\left[(1+\xi^2)-x\right] \qquad \text{(E.41)}$$

$$= 2x^2(3-2x) \qquad \text{(E.42)}$$

This agrees with equation E.36.

This result shows that the peculiar form of the virtual W mass distribution $Q(\xi)$ is a reflection of the $(V-A)\times(V-A)$ nature of the process $A \to B + \overline{X} + Y$. In other words, $G_e(x)$ is, by definition, a convolution of $Q(\xi)$ and $g(x,\xi)$. The function $g(x,\xi)$ has ξ-dependence in the denominator because of the ξ-dependence in the helicity fractions $h_i(\xi)$ (which characterizes the A-B-W vertex) and the normalization of the helicity-dependent energy distributions $g_i(x,\xi)$ (which characterizes the W-$\overline{X}$-Y vertex). The fact that $Q(\xi)$ cancels out most of the ξ-dependence in the denominator indicates that the form of $Q(\xi)$ is fixed by the nature of these two vertices.

E.6 Average W Helicity Fraction for Case 1

In case 2, where the W invariant mass is basically fixed at $\sqrt{W^2} = M_W$, the parent particle A can be thought of as decaying to the left-handed, longitudinal, and right-handed helicity state of W with probability $h_-(\xi_0)$, $h_0(\xi_0)$, and $h_+(\xi_0)$, respectively, where $h_i(\xi_0)$

is defined in equations E.13 to E.15, and $\xi_0 = M_W/m_A$.[5] For example, if A = top quark with $m_t = 175\ GeV$, then $\xi_0 = 0.459$, so that $h_- = 0.30$, $h_0 = 0.70$, and $h_+ = 0.00$. Thus, in some sense, the top quark can be thought of as decaying to a left-handed W 30% of the time, to a longitudinal W 70% of the time, and almost never to a right-handed W. This decoupling of the top quark from the right-handed W is one of the chief features of the $V-A$ coupling in the t-W-b vertex.

In case 1, the situation is not so simple because the virtual W invariant mass is spread out over a range of values. However, for a fixed value of ξ, the parent particle A can be thought of as decaying to the left-handed, longitudinal, and right-handed W with probability $h_-(\xi)$, $h_0(\xi)$, and $h_+(\xi)$, respectively. This implies that, even in case 1, the parent particle A decouples from the right-handed W. The average helicity fraction $< h_i >$ is defined as the convolution of $h_i(\xi)$ with $Q(\xi)$:

$$< h_i > = \int_0^1 d\xi Q(\xi) h_i(\xi) \tag{E.43}$$

For the case $m_A \gg m_B, m_X, m_Y$, the integral evaluates to the following: $< h_- > = 1/3$, $< h_0 > = 2/3$, and $< h_+ > = 0$.

E.7 The Helicity Structure of the Decay Vertices and the W Decay Product Kinematics

In all of the discussions on $(V-A)\times(V-A)$ decays so far, the focus has been on one of the two decay products of W. For instance, in sections E.2 to E.4, the kinematics of $\overline{X}$ in $W \to \overline{X} + Y$ was examined, while in section E.5, that of e^- (a Y-type particle) in $\mu^- \to \nu_\mu + \overline{\nu}_e + e^-$ was derived. In this section, the kinematics of both $\overline{X}$ and Y in $A \to B + W$ followed by $W \to \overline{X} + Y$ (and the charge-conjugate process) are compared. In particular, the following points are discussed:

- Determine which of the two W decay products has the *softer* energy distribution.

[5] The view that A decays to W of a definite helicity state on an event-by-event basis is, strictly speaking, incorrect because the intermediate W is, in fact, a superposition of helicity states. Yet, in an effective sense, A can be thought of as decaying to W of a definite helicity state. See appendix J for more details.

- Show that the W decay product energy distributions in $A \to B + \overline{X} + Y$ are exactly the same as those in the charge-conjugate process $\overline{A} \to \overline{B} + X + \overline{Y}$.

- Use the nature of the A-B-W and W-X-Y decay vertices to explain the relative hardness of the W decay product energy distributions, and to explain why the distributions are exactly the same in the charge-conjugate process.

- Explore what happens when one or both of the decay vertices is changed from $(V-A)$ to $(V+A)$.

E.7.1 Determining the Relative Hardness of the W Decay Product Energy Distributions

In equation E.4, $|\mathcal{M}|^2$ for the process $A \to B + W$ followed by $W \to \overline{X} + Y$ is given by the following (after summing over the parent particle spin):

$$|\mathcal{M}|^2 \sim (A \cdot X)(B \cdot Y) \tag{E.44}$$

This relation holds for both case 1 and case 2. In case 1, there is a constant factor multiplying this expression, while in case 2, there is a constant times a delta-function peaking at M_W. A notable point about this expression is that the 4-vectors of the parent and decay products are paired by dot product. The parent A is paired with $\overline{X}$, and the decay products B and Y are paired. This pairing is due to the $(V-A) \times (V-A)$ nature of the decay $A \to B + \overline{X} + Y$. This pairing can be used to determine which of the W decay products has the *softer* energy distribution in the A rest frame:

> The W decay product whose 4-vector is dotted with the parent 4-vector has the softer energy distribution.

Why this is so will be discussed in section E.7.3. For the time being, let us take this statement as a given and explore its implications.

As an illustration, let us first see what this statement implies about two specific processes: (a) $t \to b + \ell^+ + \nu_\ell$ ($\ell = e$, μ or τ); and (b) $\mu^- \to \nu_\mu + e^- + \overline{\nu}_e$. In (a), $\overline{X} = \ell^+$ and $Y = \nu_\ell$; in (b), $\overline{X} = \overline{\nu}_e$ and $Y = e^-$. This implies that in (a), ℓ^+ has the softer

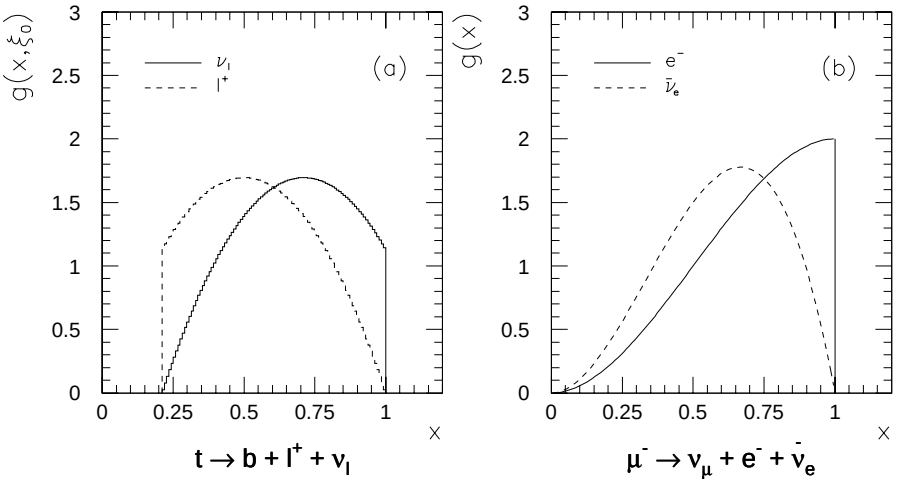


Figure E.3: The scaled energy distribution of the W decay products in: (a) $t \to b + \ell^+ + \nu_\ell$; and (b) $\mu^- \to \nu_\mu + \overline{\nu}_e + e^-$.

energy distribution, while in (b), $\overline{\nu}_e$ has the softer distribution. Figure E.3 demonstrates this point. This illustration demonstrates the fact that the lepton charge cannot be used to predict which W decay product has the softer energy distribution. In case (a), the charged lepton distribution is softer; in (b), it is the neutrino. The relative hardness of the energy distributions can be predicted using a number of different methods. One was given at the beginning of this section: the W decay product whose 4-momentum is dotted with the parent 4-momentum has the softer distribution. The following are other methods:

Weak Isospin

If the parent has weak isospin $T_3 = \pm 1/2$, then the W decay product with opposite isospin $(\mp 1/2)$ has the softer energy distribution.

Particle/Anti-particle

If the parent is a particle (anti-particle), then the anti-particle (particle) decay product of W has the softer energy distribution.

Feynman Diagram

In this method, one has to keep track of the momentum flow in Feynman diagrams. In particular, the momentum flow of anti-particles are time-reversed compared to that

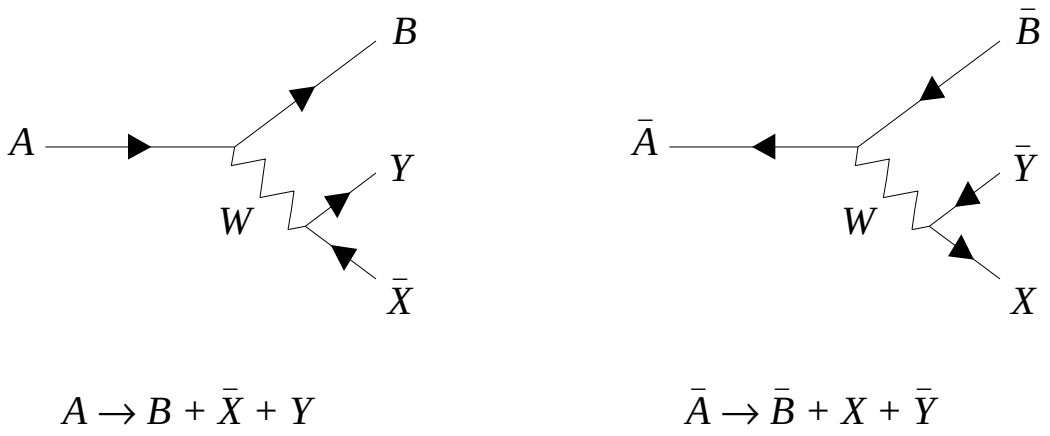


Figure E.4: Feynman diagram for $A \rightarrow B + \overline{X} + Y$, and the charge-conjugate process.

of a particle. With this in mind, if the parent particle momentum is flowing into (out of) the diagram, then the W decay product whose momentum flows into (out of) the diagram has the softer energy distribution (see figure E.4).

E.7.2 The Charge-conjugate Process $\overline{A} \rightarrow \overline{B} + X + \overline{Y}$

The charge-conjugate process $\overline{A} \rightarrow \overline{B} + X + \overline{Y}$ has exactly the same $|\mathcal{M}|^2$ as for $A \rightarrow B + \overline{X} + Y$ (assuming all spins are summed over). As a consequence, the energy distribution of X is exactly the same as for $\overline{X}$, and the same for $\overline{Y}$ and Y. There are several ways to demonstrate why this should be so. The clearest way is to examine the nature of the A-B-W and W-X-Y decay vertices. This will be done in the next section. Here, an alternative method will be introduced. This method involves the examination of the complex matrix element.

Equation E.1 shows the complex matrix element for the process $A \rightarrow B + \overline{X} + Y$. In simplified terms, it can be written as follows:

$$\mathcal{M} \sim \left[\overline{u}(B)\gamma^{\mu}(1-\gamma^5)u(A)\right] D_{\mu\nu}(W) \left[\overline{u}(Y)\gamma^{\nu}(1-\gamma^5)v(X)\right] \qquad \text{(E.45)}$$

The quantity $D_{\mu\nu}(W)$ is the W propagator. Information on constant factors and the spin of all the particles has been suppressed because it is irrelevant in this context. The Feynman rule for constructing this matrix element from figure E.4 can be found in introductory particle physics text books [1], [2], [3]. The important point to note about $\mathcal{M}$ is the order in which the 4-vectors appear, from left to right: (B, A, Y, X). When $\mathcal{M}$ is squared, the 1^{st} and 3^{rd} 4-vectors are dotted together, and the 2^{nd} and 4^{th} ones are dotted.

Using Feynman rules to obtain $\mathcal{M}$ for the charge conjugate process $\overline{A} \rightarrow \overline{B} + X + \overline{Y}$, one obtains the following:

$$\mathcal{M} \sim \left[\overline{v}(A)\gamma^{\mu}(1-\gamma^5)v(B)\right] D_{\mu\nu}(W) \left[\overline{u}(X)\gamma^{\nu}(1-\gamma^5)v(Y)\right] \tag{E.46}$$

This has the same form as equation E.45 except for the fact that: (1) the spinors u and v are swapped; and (2) the 4-vectors A and B are swapped, and so are X and Y. For the moment, it will simply be stated that the swap $u \leftrightarrow v$ is unimportant in determining the form of $|\mathcal{M}|^2$. Thus the only essential difference between the two equations is the ordering of the 4-vectors. In the charge-conjugate case, this is: (A, B, X, Y). Because $\mathcal{M}$ in the two cases have the same form, the rule for pairing the 4-vectors in $|\mathcal{M}|^2$ must be exactly the same: match the 1^{st} with the 3^{rd}, and the 2^{nd} with the 4^{th}. In other words, $(A \cdot X)(B \cdot Y)$. This is exactly the same as before. This equality is due to the fact that the swapping takes place at both the A-B-W and the W-X-Y vertices. This double-swapping behavior is something that will also be seen in the next section, where the equivalence of the charge-conjugate processes will be demonstrated using an alternative technique.

Before leaving this section, let us see what differences exist between the charge conjugate processes. It has already been stated that charge conjugation results in the swaps: (1) $u \leftrightarrow v$; (2) $A \leftrightarrow B$; and (3) $X \leftrightarrow Y$. The effects of (2) and (3) are obvious. The first swap affects the spin configuration of a particle. For example, suppose particle A were produced with 100% spin polarization. Then one must take account its spin in order to obtain the correct kinematics. The spin 4-vector of A, s_A, shows up in $\mathcal{M}$ inside A's spinor: $u(A, s_A)$. Earlier, when the spin information was ignored, the squaring of $\mathcal{M}$ resulted in the dotting of A with X. When the spin of A is taken account of, A is replaced with $A - m_A s_A$, where m_A is A's mass. This is summarized below:

	$\lvert\mathcal{M}\rvert^2$
UNPOLARIZED	$[A \cdot X][B \cdot Y]$
100% POLARIZED	$[(A - m_A s_A) \cdot X][B \cdot Y]$

The effect of swapping u with v is to change the sign of the spin term — *i.e.* $A - m_A s_A \rightarrow A + m_A s_A$. This change in sign has no effect on the decay product energy distributions. It does, however, greatly affect the angular distributions. In particular, whereas $\overline{X}$ has a strong tendency to travel *along* A's spin, X has a strong tendency to travel *away* from $\overline{A}$'s spin.[6]

E.7.3 Examining the W Decay Product Kinematics in Light of the Nature of the Decay Vertices

In the last two sections, the following observations have been made concerning the W decay product energy distributions:

- In $A \rightarrow B + \overline{X} + Y$, the energy distribution of $\overline{X}$ is softer than that of Y.
- The energy distribution of X $(\overline{Y})$ in the charge conjugate process $\overline{A} \rightarrow \overline{B} + X + \overline{Y}$ is exactly the same as that of $\overline{X}$ (Y) in $A \rightarrow B + \overline{X} + Y$.

These observations will be confirmed in this section by examining the nature of the A-B-W and the W-X-Y decay vertices. In the discussion that follows, $m_A \gg m_B$, m_X, and m_Y is assumed.

In both cases 1 and 2, the scaled energy distribution of $\overline{X}$ in $A \rightarrow B + \overline{X} + Y$ is given by the following equation:

$$\begin{aligned} g(x,\xi) &= (h_-,\, h_0,\, h_+) \cdot (g_s,\, g_m,\, g_h) \\ &= h_-(\xi) \cdot g_s(x,\xi) + h_0(\xi) \cdot g_m(x,\xi) + h_+(\xi) \cdot g_h(x,\xi) \end{aligned} \tag{E.47}$$

[6] In this discussion, no mention was made of the spin-dependence of the decay products. This is because the behavior of the decay product spin is fixed by the decay vertices A-B-W and W-X-Y. The behavior of the spin of A, however, is determined by the physics of the process that created A. Since this is not specified, one is free to choose A's spin behavior.

In case 1, the scaled W mass ξ is spread out between 0 and 1 (equation E.35); in case 2, it is fixed at $\xi_0 = M_W/M_A$. In the first line of the above equation, $g(x,\xi)$ is described in suggestive notation as a dot product of a triplet of scalars (h_-, h_0, h_+) with a triplet of functions (g_s, g_m, g_h). The three components of both triplets correspond to the three helicity states of the intermediate W boson. The correspondence between the components and the helicity states depends on the charge state of the parent. For $A \to B + \overline{X} + Y$, the 1^{st}, 2^{nd}, and 3^{rd} components correspond to the left-handed, longitudinal, and right-handed states of W. In the charge conjugate process $\overline{A} \to \overline{B} + X + \overline{Y}$, the 1^{st}, 2^{nd}, and 3^{rd} components correspond to the right-handed, longitudinal, and left-handed states. This assignment is summarized below:

$A \to B + \overline{X} + Y$	(left, long, right)
$\overline{A} \to \overline{B} + X + \overline{Y}$	(right, long, left)

The triplet of scalars (h_-, h_0, h_+) is determined by the nature of the A-B-W decay vertex. Because this vertex has $V-A$ coupling, the parent A decouples from the right-handed helicity state of W — *i.e.* $h_+ = 0$. The triplet, therefore, can be written as follows:

$$(h_-,\, h_0,\, h_+) = (r_T,\, 1 - r_T,\, 0) \tag{E.48}$$

$$r_T = \frac{2\xi^2}{1 + 2\xi^2} \tag{E.49}$$

For $\overline{A}$, $V-A$ coupling causes it to decouple from the *left-handed* helicity state. By definition, h_+ is the coupling strength of $\overline{A}$ to the left-handed state, so $h_+ = 0$. It can also be shown that $\overline{A}$'s coupling to the right-handed state is given by $h_- = r_T$. Thus the triplet (h_-, h_0, h_+) for $\overline{A}$ is the same as that of A.

The triplet (g_s, g_m, g_h) is determined by the nature of the W-X-Y vertex. This notation indicates that the energy distribution of $\overline{X}$ from the decay of a left-handed, longitudinal, and right-handed W is soft, medium, and hard (see figure E.1). The origin of this behavior is in the $V-A$ nature of the W-X-Y vertex. The following chain of logic describes how the $V-A$ nature explains this behavior:

1. In the $V - A$ decay of W, the particle (anti-particle) decay product is produced in the left-handed (right-handed) helicity state.

2. To conserve spin angular momentum, the particle (anti-particle) decay product tends to travel against (along) the W spin direction.

3. A left-handed W produced in the decay A has spin pointing against the boost direction from A's rest frame to W's rest frame. Thus, in the rest frame of a left-handed W, the particle (anti-particle) decay product tends to travel along (against) the boost direction (see figure E.5). For a right-handed W, the particle (anti-particle) tends to travel against (along) this direction. For a longitudinal W, both particle and anti-particle tend to travel perpendicular to this direction.

4. An object that tends to travel against the boost direction in the W rest frame has soft energy distribution in A's rest frame, while an object that tends to travel along this direction has a hard energy distribution. An object that tends to travel perpendicular to the boost direction has medium energy distribution. This relationship between direction of travel in W's rest frame and hardness of energy in A's rest frame is due to the nature of the Lorentz transformation (see equation 7.1 in section 7.2).

5. In the decay of a left-handed, longitudinal, and right-handed W, $\overline{X}$ tends to travel against, normal to, and along the boost direction. Thus $\overline{X}$ originating from the decay of a left-handed, longitudinal, and right-handed W has soft, medium, and hard energy distribution — *i.e.* the triplet of functions is (g_s, g_m, g_h). For Y, the trend is opposite that of $\overline{X}$, so the triplet of functions is (g_h, g_m, g_s).

The chain of reasoning used above for $A \rightarrow B + \overline{X} + Y$ is valid for the charge-conjugate process $\overline{A} \rightarrow \overline{B} + X + \overline{Y}$. According to the reasoning, the energy distribution of X from the decay of a left-handed, longitudinal, and right-handed W is hard, medium, and soft, while for $\overline{Y}$, it is soft, medium, and hard. In the charge-conjugate process, the assignment of helicity states to each component of a triplet is reversed, so the triplet of functions for X is (g_s, g_m, g_h) — this is the same as that of $\overline{X}$. Similarly, $\overline{Y}$ has triplets (g_h, g_m, g_s), which is the same as that of Y.

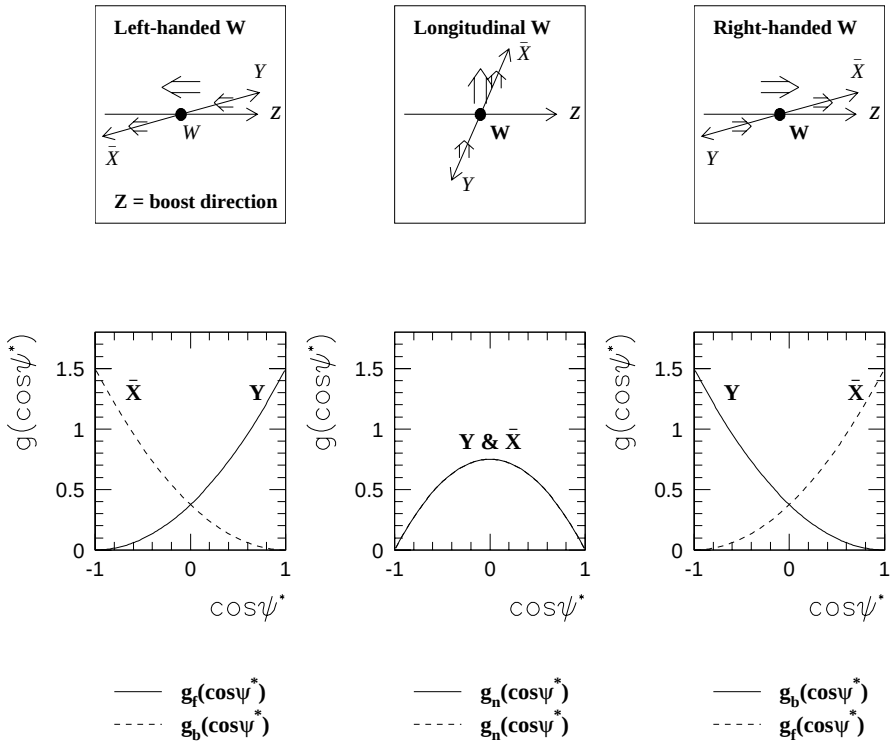


Figure E.5: Demonstrating the W rest frame angular distribution of W decay products originating from W in the three possible helicity states. The direction z is the boost direction from the parent rest frame to the W rest frame. The decay process is $W \rightarrow \overline{X} + Y$. The double-arrows represent spin, and ordinary arrows represent momentum $(\overline{X}, Y)$ or boost direction (z).

The chain of reasoning given above allows one to confirm the observations about the W decay product kinematics made in sections E.7.1 and E.7.2. First, it was stated that in $A \rightarrow B + \overline{X} + Y$, the energy distribution of $\overline{X}$ is softer than that of Y. This can be understood by examining the triplets for both objects. For both $\overline{X}$ and Y, the triplet of scalars is: $(h_-, h_0, h_+) = (r_T, 1 - r_T, 0)$. Both objects have the same triplet because W originates from the same decay vertex, A-B-W. The triplet of functions for $\overline{X}$ is (g_s, g_m, g_h), while that of Y is (g_h, g_m, g_s). When the dot product is performed to obtain the energy distributions, one obtains the following:

W DECAY PRODUCT	ENERGY DISTRIBUTION
$\overline{X}$	$r_T \cdot g_s + (1 - r_T) \cdot g_m$
Y	$r_T \cdot g_h + (1 - r_T) \cdot g_m$

This shows that, of the W decay products, $\overline{X}$ has the softer energy distribution.

Another observation that was made was the fact that the energy distribution of $\overline{X}$ (Y) is exactly the same as that of the corresponding particle X $(\overline{Y})$ in the charge-conjugate process. This equality has already been established formally — in the above discussions, it has been shown that the triplets for the scalars and the functions remain unchanged for the charge-conjugate process. This formal equality, however, obscures the physical basis, so let us examine it more closely. As an example, let us take the $\overline{X}$ and X energy distributions in the processes $A \rightarrow B + \overline{X} + Y$ and $\overline{A} \rightarrow \overline{B} + X + \overline{Y}$. Table E.1 shows that, at the A-B-W decay vertex, the $V - A$ nature of the vertex causes A to decouple from the right-handed state of W, while it causes $\overline{A}$ to decouple from the left-handed state. At the W-X-Y vertex, the $V - A$ nature of the vertex causes $\overline{X}$ from a left-handed, longitudinal, and right-handed W to have soft, medium, and hard energy distributions, while for X, they are hard, medium, and soft. The effect of charge conjugation, therefore, is to swap left with right at both vertices. The swapping at the A-B-W vertex, however, is undone by the swap in the W-X-Y vertex. The result is that the energy distributions of $\overline{X}$ and X are identical.

	A-B-W			W-X-Y		
	LEFT	LONG	RIGHT	LEFT	LONG	RIGHT
$A \rightarrow B + \overline{X} + Y$	r_T	$1 - r_T$	0	g_s	g_m	g_h
$\overline{A} \rightarrow \overline{B} + X + \overline{Y}$	0	$1 - r_T$	r_T	g_h	g_m	g_s

Table E.1: The coupling of A and $\overline{A}$ to the three helicity states of W, and the $\overline{X}$ and X energy distributions from the decay of W in the three helicity states.

E.7.4 What Happens When One or Both Decay Vertices is Changed from $V - A$ to $V + A$

The 3-component notation developed for describing the W decay product energy distributions can be used to understand in a simple way what happens when one or both of the decay vertices in $A \rightarrow B + \overline{X} + Y$ is changed from $V - A$ to $V + A$. For the sake of concreteness, let us examine the $\overline{X}$ energy distribution.

The effect of the swap $V - A \rightarrow V + A$ at the A-B-W decay vertex is to change the role of left and right. In other words, with $V - A$, A decouples from the right-handed helicity state of W; with $V + A$, it decouples from the left-handed state. The triplet of helicities, therefore, changes as follows:

	$V - A$	$V + A$
(h_-, h_0, h_+)	$(r_T, 1 - r_T, 0)$	$(0, 1 - r_T, r_T)$

The effect of this swap at the W-X-Y vertex is, again, to change left with right. In other words, with $V - A$, an $\overline{X}$ originating from a left-handed, longitudinal, and right-handed W has soft, medium, and hard energy distributions; with $V + A$, the corresponding distributions are hard, medium and soft. The triplet of functions, therefore, changes as follows:

	$V - A$	$V + A$
(g_-, g_0, g_+)	(g_s, g_m, g_h)	(g_h, g_m, g_s)

The effect of changing one or both of the vertex from $V - A$ to $V + A$ is summarized in table E.2. This table demonstrates the following:

		W-X-Y	
		$(V-A)$	$(V+A)$
A-B-W	$(V-A)$	$(r_T, 1-r_T, 0)\cdot(g_s, g_m, g_h)$	$(r_T, 1-r_T, 0)\cdot(g_h, g_m, g_s)$
	$(V+A)$	$(0, 1-r_T, r_T)\cdot(g_s, g_m, g_h)$	$(0, 1-r_T, r_T)\cdot(g_h, g_m, g_s)$

Table E.2: The $\overline{X}$ energy distribution for $(V \pm A) \times (V \pm A)$ in the decay $A \to B + W$ followed by $W \to \overline{X} + Y$.

- If just one of the decay vertex is changed (A-B-W or W-X-Y), then the $\overline{X}$ energy distribution is exactly the same as that of Y in a $(V-A)\times(V-A)$ interaction.

- Changing just the A-B-W vertex results in the same $\overline{X}$ energy distribution as that obtained by changing just the W-X-Y vertex.

- Changing both vertices results in the same $\overline{X}$ energy distribution as that in $(V-A)\times(V-A)$. In other words, the W decay product energy distributions in $(V+A)\times(V+A)$ and in $(V-A)\times(V-A)$ are the same. This equality is due to the fact that the role change of right and left at the A-B-W vertex is canceled out by a similar role change at the W-X-Y vertex.

Appendix F

The Full Matrix Element Calculation of $t\bar{t}$ Production and Decay and $t\bar{t}$ Spin Correlation

F.1 Introduction

In section 7.3, it was stated that if $\tau_{top} \ll \tau_{had}$, then t and $\bar{t}$ spin-alignment and spin-coherence in $t\bar{t}$ production must be taken account of in order to correctly simulate the t and $\bar{t}$ decay kinematics. It was decided, however, that these spin-alignment and spin-coherence effects will be ignored in the t and $\bar{t}$ decay simulations in this thesis. One reason why this decision was made is because the t and $\bar{t}$ decay kinematics with the spin effects ignored is almost identical to that which incorporates it. In this appendix, the t and $\bar{t}$ decay product kinematics in which the spin effects are taken account of (*i.e.* the full matrix element calculation) is described. This is compared to that of the independent-decay procedure, and the effect of any difference between the kinematics on the observables analyzed in this thesis is examined.

F.2 The Full Matrix Element Calculation

In order to obtain the t and $\bar{t}$ decay kinematics for $p\bar{p} \rightarrow t\bar{t}$ in which the t and $\bar{t}$ spin-alignment and spin-coherence effect are taken account of, the matrix element corresponding to the diagrams shown in figure 7.3 must be obtained. Note that the diagrams are quite

complicated because it involves all of the incoming, intermediate, and outgoing particles: $q\overline{q}$ or $gg \to t\overline{t} \to \ell + \nu_\ell + b_\ell + b_h + W_d + W_u$. One way to obtain the matrix element is to use the helicity projection technique, which is given in general terms in appendix C of [1], and specifically for $p\overline{p} \to t\overline{t}$ in [29]. In this appendix, however, a quaint but intuitive method described in [32] and [33] will be used. This alternative method will be referred to as the *phase space weighting technique* because it gives the correct kinematics by weighting the angular phase space of decay products of t and $\overline{t}$ in the t and $\overline{t}$ rest frames.

The first step is to obtain, in the $t\overline{t}$ rest frame, the tree-level differential cross section for $q\overline{q}$, $gg \to t\overline{t}$ (averaged over incoming parton spin, but *not* summed over the t and $\overline{t}$ spins). They are the following:

$$\begin{aligned}
\frac{d\sigma^{q\overline{q}}}{d\Omega_t} \;=\; \frac{2}{9}\frac{\alpha_s^2\beta}{8s^3}\{ \;& \tfrac{1}{2}[s^2 + 4sm_t^2 + (lQ)^2] \\
& -\tfrac{1}{2}s_+s_-[-s^2 + 4sm_t^2 + (lQ)^2] \\
& +lQ[Ps_- \cdot ls_+ - Ps_+ \cdot ls_-] \\
& -s[Ps_+ \cdot Ps_- - ls_+ \cdot ls_-] \qquad \}
\end{aligned} \tag{F.1}$$

$$\begin{aligned}
\frac{d\sigma^{gg}}{d\Omega_t} \;=\; & \frac{\alpha_s^2\beta}{256s}\frac{9(lQ/s)^2+7}{3[(lQ/s)^2-1]^2}\{ \\
& -(lQ/s)^4 - 2(lQ/s)^2(1-\beta^2) - 2\beta^4 + 2\beta^2 + 1 \\
& +s_+s_-[(lQ/s)^4 - 2(lQ/s)^2\beta^2 + 2\beta^4 - 2\beta^2 + 1] \\
& +\tfrac{2}{s}[\beta^2 - (lQ/s)^2][(ls_+ \cdot ls_- - Ps_+ \cdot Ps_-) - (lQ/s)(Ps_+ \cdot ls_- - Ps_- \cdot ls_+)] \qquad \}
\end{aligned} \tag{F.2}$$

The quantities appearing in the above expressions are defined as follows:

$$\begin{aligned}
s &= (p_1 + p_2)^2, && (p_1,\, p_2 = \text{incoming parton 4-momentum}) \\
P &= p_1 + p_2 \\
l &= p_1 - p_2 \\
Q &= t - \overline{t}, && (t,\, \overline{t} = t \text{ and } \overline{t} \text{ 4-momenta}) \\
s_\pm &= (s_\pm^0, \hat{s}_\pm), && (t \;\&\; \overline{t} \text{ spin 4-vectors}) \\
\beta &= |\vec{\beta}| = |\vec{t}|/E_t, && (t \;\&\; \overline{t} \text{ relativistic speed})
\end{aligned}$$

These cross sections can be converted into differential cross sections that describe the t and $\overline{t}$ decay kinematics using a special technique described in [33] and [32]. For the sake

of notational convenience, let us assume that t and $\bar{t}$ decay in the *dilepton decay channel* — *i.e.* $t \to b + \ell^+ + \nu_\ell$, and $\bar{t} \to \bar{b} + \ell^- + \bar{\nu}_\ell$ (to convert to the lepton + jets decay channel, replace ℓ with W_d and ν_ℓ with W_u for whichever of t or $\bar{t}$ that decays hadronically).

The t and $\bar{t}$ spin-dependent cross section will be denoted as follows:

$$\frac{d\sigma^{q\bar{q},gg}}{d\Omega_t}(s_+, s_-) \tag{F.3}$$

The s_+ and s_- in parenthesis are a reminder that the cross section depends on the t and $\bar{t}$ spin. This can be converted to the differential cross section $d\sigma^{q\bar{q},gg}/d\Omega_x^* d\Omega_{x'}^* d\Omega_t$ that describe the angular distribution of the decay product $x = \ell^+$ and $x' = \ell^-$ as follows:

$$\frac{d\sigma^{q\bar{q},gg}}{d\Omega_x^* d\Omega_{x'}^* \Omega_t} \sim \frac{d\sigma^{q\bar{q},gg}}{d\Omega_t}(n_+, n_-) \tag{F.4}$$

The solid angles Ω_x^* and $\Omega_{x'}^*$ are in the t and $\bar{t}$ rest frame, respectively. What equation F.4 shows is this: in order to obtain $d\sigma^{q\bar{q},gg}/d\Omega_x^* d\Omega_{x'}^* d\Omega_t$, one takes equation F.3 and replaces s_+ and s_- with n_+ and n_-. The 4-vectors $n_\pm$ are defined in two steps:

Step 1

Express the t spin 4-vector s_+ in terms of the spin 3-vector $\hat{s}_+^*$ in the t rest frame by boosting $\hat{s}_+^*$ to the $t\bar{t}$ rest frame (*i.e.* by $-\vec{\beta}$). Follow the same procedure with the $\bar{t}$ spin 4-vector s_- (boost $\hat{s}_-^*$ by $+\vec{\beta}$):

$$s_+ = (s_+^0, \vec{s}_+) = (\gamma\vec{\beta}\cdot\hat{s}_+^*,\ \hat{s}_+^* + \frac{\gamma^2}{\gamma+1}(\vec{\beta}\cdot\hat{s}_+^*)\vec{\beta}) \tag{F.5}$$

$$s_- = (s_-^0, \vec{s}_-) = (-\gamma\vec{\beta}\cdot\hat{s}_-^*,\ \hat{s}_-^* + \frac{\gamma^2}{\gamma+1}(\vec{\beta}\cdot\hat{s}_-^*)\vec{\beta}). \tag{F.6}$$

Step 2

To obtain n_+, replace the t rest frame spin 3-vector $\hat{s}_+^*$ by $\alpha_t\hat{q}^*$. For n_-, replace the $\bar{t}$ rest frame spin 3-vector $\hat{s}_-^*$ by $\alpha_{\bar{t}}\hat{q'}^*$:

$$n_+ = (n_+^0, \vec{n}_+) = \alpha_t(\gamma\vec{\beta}\cdot\hat{q}^*,\ \hat{q}^* + \frac{\gamma^2}{\gamma+1}(\vec{\beta}\cdot\hat{q}^*)\vec{\beta}) \tag{F.7}$$

$$n_- = (n_-^0, \vec{n}_-) = \alpha_{\bar{t}}(-\gamma\vec{\beta}\cdot\hat{q'}^*,\ \hat{q'}^* + \frac{\gamma^2}{\gamma+1}(\vec{\beta}\cdot\hat{q'}^*)\vec{\beta}). \tag{F.8}$$

In the equations, $\hat{q}^*$ is the unit vector pointing along x's 3-momentum in the t rest frame, and $\hat{q}'^*$ is defined similarly. The parameter α_t equals $+1$, and $\alpha_{\overline{t}}$ equals -1. The fact that $\hat{s}^*_+$ is replaced with the unit vector along the 3-momentum of ℓ^+ in the t rest frame, while $\hat{s}^*_-$ is replaced with the negative of the unit vector along the 3-momentum of ℓ^- in the $\overline{t}$ rest frame, is intuitively reasonable because the charged lepton in the semileptonic decay of the top quark is known to travel preferentially along the top quark spin, while that in the semileptonic decay of anti-top is known to travel against the $\overline{t}$ spin.

When this substitution is performed, the differential cross section comes out to be as follows:

$$
\begin{aligned}
\frac{d\sigma^{q\overline{q}}}{d\Omega^*_x d\Omega^*_{x'} d\Omega_t} \quad \sim \quad & \{8\gamma^2[(\gamma^2+1)+(\gamma^2-1)\cos^2\theta_t)] \\
& -8\gamma^2(\gamma^2-1)\alpha_t\alpha_{\overline{t}}\sin^2\theta_t[\cos\theta^*_{+-}+2\gamma^2\beta^2\cos\psi_+\cos\psi_-] \\
& -16\gamma^5\beta^2\alpha_t\alpha_{\overline{t}}[\cos\theta_t(\cos\theta^*_x\cos\psi_- + \cos\theta^*_{x'}\cos\psi_+) \\
& \qquad +2(\gamma-1)\cos^2\theta_t\cos\psi_+\cos\psi_-] \\
& +16\gamma^4\alpha_t\alpha_{\overline{t}}[((\gamma^2-1)+(\gamma-1)^2\cos^2\theta_t)\cos\psi_+\cos\psi_- + \cos\theta^*_x\cos\theta^*_{x'} \\
& \qquad +(\gamma-1)\cos\theta_t(\cos\theta^*_x\cos\psi_- + \cos\theta^*_{x'}\cos\psi_+)] \\
& \}
\end{aligned}
\tag{F.9}
$$

$$\frac{d\sigma^{gg}}{d\Omega_x^* d\Omega_{x'}^* d\Omega_t} \sim \frac{9\beta^2\cos^2\theta_t+7}{3(1-\beta^2\cos^2\theta_t)^2}\{(1+2\beta^2-2\beta^4)-2\beta^2(1-\beta^2)\cos^2\theta_t-\beta^4\cos^4\theta_t$$

$$+2\alpha_t\alpha_{\bar{t}}\beta^2\sin^2\theta_t[((\gamma^2-1)+(\gamma-1)^2\cos^2\theta_t)\cos\psi_+\cos\psi_-$$
$$+\cos\theta_x^*\cos\theta_{x'}^*$$
$$-\tfrac{\gamma-1}{\gamma}\cos\theta_t(\cos\theta_x^*\cos\psi_-+\cos\theta_{x'}^*\cos\psi_+)$$
$$-2\beta^2\gamma(\gamma-1)\cos^2\theta_t\cos\psi_+\cos\psi_-]$$

$$-\alpha_t\alpha_{\bar{t}}[\cos\theta_{+-}^*+2\gamma^2\beta^2\cos\psi_+\cos\psi_-]$$
$$\times[(1-2\beta^2+2\beta^4)-2\beta^4\cos^2\theta_t+\beta^4\cos^4\theta_t]$$

$\}$ (F.10)

The symbols appearing in the above expressions are described below:

θ_t	The angle between the momentum of t and the beam in the $t\bar{t}$ rest frame.
θ_x^*, $\theta_{x'}^*$	θ_x^* is the angle between the momentum of x *in the* t *rest frame* and the beam direction *in the* $t\bar{t}$ *rest frame*. $\theta_{x'}^*$ is defined similarly. See below about the validity of angles between vectors in different reference frames.
θ_{+-}^*	The angle between the momentum of x in the t rest frame and of x' in the $\bar{t}$ rest frame. See below about the validity of angles between vectors in different reference frames.
β	The relativistic speed of t and $\bar{t}$ in the $t\bar{t}$ rest frame.
γ	The relativistic dilation factor of t and $\bar{t}$ in the $t\bar{t}$ rest frame. $\gamma = 1/\sqrt{1-\beta^2}$.

The angles θ_x^*, $\theta_{x'}^*$, and θ_{+-}^* are peculiar because they represent angles between vectors in different reference frames. In general, such angles may suffer from ambiguities because of arbitrary rotations introduced by the freedom to choose coordinate systems. However, if angles in the t and $\bar{t}$ rest frame are defined in terms of the coordinate system used to boost from the $t\bar{t}$ to the t and $\bar{t}$ rest frame (see figure 7.8), then such ambiguities can be avoided, and angles between vectors in different rest frames can be defined meaningfully. See [33] for a justification of using mixed-reference frame angles.

The following are some observations on equations F.9 and F.10:

- The angular distribution of x in the t rest frame and of x' in the $\bar{t}$ rest frame are *correlated* — *i.e.* the direction of x's momentum in the t rest frame influences that of x' in the $\bar{t}$ rest frame, and *vice versa.*

- Both equations have the form $A + \alpha_t \alpha_{\bar{t}} \cdot B$. The first term A is independent of the x and x' kinematics; it gives the t and $\bar{t}$ momentum distribution in the $t\bar{t}$ rest frame. The second term B depends on the kinematics of x and x'; it gives the correlation between the direction of x in the t rest frame and of x' in the $\bar{t}$ rest frame.

- The independent decay procedure sets α_t and $\alpha_{\bar{t}}$ to zero. It uses the remaining expression to obtain the t and $\bar{t}$ momentum distribution in the $t\bar{t}$ rest frame. The decay of x and x' in the t and $\bar{t}$ rest frame are carried out using the matrix element for the process $t \to b + \ell + \nu_\ell$ or $b + W_d + W_u$ and assuming: (1) the t and $\bar{t}$ spin are randomly oriented; and (2) the t and $\bar{t}$ decay are independent.

- If one were to focus on the angular distribution of x alone (not the orientation of x relative to x'), then it is isotropic. In other words, the full matrix element calculation does not introduce *spin polarization.* However, if the differential cross sections are weighted by acceptance effects, it is possible, in principle, that spin-polarization effects will show up. Monte carlo studies show, however, that such polarization effects are negligible.

F.3 $t\bar{t}$ Spin Correlation at Production Threshold

In the last section, the differential cross sections $d\sigma^{q\bar{q},gg}/d\Omega_x^* d\Omega_{x'}^* d\Omega_t$ were obtained in their most general form. As equations F.9 and F.10 show, the resulting expressions are very complicated, and it is rather difficult to make sense of what kind of spin-related effects give rise to the correlation between the angular distributions of x and x'. In order to gain some insight in this regard, it is useful to examine the differential cross sections at

$t\bar{t}$ production threshold — *i.e.* in situations where the incoming partons have just enough energy to create a $t\bar{t}$ pair.

At $t\bar{t}$ production threshold, the relativistic speed β of t and $\bar{t}$ in the $t\bar{t}$ rest frame is zero. In the limit as $\beta \to 0$, the differential cross sections in equations F.9 and F.10 are (after integrating out the top quark solid angle Ω_t):

$$\frac{1}{\beta} \cdot \frac{d\sigma^{q\bar{q}}}{d\Omega_x d\Omega_{x'}} \sim 1 - \cos\theta^*_x \cos\theta^*_{x'} \tag{F.11}$$

$$\frac{1}{\beta} \cdot \frac{d\sigma^{gg}}{d\Omega_x d\Omega_{x'}} \sim 1 + \cos\theta^*_{+-} \tag{F.12}$$

The differential cross section in equation F.11 can alternatively be expressed as a function of the angle $\cos\theta^*_{+-}$. When this is done, and normalizing both distributions to 1, one obtains the following:

$$\frac{1}{\sigma^{q\bar{q}}} \cdot \frac{d\sigma^{q\bar{q}}}{d\cos\theta^*_{+-}} = \frac{1}{2}[1 - \frac{1}{3}\cos\theta^*_{+-}] \tag{F.13}$$

$$\frac{1}{\sigma^{gg}} \cdot \frac{d\sigma^{gg}}{d\cos\theta^*_{+-}} = \frac{1}{2}[1 + \cos\theta^*_{+-}] \tag{F.14}$$

These equations indicate that, in the $q\bar{q}$ production channel, ℓ^+ and ℓ^- tend to *avoid* each other, while in the gg channel, they tend to *attract* each other. Moreover, the tendency of ℓ^+ and ℓ^- to attract one another in the gg production channel is much more pronounced than the tendency for them to avoid each other in the $q\bar{q}$ channel — this can be seen by the fact that the coefficient multiplying $\cos\theta^*_{+-}$ in the $q\bar{q}$ channel is $1/3$ of that in the gg channel. These tendencies can be understood on the basis of the behavior of t and $\bar{t}$ spin at $t\bar{t}$ production threshold.

In the case of $q\bar{q}$ production channel, the matrix elements strongly favor the configuration in which the spin of q and $\bar{q}$ are *aligned along the beam direction.* See figure F.1. The angular distribution of ℓ^+ originating from a top quark with fixed spin direction is:

$$f(\cos\theta_{\ell^+}) = \frac{1}{2}(1 + \cos\theta_{\ell^+}) \tag{F.15}$$

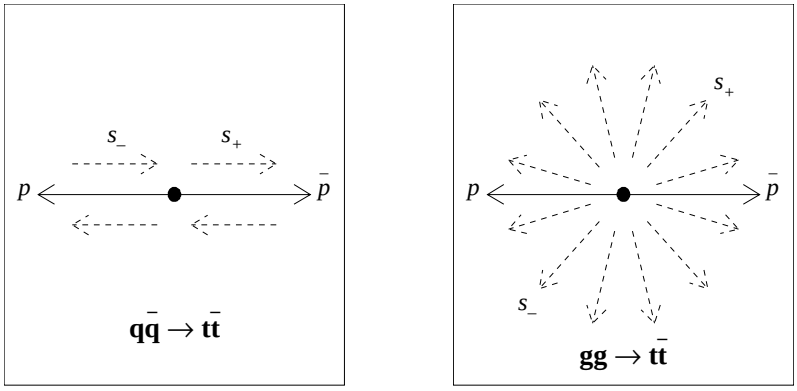


Figure F.1: The t and $\bar{t}$ spin orientation for $q\bar{q}$, $gg \to t\bar{t}$ at threshold. For the $q\bar{q}$ production channel, the spins s_+ and s_- are aligned along the beam. The direction of alignment is equally likely to be along p or $\bar{p}$. For the gg production channel, the spins are anti-aligned. Because of interference effects, the spins are anti-aligned no matter what axis they are projected onto.

The angle θ_{ℓ^+} is that between the momentum of ℓ^+ and the top quark spin. The corresponding distribution for ℓ^- originating from the decay of $\bar{t}$ is:

$$f(\cos\theta_{\ell^-}) = \frac{1}{2}(1 - \cos\theta_{\ell^-}) \tag{F.16}$$

These equations show that ℓ^+ has a strong tendency to move *along* the t spin, and ℓ^- has a strong tendency to move *against* the $\bar{t}$ spin. Since the t and $\bar{t}$ spins are aligned, these equations confirm the prediction in equation F.13 that ℓ^+ and ℓ^- avoid each other.

The factor $1/3$ in front of $\cos\theta^*_{+-}$ in equation F.13 is characteristic of situations in which the t and $\bar{t}$ spin are aligned in a definite direction — in the case of $q\bar{q}$ production, along the beam line. The fact that such a factor is missing in the gg channel indicates that one cannot understand the attraction of ℓ^+ and ℓ^- simply on the basis of spin alignment/anti-alignment. In fact, interference between matrix elements with different spin configurations are crucial in explaining equation F.14.

At the production threshold of $t\bar{t}$ via the gg channel, the $t\bar{t}$ is in the spin singlet state ${}^1S_0$ (by Yang's theorem — see [1] and [34]). The spin singlet state is an anti-symmetric combination of anti-aligned spin states of t and $\bar{t}$: $\frac{1}{\sqrt{2}}[|\uparrow\downarrow> - |\downarrow\uparrow>]$. This is a spatially isotropic wave function, the isotropy of which is attained by interference effects. The fact

that wave function is a combination of anti-aligned t and $\bar{t}$ spin states partly explains the fact that ℓ^+ and ℓ^- attract one another. But the great enhancement in attraction seen in equation F.14 is a result of the fact that the spins are not simply anti-aligned in a fixed direction, but are anti-aligned in such a way that, no matter what direction is chosen as the spin projection axis, the spins come out anti-aligned.

These illustrations of the behavior of the ℓ^+ and ℓ^- angular distributions at $t\bar{t}$ production threshold demonstrate the following:

- The attraction/repulsion between ℓ^+ and ℓ^- from the t and $\bar{t}$ decay is explainable partly from the alignment or anti-alignment of the t and $\bar{t}$ spin.
- Spin alignment alone, however, does not explain everything. Interference effects between matrix elements with different spin configuration can have a very pronounced effect on how ℓ^+ and ℓ^- attract/repel one another.

The behavior of the ℓ^+ and ℓ^- angular distributions away from $t\bar{t}$ production threshold can be understood using the same sort of arguments as given above. The only difference is that, away from threshold, the behavior of the t and $\bar{t}$ spin are more complicated.

F.4 $t\bar{t}$ Spin Correlation Away from Threshold

In the last section, the distribution of $\cos\theta^*_{+-}$ in the $q\bar{q}$ and gg production channels gave a succinct description of the effect of $t\bar{t}$ spin correlation on the angular distribution of ℓ^+ and ℓ^- at the threshold of $t\bar{t}$ production. It turns out that the effect of $t\bar{t}$ spin correlation can be expressed in terms of $\cos\theta^*_{+-}$ even away from the threshold. The $\cos\theta^*_{+-}$ distributions as a function of the top quark relativistic speed β have the following form:

$$\frac{1}{\sigma^{q\bar{q}}}\cdot\frac{d\sigma^{q\bar{q}}}{d\cos\theta^*_{+-}} = \frac{1}{2}[1 - Q(\beta)\cos\theta^*_{+-}] \tag{F.17}$$

$$\frac{1}{\sigma^{gg}}\cdot\frac{d\sigma^{gg}}{d\cos\theta^*_{+-}} = \frac{1}{2}[1 - G(\beta)\cos\theta^*_{+-}] \tag{F.18}$$

The β-dependent coefficients $Q(\beta)$ and $G(\beta)$ are defined as follows:

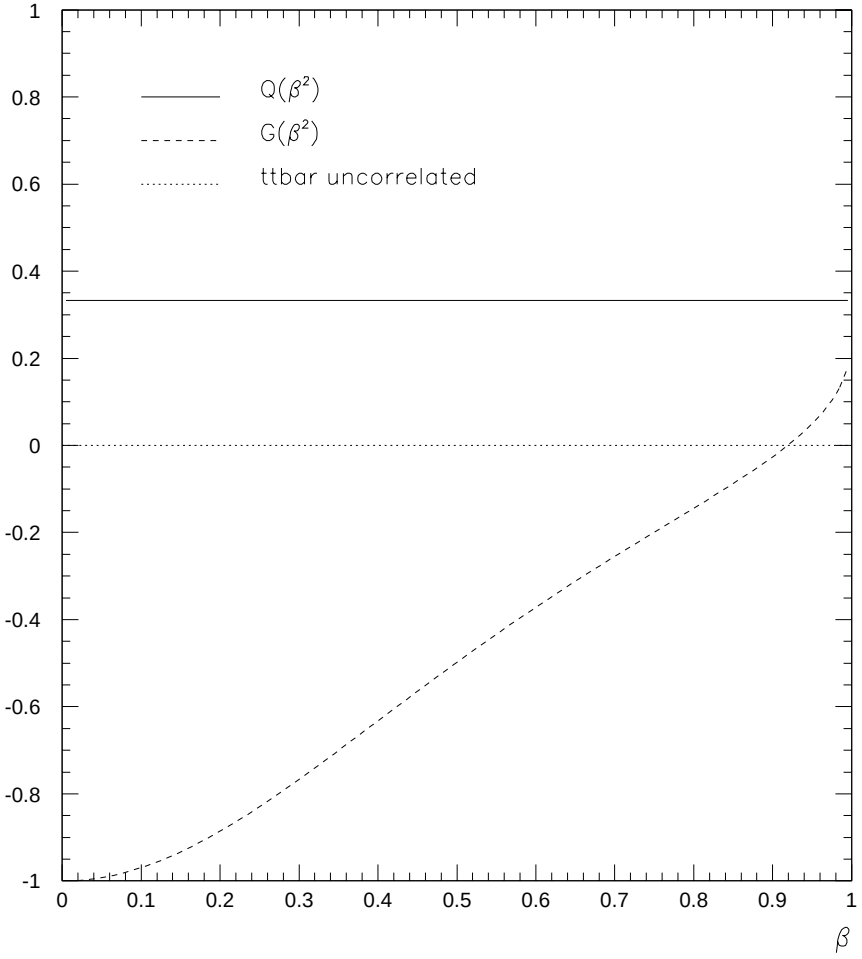


Figure F.2: The functions $Q(\beta)$ and $G(\beta)$. These functions determine the degree of asymmetry in the $\cos\theta^*_{+-}$ distribution. Positive values of the function indicate *repulsion* between ℓ^+ and ℓ^-, and negative values indicate *attraction*.

$$Q(\beta) = \frac{1}{3} \tag{F.19}$$

$$G(\beta) = \frac{-(7 + 3\beta^2 - \frac{2}{3}\beta^4) + \frac{1}{3}(35 - 20\beta^2 + \beta^4)(\frac{1}{\beta}\ln\frac{1+\beta}{1-\beta} - 2)}{(7 - 5\beta^2 + 2\beta^4) + (33 - 18\beta^2 + \beta^4)(\frac{1}{\beta}\ln\frac{1+\beta}{1-\beta} - 2)} \tag{F.20}$$

Equations F.17 and F.18 are generalizations of those appearing in equations F.13 and F.14 — the reader can verify that, in the threshold limit, $Q(\beta) = \frac{1}{3}$, and $G(\beta) = -1$. Figure F.2 shows Q and G plotted against β.

There are a couple of remarkable features about $Q(\beta)$ and $G(\beta)$. First is the fact

that $Q(\beta)$ is $1/3$ for all values of β. This suggests that, even away from the threshold, the t and $\bar{t}$ spin in the $q\bar{q}$ production channel are always aligned along some fixed axis (though not necessarily along the beam line). Second is the fact that $G(1) = 1/3$ (this may not be apparent from the graph in figure F.2 because the approach of $G(\beta)$ to $1/3$ as $\beta \rightarrow 1$ is very abrupt). In other words, in the limit $\beta \rightarrow 1$, $G(\beta) \rightarrow Q(\beta)$. Although the author is not aware of the details, this may indicate the convergence of spin properties in the gg and $q\bar{q}$ production channels in the limit of ultra-relativistic top quark.

F.5 The Top Quark Decay Product Kinematics in the Full Matrix Element Calculation Versus that in the Independent Decay Procedure

The goal of this section is to see what differences exist between the t and $\bar{t}$ decay kinematics in the full matrix element calculation on the one hand, and in the independent decay procedure on the other. Before this can be done, however, the decay kinematics must be parametrized. It was shown in section 7.2 that four parameters are necessary and sufficient to describe the top quark decay in the top rest frame. The same parameters as those described there will be used to describe the t and $\bar{t}$ decay in their respective rest frames. See figure F.3. Note that another set of parameters is needed to describe the t and $\bar{t}$ momentum distributions. However, these distributions are the same in both procedures, so this can be ignored in comparing the decay kinematics.

One thing to note in figure F.3 is the fact that the t and $\bar{t}$ decay are described in the *lepton + jets* channel — *i.e.* one of the top quarks decays semileptonically ($t_\ell \rightarrow b_\ell + \ell + \nu_\ell$) and the other decays hadronically ($t_h \rightarrow b_h + W_d + W_u$). This may seem inconsistent with the description of $t\bar{t}$ spin correlation effects given so far, where the decay was assumed to be in the *dilepton channel*: $t \rightarrow b + \ell^+ + \nu_\ell$ and $\bar{t} \rightarrow \bar{b} + \ell^- + \bar{\nu}_\ell$. The use of the dilepton channel was motivated by the fact that it is the most natural channel to use in describing the spin correlation effect. However, since this thesis deals with the lepton + jets decay channel, the discussion in this section will be made in terms of this channel. The results

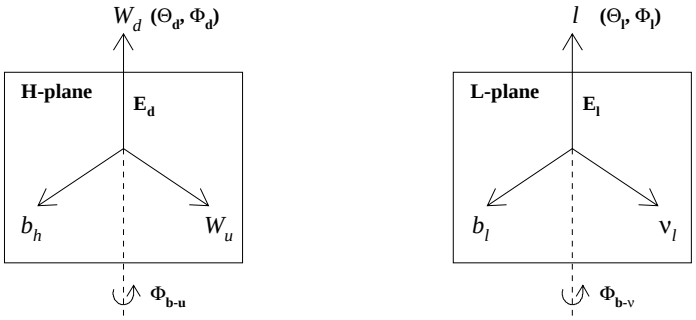


Figure F.3: The parameters used to describe the decay kinematics of t_h and t_ℓ in their respective rest frames. The t_h decay product momentum vectors lie in a plane, and the same for t_ℓ. The decay plane for t_h and t_ℓ are called the *H-plane* and the *L-plane*, respectively.

on spin correlation obtained so far for the dilepton decay channel can be translated to the lepton + jets channel by swapping one of the charged leptons in the dilepton channels with W_d. This is appropriate because both ℓ and W_d have weak isospin $-1/2$.

The t_ℓ decay product momenta lie in a plane. This plane will be referred to as the *L-plane* (because it contains the decay product momenta of the semi-**L**eptonically decaying top quark). It takes three parameters to span all possible orientation of this plane. The three parameters will be chosen to be θ_ℓ, ϕ_ℓ, and $\phi_{b\text{-}\nu}$. The angles θ_ℓ and ϕ_ℓ are the two angles required to define the direction of ℓ in the t_ℓ rest frame, while $\phi_{b\text{-}\nu}$ is the angle of rotation of the L-plane about the line containing the ℓ momentum vector. This ℓ-centeredness of defining the decay plane orientation has good reasons, as will become apparent soon. A single parameter is necessary to fix the magnitude and direction of the vectors in the decay plane. This parameter will be taken as the energy of ℓ, E_ℓ.[1]

The parameters used to describe the decay of t_h is analogous to that for t_ℓ. The parameters θ_ℓ, ϕ_ℓ, $\phi_{b\text{-}\nu}$, and E_ℓ are changed to θ_d, ϕ_d, $\phi_{b\text{-}u}$, and E_d. Also, the decay plane for t_h is referred to as the *H-plane* because it contains the decay product momenta of the **H**adronically decay top quark.

Having defined the parameters to describe the decay of t_ℓ and t_h, one can now

[1] See chapter 7.2 for details on the parameters that describe the top quark decay in the top rest frame.

compare and contrast the decay product kinematics in the two procedures. First of all, let us note the aspects of the kinematics that both procedures have in common.

- The distribution of E_ℓ is independent of the other parameters. That is, no matter what θ_ℓ, ϕ_ℓ, and $\phi_{b\text{-}\nu}$ are, the distribution of E_ℓ is the same. The same goes for the distribution of E_d. The E_ℓ and E_d distributions are given in equation 7.17 in section 7.3.

- The distribution of $\phi_{b\text{-}\nu}$ is also independent of the other parameters. The same goes for the distribution of $\phi_{b\text{-}u}$. These angles are distributed randomly.

This leaves the distribution of the angles θ_ℓ and ϕ_ℓ in the decay of t_ℓ, and θ_d and ϕ_d in the decay of t_h. The distribution of these angles are different between the full matrix element calculation and the independent decay procedure. In the independent decay procedure, all four of these angles are randomly distributed. In the full matrix element calculation, they are distributed according to equations F.9 and F.10. In other words, in the independent decay procedure, the momentum vector of ℓ and W_d point in random direction in the t_ℓ and t_h rest frame, respectively, whereas in the full matrix element calculation, the direction of ℓ in the t_ℓ rest frame affects the direction of W_d in the t_h rest frame, and *vice versa*. Also, the direction and magnitude of the top quark momentum in the $t\bar{t}$ rest frame affects the correlation between the direction of ℓ and W_d momenta.

Let us describe the kinematics of the two procedure from yet another point of view. When the direction of the ℓ and W_d momentum vectors are fixed, the decay kinematics of t_ℓ and t_h in the two procedures are identical. That is, the distribution of the direction and magnitude of the t_ℓ and t_h decay product momenta *in their respective decay planes* and the orientation of the decay planes about the ℓ and W_d axes are exactly the same in the two procedures. The distribution of the ℓ and W_d momentum vectors, however, are different in the two procedures.

F.6 Quantifying the Difference between the Kinematics in

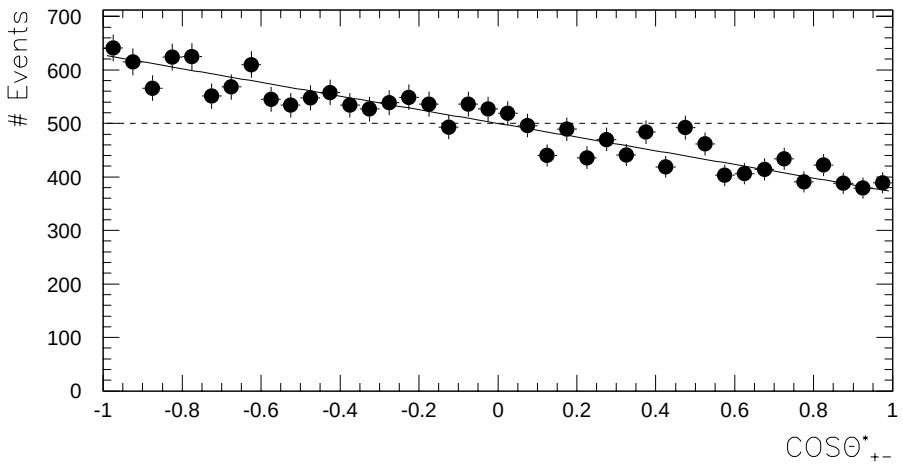


Figure F.4: The distribution of $\cos\theta^*_{+-}$ from monte carlo events generated using the full matrix element calculation (points). The solid line is the analytic distribution obtained from the full matrix element calculation. The dashed line shows the distribution for the independent decay procedure.

the Full-matrix Element Calculation and Independent Decay Method

The difference between the t and $\bar{t}$ decay kinematics in the full matrix element calculation on the one hand, and in the independent decay procedure on the other, can be summarized by the distribution of the angle $\cos\theta^*_{+-}$, which, in this context, is the angle between the ℓ momentum in the t_ℓ rest frame and the W_d momentum in the t_h rest frame. Figure F.4 shows this distribution for the two procedures. The points in the figure show the distribution of $\cos\theta^*_{+-}$ obtained from a monte carlo generator that incorporates the full matrix element calculation. The solid line shows the analytic $\cos\theta^*_{+-}$ distribution obtained from the matrix element. The dashed line is the distribution for the independent decay procedure. The formula for the $\cos\theta^*_{+-}$ distribution from the full matrix element calculation is the following:

$$f(\cos\theta^*_{+-}) = \frac{1}{2}(1 - 0.256 \cdot \cos\theta^*_{+-}) \tag{F.21}$$

The coefficient -0.256 multiplying $\cos\theta^*_{+-}$ assumes that 90% of the $t\bar{t}$ events originate from $q\bar{q}$ collisions, and the rest from gg collisions.

Figure F.4 shows qualitatively that the difference between the top quark decay kinematics in the full matrix element calculation and the independent decay procedure is not very large. In order to quantify this observation, let us imagine performing the following monte carlo exercise:

1. Generate N $t\bar{t}$ events using a monte carlo that incorporates the full matrix element calculation.

2. Obtain the distribution of $\cos\theta^*_{+-}$ from this monte carlo sample.

3. Perform a 2-component log-likelihood fit to the distribution. The two components are the following distributions:

$$f_0(\cos\theta^*_{+-}) = \frac{1}{2} \tag{F.22}$$

$$f_1(\cos\theta^*_{+-}) = \frac{1}{2}(1 - 0.256 \cdot \cos\theta^*_{+-}) \tag{F.23}$$

   The combined function is:

$$f(\cos\theta^*_{+-}) = h \cdot f_1(\cos\theta^*_{+-}) + (1-h) \cdot f_0(\cos\theta^*_{+-}) \tag{F.24}$$

   The function f_1 is the distribution from the full matrix element calculation, and f_0 is that from the independent decay procedure. The fit parameter is h, which can be interpreted as the fraction of events originating from the distribution f_1.

4. Repeat the above procedures N_{exp} times. Each trial is referred to as a *pseudo-experiment.*

When this procedure is carried out, one obtains N_{exp} number of fit parameters h_0 that minimize the log-likelihood function. These values are distributed as a gaussian[2] centered at $h_0 = 1.0$ with width $\sigma = K/\sqrt{N}$, where K is given approximately by the following:

[2]This statement is only approximately correct. In most instances, the approximation is excellent. See appendix M for a discussion on conditions under which the approximation fails.

$$\frac{1}{K^2} = \int dx \, \frac{(f_1(x) - f_0(x))^2}{f_1(x)} \tag{F.25}$$

The variable x is equal to $\cos\theta^*_{+-}$. The result of the integral is:

$$K = 6.63 \tag{F.26}$$

In other words, the statistical error in a 2-component fit is $\sigma = 6.63/\sqrt{N}$.

In order to get a feel for the size of K, consider the error on a binomial distribution, which is $\sigma = 0.5/\sqrt{N}$. A binomial distribution would correspond to an hypothetical situation where the $\cos\theta^*_{+-}$ distribution from the full matrix element calculation has no overlap with that from the independent decay procedure. In this case, one can distinguish events originating from the two methods on an event-by-event basis. In reality, however, figure F.4 shows that the $\cos\theta^*_{+-}$ distributions overlap a great deal. This large overlap makes it difficult for one to distinguish between the two event generation models. This difficulty is quantified by the size of the K-factor: $K = 6.63$. This is $6.63/0.5 = 13.26$ times greater than in a binomial distribution. One consequence of the size of the K-factor is this: for the 2-component fit to have equal statistical significance as the binomial distribution, one needs $13.26^2 = 175.8$ times as many events. This, in some measure, indicates how difficult it is to distinguish — *i.e.* how similar are — the decay kinematics in the two procedures.

Another way to get a feel of how similar the two decay kinematics are is to consider this question:

> Assume that a sample of N $t\bar{t}$ events originate from the full matrix element calculation. What is the minimum number of events required in order to have at least a 97.5% chance that the fit value h_0 in a given experiment lies outside of 2-σ region of the h distribution from the independent decay procedure?

This question can be restated as follows. Suppose one is given two gaussians of approximately equal width $\sigma = K/\sqrt{N}$. One peaks at $h = 0$, the other at $h = 1$. The two gaussians intersect each other at the midpoint $h = 1/2$ for all values of N. How large does N have to be in order for the distance from the peaks to the intersection point is equal to 2-σ? See figure F.5. The equation corresponding to this question is:

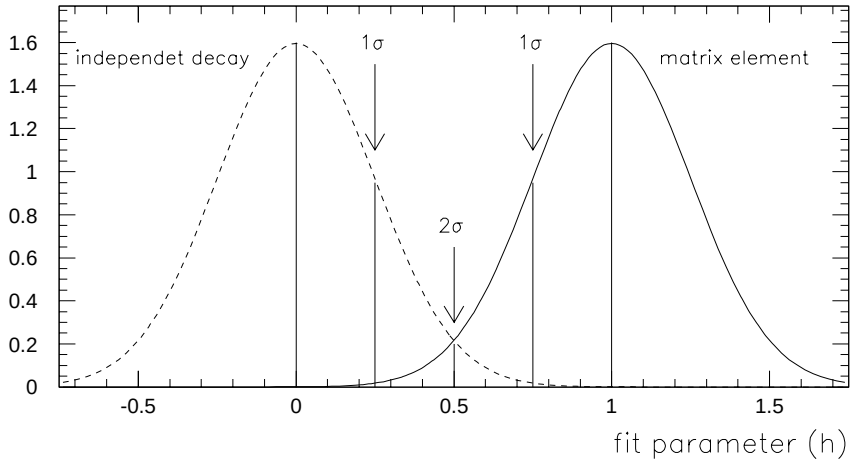


Figure F.5: The distribution of the fit parameter h_0 when the parent distribution is the full matrix element calculation (solid curve) and the independent decay procedure (dashed curve). The distributions are gaussians with approximately the same width, $\sigma = K/\sqrt{N}$. The distributions intersect at the midpoint between the two peaks — *i.e.* $h_0 = 0.5$. The number of events N is chosen so that the distance from the peak to the intersection point is 2-σ.

$$2\sigma = \frac{1}{2} \tag{F.27}$$

Substituting $\sigma = K/\sqrt{N}$ and solving for N, one obtains the following:

$$\begin{aligned} N &= 16K^2 \\ &= 16 \times (6.63)^2 \\ &= 703 \end{aligned} \tag{F.28}$$

Thus about 700 events are needed before the two kinematics can be distinguished from one another at the 2-σ level.

What has been shown up to this point gives a good indication that the kinematics from the two procedures are similar enough so that the choice of which procedure to use to model t and $\bar{t}$ decay is unimportant. Since the conclusions reached above concerned the parton-level distributions, one would expect that after performing the full simulation

— *i.e.* taking account of gluon radiation, converting outgoing quarks and gluons to jets, performing detector simulation, *etc.* — the difference in the kinematics between the two procedures would be even more difficult to detect.

F.7 The Effect of the Difference in Kinematics on the Observables Analyzed in this Thesis

The ultimate test of whether the difference in the kinematics from the two procedures can be ignored or not is to examine the difference in the distributions of the observables examined in this thesis ($\cos U_i$ $(i = 1, 2, 3)$ and E_ℓ). Another quantity that should be compared between the two procedures is $f_{b\ell}$, the fraction of events where the $t\bar{t}$ reconstruction algorithm assigns correctly assigns the correct jet to the b-quark from the decay of t_ℓ.

Figure F.6 shows the observable distributions. The solid histogram is from the independent decay procedure, and the points are from the full matrix element calculation. The distributions are statistically indistinguishable. The b-quark matching fraction $f_{b\ell}$ for the two procedures are:

	$f_{b\ell}$
independent decay	(47.0 ± 0.8) %
full matrix element	(44.9 ± 0.8) %

The difference between them is 2.1%. The statistical error of the *difference* between two quantities with independent statistical errors is $\sigma = \sqrt{\sigma_1^2 + \sigma_2^2}$. In this case, the statistical error of the difference is $\sqrt{2} \times 0.8\% = 1.1\%$. Thus the two values of $f_{b\ell}$ are 1.9 standard deviations away from each other. This may indicate a slight difference between the two values, though it is not unlikely to be a fluctuation.

F.8 Conclusion

The t and $\bar{t}$ decay kinematics in the full matrix element calculation was examined, and it was compared to that in the independent decay procedure. The full matrix element

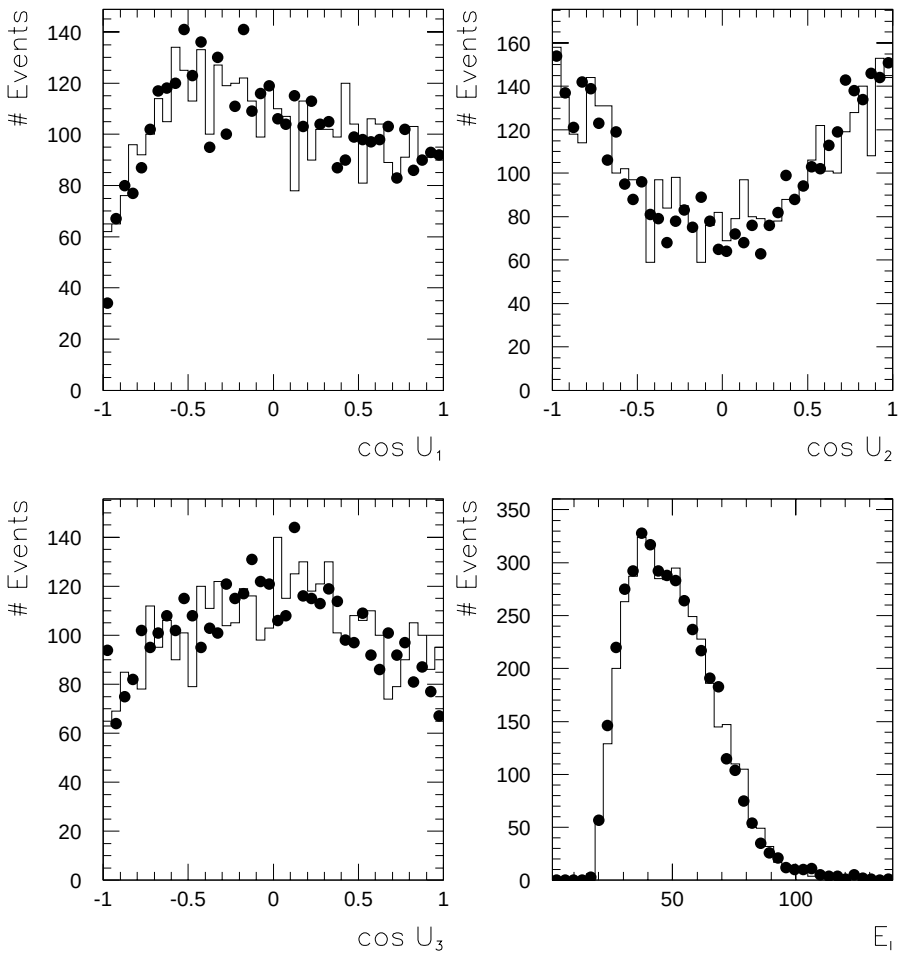


Figure F.6: The distribution of the observables $\cos U_i$ $(i = 1, 2, , 3)$ and E_ℓ from the independent decay procedure (solid histogram) and the full matrix element calculation (points).

calculation takes account of t and $\bar{t}$ spin-alignment and coherence between matrix elements of different spin configurations, whereas the independent decay procedure does not. The difference in the t and $\bar{t}$ decay kinematics for the two procedure is this: in the independent decay procedure, the momentum of ℓ^+ in the t rest frame and that of ℓ^- in the $\bar{t}$ rest frame are randomly distributed; in the full matrix element calculation, they are correlated. In all other respects, however, the t and $\bar{t}$ decay kinematics is the same for the two procedures. In other words, if one focused on the decay kinematics *keeping the direction of* ℓ^+ *and* ℓ^- *fixed*, it is the same in both procedures.

When one plots the cosine of the angle between the momentum of ℓ^+ in the t rest frame and that of ℓ^- in the $\bar{t}$ rest frame ($\cos\theta^*_{+-}$), one finds that the distribution is flat in the independent decay procedure, whereas it is skewed toward negative values in the full matrix element calculation — *i.e.* the ℓ^+ and ℓ^- momenta have some tendency to avoid one another.

The statements made above about the kinematics in the two procedures applies to the lepton + jets decay channel too. To translate from the dilepton channel to the lepton + jets channel, one replaces one of the charged leptons with W_d in the hadronically decaying top quark t_h.

The difference in the t and $\bar{t}$ decay kinematics in the two procedures is small: even at the parton-level, 700 events are necessary to distinguish the $\cos\theta^*_{+-}$ distribution from the two procedures at the 2-σ level. After taking account of gluon radiation, conversion of quarks and gluons to jets, detector simulation, *etc.*, the difference between the kinematics in the two procedures become smaller yet. As far as the observables considered in this thesis are concerned, the difference between the two procedures is undetectable. Thus it is safe to conclude that the choice of which procedure to use to simulate the t and $\bar{t}$ decay kinematics is unimportant.

Appendix G

The Formulas for $d\sigma^{q\overline{q}}/d\hat{t}$ and $d\sigma^{gg}/d\hat{t}$

In this appendix, the differential cross section formulas for obtaining the t and $\overline{t}$ momentum distribution are presented.

The Feynman diagrams used to obtain $d\sigma^{q\overline{q}}/d\hat{t}$ and $d\sigma^{gg}/d\hat{t}$ are shown in figure G.1. The differential cross sections formulas follow:

$$\frac{d\sigma^{q\overline{q}}}{d\hat{t}} = \frac{4\pi\alpha_s}{9s^4}\left[({m_t}^2 - \hat{t})^2 + ({m_t}^2 - \hat{u})^2 + 2{m_t}^2\hat{s}\right] \tag{G.1}$$

$$\begin{aligned}\frac{d\sigma^{gg}}{d\hat{t}} = \frac{\pi\alpha_s^2}{8\hat{s}^2}\ [\ & \tfrac{6(m_t^2-\hat{t})(m_t^2-\hat{u})}{\hat{s}^2} \\ & -\tfrac{m_t^2(\hat{s}-4m_t^2)}{3(m_t^2-\hat{t})(m_t^2-\hat{u})} \\ & +\tfrac{4}{3}\tfrac{(m_t^2-\hat{t})(m_t^2-\hat{u})-2m_t^2(m_t^2+\hat{t})}{(m_t^2-\hat{t})^2} \\ & +\tfrac{4}{3}\tfrac{(m_t^2-\hat{t})(m_t^2-\hat{u})-2m_t^2(m_t^2+\hat{u})}{(m_t^2-\hat{u})^2} \\ & -3\tfrac{(m_t^2-\hat{t})(m_t^2-\hat{u})+m_t^2(\hat{u}-\hat{t})}{\hat{s}(m_t^2-\hat{t})} \\ & -3\tfrac{(m_t^2-\hat{t})(m_t^2-\hat{u})+m_t^2(\hat{t}-\hat{u})}{\hat{s}(m_t^2-\hat{u})} \\ ]\end{aligned} \tag{G.2}$$

The symbols $\hat{s}$, $\hat{t}$, and $\hat{u}$ represent the Mandelstam variables:

$$\hat{s} = (p_1 + p_2)^2 \tag{G.3}$$

$$\hat{t} = (p_1 - t)^2 \tag{G.4}$$

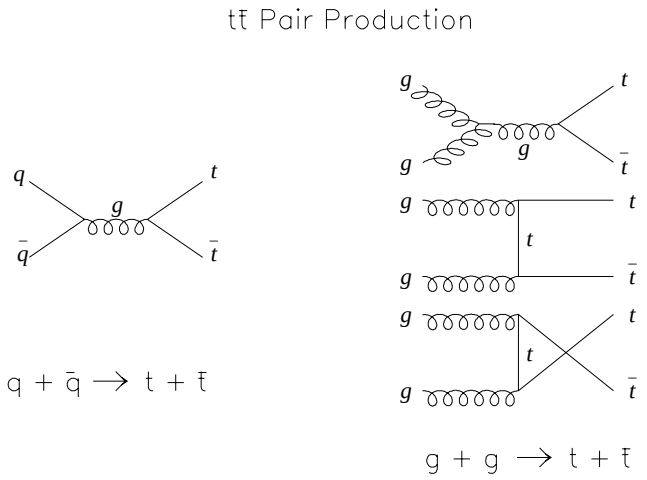


Figure G.1: The tree-level Feynman diagrams for $q\overline{q}$, $gg \to t\overline{t}$.

$$\hat{u} = (p_1 - \overline{t})^2 \tag{G.5}$$

In the formulas above, p_1 and p_2 are the 4-momenta of q and $\overline{q}$ for $q\overline{q} \to t\overline{t}$; for $gg \to t\overline{t}$, they are the 4-momentum of g_1 and g_2. The symbols t and $\overline{t}$ denote the 4-momentum of the top and anti-top quark.

Appendix H

Obtaining the Standard Model Distribution of the Four Parameters from the Matrix Elements

The standard model distribution of the four parameters that describe the top quark decay in the top rest frame can be obtained from the matrix element for the process $t \to \ell + \nu_\ell + b$. For the sake of brevity, the calculation will be performed only for the charge state $t \to \ell^+ + \nu_\ell + b$; the results for the charge conjugate process can be obtained by invoking known C and P asymmetry and CP symmetry. Since the independent decay assumption is used, the top quark is predicted to have no spin polarization. This assumption will be relaxed in the calculations that follow — *i.e.* the possibility that the top quark spin points in a preferred direction will be allowed.

The spin-dependent matrix element for this process is as follows:

$$\begin{aligned} \mathcal{M}_{st,sb,s\ell,s\nu} &= i(\frac{-ig}{2\sqrt{2}})V_{tb}\,\overline{u}(b,s_b)\gamma^\mu(1-\gamma^5)u(t,s_t)\frac{i(g_{\mu\nu}-W_\mu W_\nu/M_W^2)}{W^2-M_W^2+i\Gamma_W M_W} \\ &\quad \times(\frac{-ig}{2\sqrt{2}})\overline{u}(\nu,s_\nu)\gamma^\nu(1-\gamma^5)v(\ell,s_\ell) \end{aligned} \tag{H.1}$$

In this formula, the 4-momentum of each particle is specified, in obvious notation, as b, t, ℓ, ν, and W, and the spin 4-vectors are denoted by s with subscripts for each particle. The numerator of the W propagator can be written as follows:

$$\sum_\alpha \not{\epsilon}^*_\alpha \not{\epsilon}_\alpha = -(g_{\mu\nu} - W_\mu W_\nu/M_W^2), \tag{H.2}$$

where ϵ_α are the polarization vectors of W, and the index α spans the three helicity states

of W. When the summation over helicity states is placed in the numerator of the W propagator, the matrix element becomes:

$$\begin{aligned}\mathcal{M}_{st,sb,s\ell,s\nu} &= -(\frac{g^2}{8})V_{tb}\frac{1}{W^2 - M_W^2 + i\Gamma_W M_W} \\ &\quad \times \sum_{\alpha} \overline{u}(b,s_b)\gamma^{\mu}(1-\gamma^5)u(t,\frac{1}{2})\ \not{\epsilon}_{\alpha}^{*}\ \not{\epsilon}_{\alpha}\overline{u}(\nu,s_\nu)\gamma^{\nu}(1-\gamma^5)v(\ell,s_\ell) \quad \text{(H.3)}\end{aligned}$$

The spin 4-vector for the top quark has been replaced by $1/2$ to indicate the fact that it is *not* being summed over, but fixed in the $+$ direction along some arbitrarily chosen z-direction. When the matrix element is squared, summed over the spins of b, ν, and ℓ, and the resulting expression simplified, it looks like the following:

$$|\mathcal{M}|^2 \sim \sum_{\alpha,\beta} Tr\left[(\not{t} - m_t\ \not{s}_t)\ \not{\epsilon}_{\beta}\ \not{b}\ \not{\epsilon}_{\alpha}{}^{*}(1-\gamma^5)\right] Tr\left[\not{\ell}\ \not{\epsilon}_{\beta}{}^{*}\ \not{\nu}\ \not{\epsilon}_{\alpha}(1-\gamma^5)\right] \qquad \text{(H.4)}$$

The summation indices over the W helicity states, α and β, take on the values $+$, $-$, and 0 for right-handed, left-handed, and longitudinal helicity states, respectively. The first trace describes the decay $t \to b + W$, and the second trace describes $W \to \ell + \nu$. Term-by-term, $|\mathcal{M}|^2$ is given as follows (m_b is set to 0):

α	β	$\underline{t \to b + W^+}$	$\underline{W^+ \to \ell^+ + \nu_\ell}$
$+$	$+$	≈ 0	$\frac{M_W^2}{4}(1+\cos\psi_\ell^*)^2$
$-$	$-$	$2m_tE_b(1-\cos\theta_W)$	$\frac{M_W^2}{4}(1-\cos\psi_\ell^*)^2$
0	0	$m_tE_b(\frac{m_t}{M_W})^2(1+\cos\theta_W)$	$\frac{M_W^2}{2}(1-\cos^2\psi_\ell^*)$
0	$+$	≈ 0	$-\frac{e^{i\delta}}{\sqrt{2}}(\frac{M_W^2}{2})\sin\psi_\ell^*(1+\cos\psi_\ell^*)e^{-i\alpha_\ell}$
0	$-$	$i\frac{e^{i\delta}}{\sqrt{2}}(\frac{m_t}{M_W})2m_tE_b\sin\theta_W$	$\frac{e^{-i\delta}}{\sqrt{2}}(\frac{M_W^2}{2})\sin\psi_\ell^*(1-\cos\psi_\ell^*)e^{i\alpha_\ell}$
$+$	0	≈ 0	$-\frac{e^{-i\delta}}{\sqrt{2}}(\frac{M_W^2}{2})\sin\psi_\ell^*(1+\cos\psi_\ell^*)e^{i\alpha_\ell}$
$-$	0	$-i\frac{e^{-i\delta}}{\sqrt{2}}(\frac{m_t}{M_W})2m_tE_b\sin\theta_W$	$\frac{e^{i\delta}}{\sqrt{2}}(\frac{M_W^2}{2})\sin\psi_\ell^*(1-\cos\psi_\ell^*)e^{-i\alpha_\ell}$
$+$	$-$	0	$-e^{-2i\delta}(\frac{M_W^2}{2})\sin^2\psi_\ell^*e^{2i\alpha_\ell}$
$-$	$+$	0	$-e^{2i\delta}(\frac{M_W^2}{2})\sin^2\psi_\ell^*e^{-2i\alpha_\ell}$

The symbols that appear in the terms are described below:

E_b — b-quark energy in the top rest frame

θ_W — Angle between W's momentum and top spin, top rest frame.

ψ_ℓ^*	Angle between the ℓ and W boost vector, W rest frame.
α_ℓ	Azimuthal angle of ℓ in the plane perpendicular to the W boost direction.
δ	Arbitrary phase factor from the freedom to choose the azimuthal coordinates in the W rest frame

See appendix I for a description of the coordinate systems used to define the angles. It should be noted that the terms above for $t \to b + W^+$ are approximations because m_b was set to zero. This introduces fractional errors on the order of $\mathcal{O}(m_b^2/M_W^2) \approx 0.004$ — *i.e.* the approximate value is on the order of 0.4% off from the correct value. For the same reason, the terms $(\alpha, \beta) = (+,+)$, $(0,+)$, and $(+,0)$ for $t \to b + W^+$ vanish only approximately — *i.e.* they are on the order $m_b^2/M_W^2 \approx 0.004$ times the magnitude of the other terms. The fact that they are approximately zero is due to the $V - A$ coupling of t to W^+ and b — *i.e.* t couples to a left-handed W^+, but is almost decoupled from the right-handed W^+. On the other hand, the terms $(\alpha, \beta) = (+,-)$ and $(-,+)$ vanish exactly: the left-handed and right-handed states don't interfere. The terms $(\alpha, \beta) = (0,-)$ and $(-,0)$ represent the longitudinal-transverse interference; they are comparable in magnitude to the non-interference terms.

The terms in $|\mathcal{M}|^2$ are described by three independent parameters: θ_W, ψ_ℓ^*, and α_ℓ (E_b is constant, and δ is an arbitrary phase factor). These parameters are related the four introduced in chapter 7 in some complicated manner; by appropriately manipulating terms, the results can be expressed in terms of those parameters. The fact that only three parameters are needed implies that at least one of the four parameters is randomly distributed.

When the terms in $|\mathcal{M}|^2$ are added, the resulting expression is:

$$\begin{aligned}\mathcal{F}(\cos\theta_W, \cos\psi_\ell^*, \alpha_\ell) \quad = \quad & 2m_t E_b (\frac{M_W}{2})^2 \{ \\ & (1 - \cos\theta_W)(1 - \cos\psi_\ell^*)^2 \\ & +(\frac{m_t}{M_W})^2 (1 + \cos\theta_W)(1 - \cos^2\psi_\ell^*) \\ & -2\frac{m_t}{M_W} \sin\theta_W \sin\psi_\ell^* (1 - \cos\psi_\ell^*) \sin\alpha_\ell \\ & \} \end{aligned} \tag{H.5}$$

The first two terms inside the curly braces in this equation are non-interference terms, while the third is from interference. In order to express the parameter distribution in terms of those introduced in chapter 7, it is useful to convert the angle ψ_ℓ^* to ψ_ℓ using equation 7.3. When the converted angle is substituted in equation H.5, one obtains the following:

$$\mathcal{G}(\cos\theta_W, \cos\psi_\ell, \alpha_\ell) = \frac{M_W^4}{2}(x^2-1)\frac{8x^4}{(x^2+1)^2}\frac{1-\cos\psi_\ell}{(1-\beta_W\cos\psi_\ell)^2}(1+\hat{\ell}\cdot\hat{s}), \tag{H.6}$$

where $x = m_t/M_W$, β_W is W's relativistic velocity in the top rest frame, and $\hat{\ell}$ and $\hat{s}$ are unit vectors along ℓ's momentum and the top spin, both in the top rest frame. In arriving at this formula, the following relations were used:

$$\hat{s} = (0, -\sin\theta_W, \cos\theta_W) \tag{H.7}$$

$$\hat{\ell} = (\sin\psi_\ell\cos\alpha_\ell, \sin\psi_\ell\sin\alpha_\ell, \cos\psi_\ell) \tag{H.8}$$

$$\hat{s}\cdot\hat{\ell} = -\sin\theta_W\sin\psi_\ell\sin\alpha_\ell + \cos\theta_W\cos\psi_\ell \tag{H.9}$$

Defining θ_ℓ as the angle between $\hat{\ell}$ and $\hat{s}$, it is seen that:

$$G(\cos\theta_\ell, \cos\psi_\ell) = \frac{M_W^4}{2}(x^2-1)\frac{8x^4}{(x^2+1)^2}\frac{1-\cos\psi_\ell}{(1-\beta_W\cos\psi_\ell)^2}(1+\cos\theta_\ell) \tag{H.10}$$

For aesthetic reasons, the variable $\cos\psi_\ell$ will be changed back to $\cos\psi_\ell^*$. The result after the substitution is the following:

$$F(\cos\theta_\ell, \cos\psi_\ell^*) = \frac{8}{3}M_W^4(x^2-1)(1+\frac{x^2}{2})g(\cos\psi_\ell^*)f(\cos\theta_\ell) \tag{H.11}$$

$$f(\cos\theta_\ell) = \frac{1}{2}(1+\cos\theta_\ell) \tag{H.12}$$

$$g(\cos\psi_\ell^*) = h_-\cdot g_b(\cos\psi_\ell^*) + h_0\cdot g_n(\cos\psi_\ell^*) + h_+\cdot g_f(\cos\psi_\ell^*) \tag{H.13}$$

$$h_- = \frac{1}{1 + x^2/2} \tag{H.14}$$

$$h_0 = \frac{x^2/2}{1 + x^2/2} \tag{H.15}$$

$$h_+ = 0 \tag{H.16}$$

$$x = \frac{m_t}{M_W} \tag{H.17}$$

The quantities h_-, h_+, and h_0 are referred to as the *helicity fractions*. The formulas for the helicity fractions are approximations obtained by setting $m_b = 0$. This introduces errors on the order of $\mathcal{O}(m_b^2/M_W^2) \approx 0.004$. The functions $g_i(\cos\psi_\ell^*)$, $i = b, n, f$ are the angular distributions (in the W rest frame) of the charged lepton from the decay of W in the left-handed, longitudinal, and right-handed helicity states:

$$g_b(\cos\psi_\ell^*) = \frac{3}{8}(1 - \cos\psi_\ell^*)^2 \tag{H.18}$$

$$g_n(\cos\psi_\ell^*) = \frac{3}{4}(1 - \cos^2\psi_\ell^*) \tag{H.19}$$

$$g_f(\cos\psi_\ell^*) = \frac{3}{8}(1 + \cos\psi_\ell^*)^2 \tag{H.20}$$

The subscript b, n, and f stand for *backward*, *normal*, and *forward*. They indicate where each function reaches a maximum. See figure 7.6 in section 7.3.

Before leaving this appendix, a couple of points on $f(\cos\theta_\ell)$ and $g(\cos\psi_\ell^*)$ are in order. First, the form of $f(\cos\theta_\ell)$ given in equation H.12 is valid for a 100% spin-polarized top quark. If the degree of spin polarization is α $(0 < \alpha < 1)$, then equation H.12 must be generalized as follows:

$$f(\cos\theta_\ell) = \frac{1}{2}(1 + \alpha \cdot \cos\theta_\ell) \tag{H.21}$$

When $\alpha = 0$, the top quark spin is unpolarized; when $\alpha = 1$, it is 100% polarized. To further generalize the result to accommodate the charge conjugate process $\overline{t} \to \ell^- + \overline{\nu}_\ell + \overline{b}$, the following modification is made:

$$f(\cos\theta_\ell) = \frac{1}{2}(1 + q_\ell\, \alpha \cdot \cos\theta_\ell) \tag{H.22}$$

The quantity q_ℓ is the sign of the charge of ℓ. From this formula, it is seen that ℓ^+ tends to decay toward the t spin vector, whereas ℓ^- tends to decay away from the $\overline{t}$ spin vector.

The second point is about the $\cos\psi_\ell^*$ distribution in the charge-conjugate process $\overline{t} \to \ell^- + \overline{\nu}_\ell + \overline{b}$. In section E.7 of appendix E, it is shown that $g(\cos\psi_\ell^*)$ in the charge-conjugate process is exactly the same as it is in the original process. See appendix E.7 for more details.

Appendix I

The Coordinate System in the Top and the W Rest Frames

The coordinate systems used to describe the angles that appear in formulas in chapter 7 and appendixes H and J will be described here.

Since the decays $t \to W^+ + b$ and $W^+ \to \ell^+ + \nu_\ell$ are being examined, it is useful to set up coordinate systems in the t and the W rest frames. The coordinate system in the t rest frame is illustrated in figure I.1. Note that it is defined in terms of the W momentum and the top spin. This is a natural choice for two reasons. First, since the angular distribution of the charged lepton is of interest in both the W and the top rest frames, defining one of the coordinate axes as the boost direction from the top to the W rest frame simplifies the conversion of angles between the two frames. Second, the use of the top spin vector to define the azimuthal coordinates '1' and '2' allows one to use the azimuthal angle as a convenient means of locating the ℓ^+ momentum relative to the top quark spin.

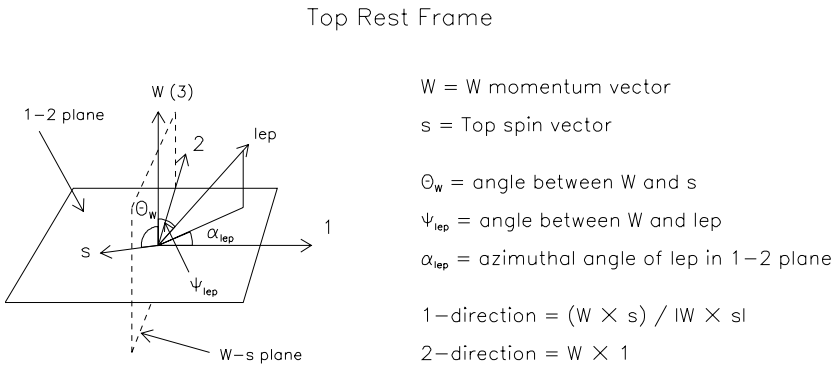


Figure I.1: The coordinate system used in the top rest frame.

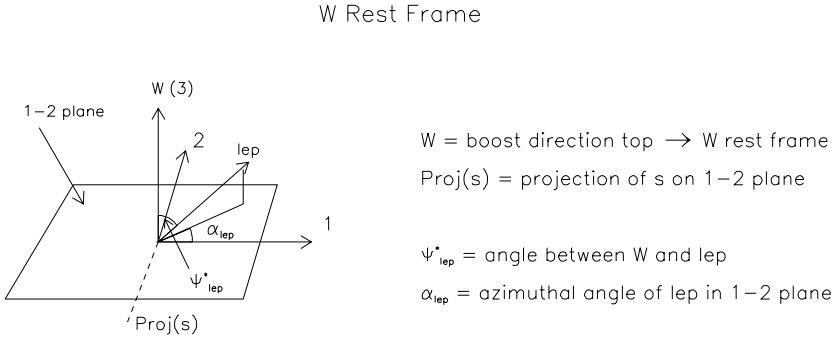


Figure I.2: The coordinate system used in the W rest frame. The azimuthal angle α_ℓ remains unchanged from the top rest frame, but the polar angle ψ^*_ℓ is different from the value ψ_ℓ in the top rest frame. The projection of the top spin vector $\vec{s}$ onto the 1-2 plane is in the negative '2' direction, or $\alpha = 270^o$.

The coordinate system in the W rest frame is shown in figure I.2. Since one boosts along the direction 'W' to go from the top to the W rest frame, the components of momentum in the 1-2 plane remain unchanged. In particular, the azimuthal angle α_ℓ stays the same between the top and the W rest frames. On the other hand, the angle between W and ℓ, ψ_ℓ, changes between the two frames according to the following relation:

$$\cos\psi^*_\ell = \frac{\cos\psi_\ell - \beta_W}{1 - \beta_W \cos\psi_\ell}, \tag{I.1}$$

where β_W is the relativistic velocity of W in the top rest frame (note: β_W is constant in the zero-width approximation for the top and W masses). The asterisk is intended to denote quantities in the W rest frame. Finally, it should be noted that the projection of the top spin vector onto the 1-2 plane is at 270^o in azimuth, or opposite the '2' direction.

Appendix J

The W Interference Effect

The angle $\cos\psi_\ell^*$ was defined in chapter 7 as that between the charged lepton momentum vector in the W rest frame and the boost direction from the top to the W rest frame. This angle is distributed according to the following formula:

$$g(\cos\psi_\ell^*) = h_- \cdot g_b(\cos\psi_\ell^*) + h_0 \cdot g_n(\cos\psi_\ell^*) + h_+ \cdot g_f(\cos\psi_\ell^*) \qquad \text{(J.1)}$$

The functions $g_b(\cos\psi_\ell^*)$, $g_n(\cos\psi_\ell^*)$, and $g_f(\cos\psi_\ell^*)$ are the $\cos\psi_\ell^*$ distributions from the intermediate W in the left-handed, longitudinal, and right-handed helicity state in $t \to W^+ + b$; h_-, h_0, and h_+ are the helicity fractions for the left-handed, longitudinal, and right-handed states. These are all defined in equations 7.13 through 7.15 and 7.9 through 7.11.

The quantities h_-, h_+, and h_0 have these properties: (1) they are non-negative; and (2) $h_- + h_+ + h_0 = 1$. These, together with equation J.1, suggest the following model for the decay of the top quark:

1. First, $t \to W^+ + b$. The resulting W^+ is in the left-handed helicity state a fraction h_- of the time, in the longitudinal state h_0 of the time, and in the right-handed state h_+ of the time. For $m_t = 175\ GeV$, W^+ is in the left-handed, longitudinal, and the right-handed helicity states 30%, 70%, and 0% of the time.

2. The intermediate W^+ with helicity $-$, 0, and $+$ decays to ℓ^+ and ν_ℓ with $\cos\psi_\ell^*$ distributed according to the distribution $g_b(\cos\psi_\ell^*)$, $g_n(\cos\psi_\ell^*)$, and $g_f(\cos\psi_\ell^*)$, respectively.

According to this model, the parameters h_-, h_0, and h_+ are the branching fraction of t to decay into a b and a W^+ in the left-handed, longitudinal, and right-handed helicity states.

This model for the top quark decay, however, is incorrect. If it were correct, then the intermediate W must always be in a helicity eigenstate. The matrix element for $t \to \ell^+ + \nu_\ell + b$, however, is a coherent sum over the W helicity states:

$$\begin{aligned}\mathcal{M}_{st,sb,s\ell,s\nu} &= -(\frac{g^2}{8})V_{tb}\frac{1}{W^2 - M_W^2 + i\Gamma_W M_W} \\ &\quad \times \sum_\alpha \overline{u}(b,s_b)\gamma^\mu(1-\gamma^5)u(t,\frac{1}{2})\ \not{\epsilon}_\alpha^*\ \not{\epsilon}_\alpha \overline{u}(\nu,s_\nu)\gamma^\nu(1-\gamma^5)v(\ell,s_\ell) \quad \text{(J.2)}\end{aligned}$$

The details of this formula can be found in appendix H. What is important here is the fact that the matrix element, before squaring, is a sum over the three helicity states (represented by the summation index α). Upon squaring, and summing over the spin of b, ν_ℓ, and ℓ^+, one obtains the following:

$$|\mathcal{M}|^2 \sim \sum_{\alpha,\beta} Tr\left[(\not{t} - m_t\ \not{s}_t)\ \not{\epsilon}_\beta\ \not{b}\ \not{\epsilon}_\alpha{}^*(1-\gamma^5)\right] Tr\left[\not{\ell}\ \not{\epsilon}_\beta{}^*\ \not{\nu}\ \not{\epsilon}_\alpha(1-\gamma^5)\right] \qquad \text{(J.3)}$$

The summation index α and β is over the W helicity states $-$, 0, and $+$. The terms $\alpha = \beta$ are the non-interference terms, while $\alpha \neq \beta$ are the interference terms. The first trace in the sum describes $t \to W^+ + b$, and the second trace describes $W^+ \to \ell^+ + \nu_\ell$. Term-by-term, the matrix element squared is as follows:

α	β	$t \to b + W^+$	$W^+ \to \ell^+ + \nu_\ell$
$+$	$+$	≈ 0	$\frac{M_W^2}{4}(1+\cos\psi_\ell^*)^2$
$-$	$-$	$2m_tE_b(1-\cos\theta_W)$	$\frac{M_W^2}{4}(1-\cos\psi_\ell^*)^2$
0	0	$m_tE_b(\frac{m_t}{M_W})^2(1+\cos\theta_W)$	$\frac{M_W^2}{2}(1-\cos^2\psi_\ell^*)$
0	$+$	≈ 0	$-\frac{e^{i\delta}}{\sqrt{2}}(\frac{M_W^2}{2})\sin\psi_\ell^*(1+\cos\psi_\ell^*)e^{-i\alpha_\ell}$
0	$-$	$i\frac{e^{i\delta}}{\sqrt{2}}(\frac{m_t}{M_W})2m_tE_b\sin\theta_W$	$\frac{e^{-i\delta}}{\sqrt{2}}(\frac{M_W^2}{2})\sin\psi_\ell^*(1-\cos\psi_\ell^*)e^{i\alpha_\ell}$
$+$	0	≈ 0	$-\frac{e^{-i\delta}}{\sqrt{2}}(\frac{M_W^2}{2})\sin\psi_\ell^*(1+\cos\psi_\ell^*)e^{i\alpha_\ell}$
$-$	0	$-i\frac{e^{-i\delta}}{\sqrt{2}}(\frac{m_t}{M_W})2m_tE_b\sin\theta_W$	$\frac{e^{i\delta}}{\sqrt{2}}(\frac{M_W^2}{2})\sin\psi_\ell^*(1-\cos\psi_\ell^*)e^{-i\alpha_\ell}$
$+$	$-$	0	$-e^{-2i\delta}(\frac{M_W^2}{2})\sin^2\psi_\ell^*e^{2i\alpha_\ell}$
$-$	$+$	0	$-e^{2i\delta}(\frac{M_W^2}{2})\sin^2\psi_\ell^*e^{-2i\alpha_\ell}$

See appendix H for details. The important point to note here is the fact that the interference terms have magnitudes that are comparable to the non-interference terms.

If the view that h_-, h_0, and h_+ are branching fraction were correct, then only the first three terms would be needed to describe the top quark decay kinematics in the top rest frame. The fact that extra terms make large contributions to the sum imply that the interference terms must be influencing the top quark decay in important ways. Just how are these terms influencing the top quark decay? To understand this, it will be useful to examine closely the angular distribution of ℓ^+ in the W^+ rest frame.

The first thing that should be done is to define a coordinate system. The system defined in appendix I is well suited for this purpose. The coordinate system in the top rest frame is defined by the three unit vectors $\hat{u}_1$, $\hat{u}_2$, and $\hat{u}_3$, where $\hat{u}_3$ is the unit vector along the momentum of W, $\hat{u}_1$ is along the cross product $\hat{u}_3 \times \hat{s}$, where $\hat{s}$ is the top quark spin vector, and $\hat{u}_2$ is along $\hat{u}_3 \times \hat{u}_1$. See figure I.1. This coordinate system is used to boost the momentum vector of ℓ and ν_ℓ to the W rest frame. In the W rest frame, the same set of unit vectors are used to define a coordinate system. See figure I.2.

In figure I.1, θ_W is the angle between $\hat{s}$ and $\hat{u}_3$, ψ_ℓ is the angle between $\vec{p}_\ell$ and $\hat{u}_3$, and α_ℓ is the azimuthal angle of $\vec{p}_\ell$ in the 1-2 plane. In the W rest frame, ψ_ℓ^* is the angle between the ℓ momentum vector and $\hat{u}_3$, and α_ℓ is the azimuthal angle. It should be noted that since the boost is defined to be along the 3-direction, the $\psi_\ell \neq \psi_\ell^*$, but the azimuthal angle α_ℓ is the same between the two frame.

One final point about figures I.1 and I.2: by comparing the two coordinate systems, it is clear that the projection of the top quark spin onto the 1-2 plane (denoted $proj(s)$) is in the negative 2-direction. In other words, the $proj(s)$ is at 270^o in the 1-2 plane.

Having defined the coordinate system, let us examine how W decays. First, let us assume that the naive model of the top quark decay is correct — *i.e.* that $t \rightarrow W^+ + b$, where W^+ is in a definite helicity state, and W^+ subsequently decays to $\ell^+ + \nu_\ell$. If this model were correct, then $\cos\psi_\ell^*$ would be distributed according to equation J.1. The azimuthal angle α_ℓ would be distributed randomly.

Next, let us examine the distribution of $\cos\psi_\ell^*$ and α_ℓ when the interference effect is taken account of. When the terms in $|\mathcal{M}|^2$ are added, the following is the result:

$$\mathcal{F}(\cos\theta_W, \cos\psi_\ell^*, \alpha_\ell) \quad = \quad 2m_t E_b (\frac{M_W}{2})^2 \{$$

$$
\begin{aligned}
&(1-\cos\theta_W)(1-\cos\psi_\ell^*)^2 \\
&+(\frac{m_t}{M_W})^2(1+\cos\theta_W)(1-\cos^2\psi_\ell^*) \\
&-2\frac{m_t}{M_W}\sin\theta_W\sin\psi_\ell^*(1-\cos\psi_\ell^*)\sin\alpha_\ell \\
&\}
\end{aligned}
\tag{J.4}
$$

It was stated in appendix H that the first two terms in the curly braces come from the non-interference terms (the terms that one would get from the naive model), and the third term comes from W interference. It is seen that only the interference term depends on α_ℓ. Therefore it is clear that the interference effect is somehow influencing the azimuthal distribution. To understand clearly how the azimuthal distribution is being influenced, let us integrate out the $\cos\theta_W$ and $\cos\psi_\ell^*$ dependence in equation J.4. The result is the following:

$$
\begin{aligned}
\mathcal{A}(\alpha_\ell) &\sim \frac{1}{2\pi}\left\{1-\frac{3}{16}\left(\frac{\frac{\pi^2}{2}(\frac{m_t}{M_W})}{1+\frac{1}{2}(\frac{m_t}{M_W})^2}\right)\sin\alpha_\ell\right\} \\
&= \frac{1}{2\pi}(1-0.598\cdot\sin\alpha_\ell)
\end{aligned}
\tag{J.5}
$$

The numerical coefficient multiplying $\sin\alpha_\ell$ in the second line of the equation assumes $m_t = 175\ GeV$. It is seen that the azimuthal angular distribution peaks at $\alpha_\ell = 270^o$ — *i.e.* toward $proj(s)$.

Let us summarize the effect of W interference:

- The distribution of $\cos\psi_\ell^*$ is the same, whether or not the interference effect is taken account of.
- The distribution of α_ℓ is random without the interference effect. With interference, α_ℓ is pulled toward the projection of the top quark spin in the 1-2 plane.

These observations lead to the following conclusion:

If the top quark spin is unpolarized, then the top quark decay kinematics obtained using the naive model is indistinguishable from that obtained using the full calculation that takes account of the W interference effect.

This conclusion follows because if the top quark spin is unpolarized, then, event-by-event, the top quark spin orientation is random. Thus even though α_ℓ is pulled toward $proj(s)$, since $proj(s)$ is randomly oriented, the distribution of α_ℓ is statistically indistinguishable from a random distribution.

On the other hand, if the top quark spin is polarized, then the top quark decay kinematics obtained using the naive model is different from that obtained using the full calculation. If the naive model is used, the charged lepton angular distribution in the top rest frame is isotropic. If the full calculation is used, the angular distribution of ℓ^+ is pulled toward (ℓ^- is pushed away from) the t ($\bar{t}$) spin polarization vector, which points in a fixed direction in space. Therefore, using the full calculation, the angular distribution is asymmetric relative to the spin polarization vector.

Appendix K

An Important Assumption Underlying the Log-likelihood Fit

K.1 Introduction

In this thesis, the observable distributions from the experimental data are compared to those predicted by monte carlo simulations using the 2-component log-likelihood technique introduced in chapter 8. If the observable is denoted by x, the component functions denoted by $f_0(x)$ and $f_1(x)$, and the fit parameter is α, then the argument of the logarithm to be minimized has the following form:

$$f_\alpha(x) = \alpha \cdot f_1(x) + (1-\alpha) \cdot f_0(x) \tag{K.1}$$

For instance, if x is one of the angular observables $\cos U_i$, then $f_1(x)$ is chosen in this thesis as the distribution of x from a 100% polarized top quark, $f_0(x)$ is that from an unpolarized top quark, and α is the degree of polarization. Or, if $x = E_\ell$, then $f_1(x)$ is the soft energy distribution $g_s(E_\ell)$, $f_0(x)$ is the medium energy distribution $g_m(E_\ell)$, and α is the transverse helicity fraction (*i.e.* left-handed for t decay, right-handed for $\bar{t}$ decay).

At first glance, the validity of equation K.1 may seem obvious. For example, if the degree of polarization is 50%, then it may seem obvious that the distribution of x is obtained by combining 0.5 times the 100% polarized distribution $f_1(x)$ and 0.5 times the unpolarized distribution $f_0(x)$:

$$f_{0.5}(x) = 0.5 \cdot f_1(x) + 0.5 \cdot f_0(x) \tag{K.2}$$

This is certainly true at the analytic stage, since $f_1(x) = (1/2)(1 + q_\ell \cdot x)$, and $f_0(x) = 1/2$. But at the parton-level and reconstructed stage, this simple rule of combination is not necessarily valid. At these stages, equation K.1 is valid if and only the acceptance curve and the smearing matrices are independent of α. The objectives of this appendix are the following:

- Demonstrate the fact that the simple combination rule in equation K.1 is valid if and only if the acceptance curve and smearing matrices are independent of the fit parameter α.
- Demonstrate that, in the standard model top quark decay, the acceptance curve and smearing matrices for the observables $\cos U_i$ and E_ℓ are independent of α.
- Explain what it is about the standard model top quark decay that makes the acceptance curves and smearing matrices independent of α.

K.2 Notation

In chapter 7, the notions of *acceptance curves* and *smearing matrices* were introduced. In that context, they were treated as discrete objects — *i.e.* the acceptance curves were viewed of as histograms with n bins each, and the smearing matrices were viewed of as $n \times n$ matrices. In the present context, it will be more useful to treat them as continuous objects. Thus an acceptance curve will be denoted as $\mathcal{A}(x)$, and the smearing matrix will be written as $\mathcal{M}(x, x')$. In order to express the possibility that the acceptance curves and smearing matrices can depend on α, $\mathcal{A}$ and $\mathcal{M}$ will be given a subscript α: $\mathcal{A}_\alpha(x)$ and $\mathcal{M}_\alpha(x, x')$. In situations where $\mathcal{A}$ and $\mathcal{M}$ are independent of α, the subscript will not be written.

The observable distributions at each stage of analysis — analytic, parton-level, and reconstructed — will be denoted by the superscripts *ana*, *part*, and *recn*, respectively. For

example, the analytic stage distribution of x from a 100% polarized top quark will be denoted $f_1^{ana}(x)$. Using this notation, the relation between the analytic, parton-level, and reconstructed distributions are given as follows:

$$f_\alpha^{part}(x) = \mathcal{A}_\alpha(x)\,[\alpha f_1^{ana}(x) + (1-\alpha) f_0^{ana}(x)] \tag{K.3}$$

$$f_\alpha^{recn}(x) = \int dx'\,\mathcal{M}_\alpha(x,x')\,f_\alpha^{part}(x') \tag{K.4}$$

K.3 The Acceptance Curves

Equation K.3 shows the most general relationship between the observables at the parton-level and analytic stages. Note that, in general, the simple rule of combination given in equation K.1 is not valid:

$$f_\alpha^{part}(x) \neq \alpha \cdot f_1^{part}(x) + (1-\alpha)\cdot f_0^{part}(x) \tag{K.5}$$

The expression for $f_\alpha^{part}(x)$ that is generally valid is the following:

$$\begin{aligned} f_\alpha^{part}(x) \;=\; & \alpha \cdot f_1^{part}(x) + (1-\alpha)\cdot f_0^{part}(x) \\ & + [\mathcal{A}_\alpha(x) - \mathcal{A}_1(x)]\cdot \alpha \cdot f_1^{ana}(x) \\ & + [\mathcal{A}_\alpha(x) - \mathcal{A}_0(x)]\cdot (1-\alpha)\cdot f_0^{ana}(x) \end{aligned} \tag{K.6}$$

The first line is the simple combination rule; the second and third lines are correction terms that take account for the α-dependence of the acceptance curves. It is clear from this equation that the simple combination rule is valid if and only if $\mathcal{A}_\alpha(x) = \mathcal{A}_1(x) = \mathcal{A}_0(x)$ for all α — *i.e.* that $\mathcal{A}_\alpha(x)$ is independent of α.

Now it will be shown that, in the standard model decay of the top quark, the acceptance curves for $\cos U_i$ and E_ℓ satisfy this property. Let us first consider the angular observables $\cos U_i$. In chapter 7.2, it was shown that the top quark decay can be described by four parameters. These four parameters can be taken as $\cos\theta_\ell$, ϕ_ℓ, $\phi_{b\text{-}\nu}$, and E_ℓ. In

this context, $\cos\theta_\ell$ will be taken as the angle between the charged lepton momentum vector $\vec{p}_\ell$ and the direction along the top quark spin polarization vector, both in the top quark rest frame (in other words, $\cos\theta_\ell = \cos U_i$). According to the standard model, these four parameters are distributed according to the following equation:

$$F(\cos\theta_\ell, E_\ell) = f(\cos\theta_\ell)g(E_\ell) \tag{K.7}$$

The dependence of F on the two other parameters ϕ_ℓ and $\phi_{b\text{-}\nu}$ is implied but suppressed because they are randomly distributed. The functions f and g are given in equations 7.7 and 7.17. The special feature of equation K.7 is this: only the function f depends on the degree of spin polarization α.

Let us see what this implies about the acceptance curve for $\cos\theta_\ell$. Imagine a monte carlo generator generating top quark decay using equation K.7. Suppose N events are generated, and $\cos\theta_\ell$ from each event is placed in a histogram. Equation K.7 implies that the top quark spin polarization affects how many events populate each bin. However, for $\cos\theta_\ell$ in a fixed bin, the kinematics of the top quark is independent of α. This is because in a fixed bin, $\cos\theta_\ell$ is approximately constant; equation K.7 shows that when $\cos\theta_\ell$ is fixed, the distribution of the other parameters is independent of the degree of polarization. By definition, the acceptance for events with $\cos\theta_\ell$ in a given bin is the ratio of the number of events in the bin passing event selection cuts to the total number of events in the bin. This ratio can depend only on the kinematics of the top quark decay when $\cos\theta_\ell$ is fixed. But this kinematics was shown above to be independent of α. Therefore the acceptance is independent of α.

Let us now show that the acceptance curve for the observable E_ℓ is independent of the fit parameter α, which, in this case, is the transverse helicity fraction r_T. Before proceeding with the proof, let us modify equation K.7 so that $g(E_\ell)$ is changed to the following:

$$\tilde{g}(E_\ell) = r_T \cdot g_s(E_\ell) + (1 - r_T)g_m(E_\ell) \tag{K.8}$$

In other words, $g(E_\ell)$ in K.7 has r_T fixed to $1/(1+x^2/2)$, with $x = m_t/M_W$. In $\tilde{g}(E_\ell)$, r_T is a fit parameter which is allowed to vary. With this change, let us imagine generating N monte

carlo events. Let us imagine placing E_ℓ from each event into a histogram. Equation K.7 implies that the parameter r_T affects how many events populate each E_ℓ bin. However, for E_ℓ in a fixed bin, the kinematics of the top quark decay is independent of r_T. The rest of the argument leading to the final result — that the acceptance curve for E_ℓ independent of r_T — proceeds in exactly the same manner as for $\cos U_i$.

The results above on the independence of the acceptance curves on the respective fit parameters can be expressed succinctly as follows:[1]

> If the top quark decay is described by the parameters $\cos\theta_\ell$, ϕ_ℓ, $\phi_{b\text{-}\nu}$, and E_ℓ, then the presence or absence of top quark spin polarization can be determined only by examination of $\cos\theta_\ell$. If $\cos\theta_\ell$ is fixed, there is no way to determine the existence or absence of top quark spin polarization. Similarly, the transverse helicity fraction r_T can be determined only by examining the distribution of E_ℓ. If E_ℓ is fixed, one cannot tell what value r_T has.

K.4 The Smearing Matrices

The proof that the simple combination rule applies to the reconstructed observable distributions follows the same line of argument as that for the parton-level distributions. In fact, the simple combination rule applies to the reconstructed observable distributions if and only if it applies to the parton-level distributions. The analog of equation K.6 for the reconstructed distributions is the following:

$$\begin{aligned} f_\alpha^{recn}(x) &= \alpha \cdot f_1^{recn}(x) + (1-\alpha)\cdot f_0^{recn}(x) \\ &\quad + \int dx' \left[\mathcal{M}_\alpha(x,x') - \mathcal{M}_1(x,x')\right]\cdot \alpha \cdot f_1^{part}(x') \\ &\quad + \int dx' \left[\mathcal{M}_\alpha(x,x') - \mathcal{M}_0(x,x')\right]\cdot (1-\alpha) \cdot f_0^{part}(x') \end{aligned} \tag{K.9}$$

This is the most general relationship between the reconstructed and parton-level distributions. The first line in this equation is the simple combination rule; the second and third lines are modifications that result from the α dependence of the smearing matrices. It is

[1] This statement assumes that the α and r_T are to be determined by the shape of distributions; a counting experiment would allow one to determine α and r_T even if the respective observables are fixed.

clear that the simple combination rule is valid if and only if the smearing matrices are independent of α.

The proof that, according to the standard model, the smearing matrices for the observables $\cos U_i$ and E_ℓ are independent of the fit parameters α and r_T, again, follows the same line of argument as that for the acceptance curves. The only difference between the two situations is that an acceptance curve assigns to each input observable value x a single number $\mathcal{A}(x)$, whereas a smearing matrix assigns to each input observable value x' a distribution of output values x given by the function $\mathcal{M}(x, x')$. Since the wording of the proof is almost exactly the same in both cases, the statement of the proof will be omitted.

K.5 Counterexamples

So far, it has been shown that, in the standard model, the simple combination rule of equation K.1 is valid at every stage of analysis — analytic, parton-level, and reconstructed. The very intuitive nature of the simple combination rule may give the impression that the rule must be true in most practical cases, and that deviations from the rule are exceptions, or, in other words, pathologies. But this is not the case. Simple and plausible changes to the standard model distribution of the parameters $\cos\theta_\ell$, ϕ_ℓ, $\phi_{b\text{-}\nu}$, and E_ℓ can destroy the conditions required for the simple combination rule to be valid.

For example, let us take the angular observables $\cos U_i$. The validity of the simple combination rule for this observable hinges on the fact that, in equation K.7, the spin polarization dependence was contained only in the function $f(\cos\theta_\ell)$. In other words, the function $g(E_\ell)$ is independent of α. This function has the following form:

$$g(E_\ell) = r_T \cdot g_s(E_\ell) + (1 - r_T) g_m(E_\ell) \tag{K.10}$$

The parameter r_T in this equation is constant — it depends only on the masses of t, b, and W (the b-quark mass in this equation is approximated as being zero). In a more complicated world, however, it is not hard to imagine that r_T might be influenced by the top quark spin. If this were the case, r_T would depend on α, and therefore the E_ℓ distribution would depend on α:

$$g(E_\ell) = r_T(\alpha) \cdot g_s(E_\ell) + (1 - r_T(\alpha)) g_m(E_\ell) \tag{K.11}$$

In a world where this is the case, the simple combination rule will be invalid, and the measurement of top quark spin polarization would be much more complicated.

Similarly, for the observable E_ℓ, the simple combination rule hinges on the fact that the distribution $\cos\theta_\ell$ and E_ℓ are separable — *i.e.* that the function $F(\cos\theta_\ell, E_\ell)$ can be expressed as a product of functions of $\cos\theta_\ell$ and E_ℓ. In effect, this rule says that no matter what helicity state the intermediate W is in, the distribution of parameters other than E_ℓ have the same form. Specifically, in the standard model, the function $F(\cos\theta_\ell, E_\ell)$ can be written as follows:

$$F(\cos\theta_\ell, E_\ell) = f(\cos\theta_\ell)\ [r_T \cdot g_s(E_\ell) + (1 - r_T) g_m(E_\ell)] \tag{K.12}$$

This shows that, regardless of whether W is in the transverse or longitudinal helicity state, the distribution of $\cos\theta_\ell$ is given by the function $f(\cos\theta_\ell)$. In a more complicated world, however, the transverse and longitudinal states of W may have different distributions of $\cos\theta_\ell$:

$$F(\cos\theta_\ell, E_\ell) = [r_T \cdot g_s(E_\ell)\ f_s(\cos\theta_\ell) + (1 - r_T) g_m(E_\ell)\ f_m(\cos\theta_\ell)] \tag{K.13}$$

In this world, again, the conditions necessary to validate the simple combination rule for E_ℓ fail, and the measurement of the transverse helicity fraction r_T would be much more complicated.

Appendix L

Measurement Resolution: Top Rest Frame Versus Lab Frame Observables

L.1 Introduction

In section 7.5.3, the benefits and liabilities of using observables defined in the lab frame on the one hand, and in the top rest frame on the other, were given. One of the liabilities of top rest frame observables is the fact that the $t\bar{t}$ candidate sample is restricted to *reconstructable* events — *i.e.* events with e or μ, large $\not\!\!E_T$, and four or more jets. Lab frame observables, in contrast, do not require full event reconstruction, so the candidate events satisfying looser cuts and events in the dilepton decay channel can be included. In other words, the candidate data sample for the lab frame observables is a superset of that for the top rest frame. This translates to increased sample size, which, in turn, implies smaller statistical error. In this appendix, the measurement resolutions of lab frame and top rest frame observables are estimated, and they are compared with each other. The goal of this appendix is to estimate the degree of improvement in measurement resolution obtained by using the lab frame observables instead of those based in the top rest frame.

L.2 The $t\bar{t}$ Candidate Event Sample

The $t\bar{t}$ candidate event sample for top rest frame observables consist, basically, of e

or μ, large $\not\!\!E_T$, and four or more jets. For the purpose of optimizing the measurement resolution, this event sample is subdivided into eight mutually exclusive subsets (see section 8.2.4). The event sample is first categorized according to the number of *stiff jets*. Events with only three stiff jets and one or more loose jets are referred to as "3.5-jet events", and are given the label NJ3.5. Events with four or more stiff jets are called "4-jet events", and are labeled NJ4. These two subsets are further subdivided according to the type of b-tagged jets in the event: (1) SVX only; (2) SLT only; (3) SVX and SLT; and (4) No Tags. These categories are labeled XO, TO, XT, and NT, respectively. The number of events and the estimated background fraction in each subsample is given in table L.1.

In analyses using lab frame observables, one can, in addition to the above, include $t\bar{t}$ candidate events satisfying looser cuts and those decaying in the dilepton channel. Exactly which events to add is arguable; for the sake of concreteness, those given in [63] will be used[1]. In that report, there are two additional classes of events: (1) those with e or μ, large $\not\!\!E_T$, and exactly three stiff jets, no loose jets, and at least one SVX b-tag; and (2) those in the dilepton decay channel. The number of events and the estimated background fraction in these non-reconstructable $t\bar{t}$ candidate sample is shown at the bottom of table L.1. A couple of words of caution about the dilepton channel is in order. First, only dilepton events in the e-μ decay channel is included because the modeling of the background shapes in the e-e and the μ-μ channel was found by the principal authors of [63] to be questionable. Second, there are 7 e-μ events, but table L.1 gives 14. This is because each dilepton event contributes two charged leptons, so one has two measurements of charged lepton observables for each event.

L.3 The Measurement Resolution

The following material on measurement resolution is based on the results of chapter 8. In that chapter, the combined error of a data sample subdivided into subsamples is given by the following formula:

[1] Reference [63] presents a measurement of the longitudinal helicity fraction in semileptonic top quark decay using the $P_T(\ell)$ distribution.

$t\bar{t}$ CANDIDATE EVENT SAMPLE			N_i	β_i	$(1-\beta_i)^2 N_i$
FULLY	NJ3.5	XO	4	0.29	2.02
RECONSTRUCTABLE		TO	5	0.66	0.58
		XT	4	0.18	2.69
		NT	77	0.90	0.77
	NJ4	XO	11	0.07	9.51
		TO	9	0.25	5.06
		XT	4	0.04	3.69
		NT	45	0.58	7.94
				TOTAL:	32.26
NOT FULLY	W + 3 jets + SVX		11	0.58	1.94
RECONSTRUCTABLE	DILEPTON		14	0.11	11.1
				TOTAL:	13.04

Table L.1: The number of events N_i and the estimated background fraction β_i in each subset of the fully and not-fully reconstructable $t\bar{t}$ candidate event samples. The quantity $(1-\beta_i)^2 N_i$ is a measure of the statistical importance of given subsample — the larger this quantity is, the smaller the measurement error, and, therefore, the more important this subsample is.

$$\frac{1}{\sigma^2} = \sum_i \frac{1}{\sigma_i^2} \tag{L.1}$$

The index i spans all of the subsamples, σ_i is the measurement error for subsample i, and σ is the combined error. The subsample error σ_i is given approximately by the following formula:

$$\sigma_i = \frac{K_i}{(1-\beta_i)\cdot\sqrt{N_i}} \tag{L.2}$$

The quantities N_i and β_i are the number of events and estimated background fraction for subsample i. The quantity K_i depends on the shape of the component functions in a 2-component fit; the more distinguishable the component functions are, the smaller the K_i, and, therefore, the smaller the error. Since K_i depends on the shape of the component functions, it is observable-dependent.

L.4 The Measurement Resolution for Top Rest Frame and Lab Frame Observables

Since the lab frame observables are based on a data sample that is a superset of that for top rest frame observables, the measurement error of the former is, by necessity, smaller than that of the latter. In this section, this fact is examined quantitatively.

The results of the last section show that the measurement error σ is given by the following:

$$\frac{1}{\sigma^2} = \sum_i \frac{(1-\beta_i)^2 \cdot N_i}{K_i^2} \tag{L.3}$$

If the observable in question is of the top rest frame, the sum is over the reconstructable subsamples; if it is of the lab frame, then the sum is over both reconstructable and non-reconstructable events.

Before proceeding, let us make the following approximations/assumptions about the K-factors:

- For a given observable, the K-factors of all the subsamples are approximately equal. Table 8.2 shows that this is a fairly reasonable assumption.

- The K-factor of a top rest frame observable is approximately the same as that of the corresponding lab-frame observable. This fact is born out in table 7.2 of chapter 7. For example, the K-factor for the r_T measurement using E_ℓ is about the same as that of the same measurement using $P_T(\ell)$.

- The K-factor of a lab frame observable in non-reconstructable events is approximately the same as that in reconstructable events. No attempt will be made to justify this quantitatively. However, since lab frame observables are smeared by the top quark momentum, and the underlying top quark momentum should be about the same regardless of reconstructability, this assumption is believed to be a fairly reasonable one.

These approximations/assumptions allow one to factor out K_i from equation L.3. Thus one obtains for σ the following:

$$\frac{1}{\sigma^2} = \frac{1}{K^2} \cdot \sum_i (1-\beta_i)^2 \cdot N_i \tag{L.4}$$

Let us denote the error for a top rest frame observable as σ_{TRF}, and that for the corresponding lab frame observable as σ_{LF}. Let us express the errors as follows:

$$\frac{1}{\sigma_{TRF}^2} = \frac{1}{K^2} \cdot \sum_R (1-\beta_i)^2 \cdot N_i \tag{L.5}$$

$$\frac{1}{\sigma_{LF}^2} = \frac{1}{K^2} \left[\sum_R (1-\beta_i)^2 \cdot N_i + \sum_{NR} (1-\beta_j)^2 \cdot N_j \right] \tag{L.6}$$

The summation $\sum_R$ is over reconstructable events, and $\sum_{NR}$ is over non-reconstructable events. The ratio of these two errors is the following:

$$\frac{\sigma_{LF}^2}{\sigma_{TRF}^2} = \frac{1}{1+R} \tag{L.7}$$

$$R = \frac{\sum_{NR}(1-\beta_j)^2 \cdot N_j}{\sum_R (1-\beta_i)^2 \cdot N_i} \tag{L.8}$$

Table L.1 shows that $\sum_R (1-\beta_i)^2 \cdot N_i = 32.26$, while $\sum_{NR}(1-\beta_j)^2 \cdot N_j = 13.04$. Thus $R = 0.40$, so $\sigma_{LF}^2/\sigma_{TRF}^2 = 1/1.40 = 0.71$. The ratio of the error of the lab frame observable to that of the corresponding top rest frame observable is, therefore, $\sigma_{LF}/\sigma_{TRF} = 0.84$ — the use of lab frame observables allows one to decrease the measurement error by 16%. This is equivalent to an increase in statistics of non-reconstructable events by a factor of $1/0.84^2 = 1.40$, or a 40% increase. It should be noted that, because of the assumptions/approximations about the K-factors made above, this result is applicable to all observables.

Appendix M

Non-ideal Behaviors in the Statistical Properties of the Minimum Log-likelihood Method

M.1 Introduction

It was stated in section 8.2.2 that, for the observables considered in this thesis, the two-component log-likelihood method has, to a good approximation, "ideal" statistical behaviors. In this context, "ideal" is defined as follows:

> In a set of pseudo-experiments, the distribution of parameter values $\{\alpha_i\}$ that minimize the log-likelihood function in each experiment is a gaussian centered at the true parameter value α_0 with width σ given by:
>
> $$\frac{1}{\sigma^2} = \left.\frac{\partial^2 \mathcal{L}}{\partial \alpha^2}\right|_{\alpha=\alpha_0} \qquad \text{(M.1)}$$
>
> Deviations from this ideal can take place in the following ways:

- A significant fraction of experiments have log-likelihood functions with no minimum.
- The parameter distribution has long, non-gaussian tails.
- The distribution is biased — *i.e.* the mean of the parameter distribution is not equal to α_0.
- The width of the parameter distribution is not given by equation M.1.

In most situations, these non-ideal behaviors decrease with increased statistics. Under certain conditions, however, the deviations from the ideal behaviors persist no matter how large the statistics.

In this appendix, the first three non-ideal behaviors are examined. In appendix O, the deviation of σ from the prediction in equation M.1 is examined. The following is an overview of this appendix:

- The non-ideal behaviors are illustrated with examples.
- The log-likelihood function is analyzed to determine factors responsible for the non-ideal behaviors.
- The necessary condition for the existence of pathologic behavior is shown.
- The fact that the observables in this thesis satisfy ideal conditions is shown.

M.2 Illustrating the Non-ideal Behaviors

For the purpose of illustration, let us take the top polarization measurement at the analytic level. Then the component functions $f_0(x)$ and $f_1(x)$ for unpolarized and 100% polarized top quark are:

$$f_0(x) = \frac{1}{2} \tag{M.2}$$

$$f_1(x) = \frac{1}{2}(1+x) \tag{M.3}$$

If the true parameter value is α_0, then the parent distribution is:

$$\begin{aligned} f_{\alpha_0}(x) &= \alpha_0 \cdot f_1(x) + (1-\alpha_0) \cdot f_0(x) \\ &= \frac{1}{2}(1+\alpha_0 \cdot x) \end{aligned} \tag{M.4}$$

Figure M.1 shows the distribution of α_{min} in one thousand pseudo-experiments with five events per experiment ($N_{exp} = 1000$, $N_{ev} = 5$), where α_{min} is the value of α

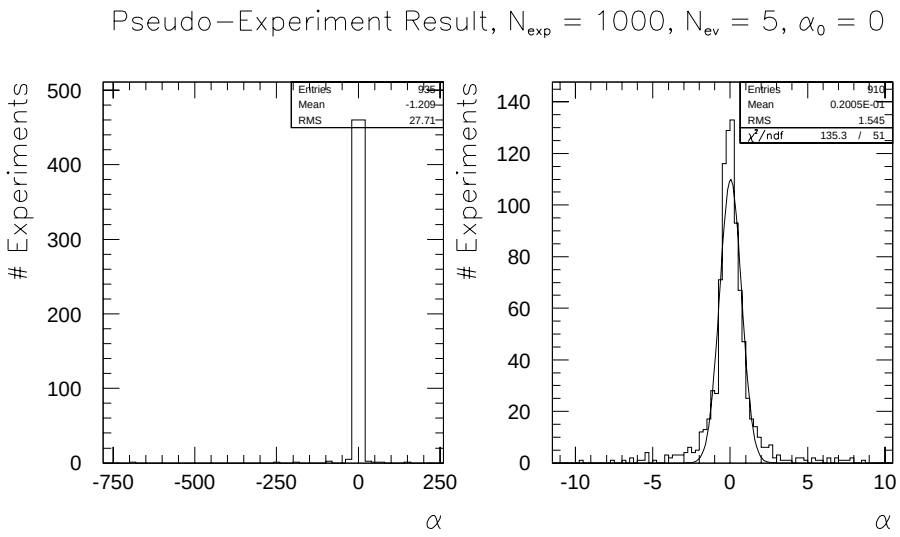


Figure M.1: The distribution of α from a set of 1000 pseudo-experiments with 5 events per experiment. The true parameter value α_0 is zero. Left plot: the full range of the distribution. Right plot: restricting attention to the region $-10 < \alpha < 10$. The curve in the right-hand plot is a best-fit gaussian for the distribution.

that minimizes the log-likelihood function in each experiment. The true parameter value α_0 is zero — *i.e.* the top quark is unpolarized. The following are some observations on the distribution of α:

- The distribution has very long tails, extending to $\alpha \sim 10^2$.
- Even when attention is restricted to $-10 < \alpha < 10$, a significant non-gaussian tail is seen.
- In the left-hand plot in figure M.1, the number of histogram entries is 935. This is 65 short of the 1000 pseudo-experiments performed. These missing experiments have log-likelihood functions with no minimum.
- The mean of the distribution is, within statistical error, consistent with zero. The distribution, therefore, is unbiased.

It is seen that, at $N_{ev} = 5$, all but one of the non-ideal behaviors described earlier are present. The fact that the distribution is not biased is due to the fact that $\alpha_0 = 0$. This is discussed further in a later example.

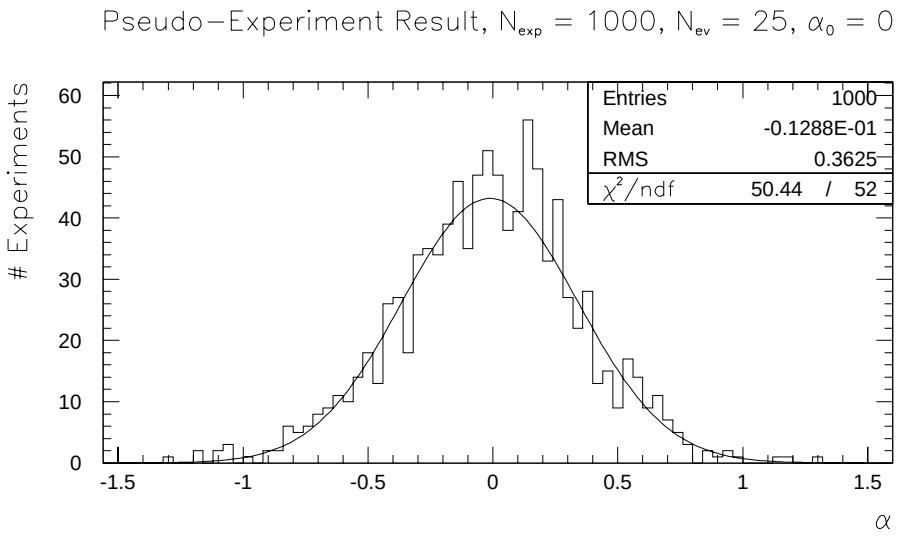


Figure M.2: The distribution of α for $N_{exp} = 1000$, $N_{ev} = 25$, $\alpha_0 = 0.0$.

Figure M.2 shows the distribution of α when N_{ev} is increased to 25. It is seen that the non-gaussian tail is no longer present, a gaussian fit to the distribution gives a good χ^2 value, and all experiments have log-likelihood functions with minimum value.

Figure M.3 shows the distribution of α_{min} when $N_{ev} = 5$ and $\alpha_0 = 0.5$. This situation has all of the non-ideal behaviors in the earlier example with $N_{ev} = 5$, $\alpha_0 = 0$. This case, however, is different from the previous one in the following ways:

- The distribution is biased. Specifically, the mean of the distribution is larger by many standard deviations compared to the true parameter value.
- The number of experiments with no minimum is 103. This is significantly larger than 65 for $\alpha_0 = 0$. The expected fraction of events with no minimum, can, in fact, be predicted using a simple formula, which is described in section M.3.2.

Figure M.4 shows the distribution of α_{min} when N_{ev} is increased to 40. With this level of statistics, none of the non-ideal behaviors exist.

As a final illustration, figure M.5 shows the α distribution when $\alpha_0 = 1$ and $N_{ev} = 1000$. $\alpha_0 = 1$ corresponds to the largest possible physical value of top quark spin polarization. Even though the number of events per experiment is very large, the distribution has moderate, but statistically significant deviations from ideal behaviors. Specifically, the

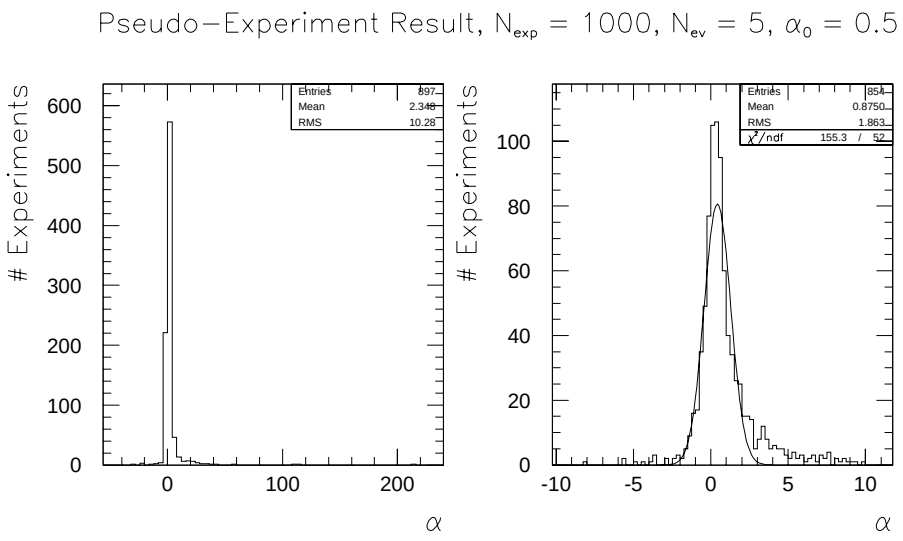


Figure M.3: The distribution of α from a set of 1000 pseudo-experiments with 5 events per experiment. The true parameter value α_0 is 0.5. Left plot: the full range of the distribution. Right plot: restricting attention to the region $-10 < \alpha < 10$. The curve in the right-hand plot is a best-fit gaussian for the distribution.

distribution has a non-gaussian tail in the region $\alpha > 1$, and the distribution is biased. It turns out that, when $\alpha_0 = 1$, the non-ideal behaviors persist, no matter how large N_{ev} is made. This sort of situation is exceptional, and occurs only when the component functions $f_0(x)$ and $f_1(x)$ satisfy a certain condition. This is discussed further in section M.3.2.

M.3 Origin of the Non-ideal Behaviors

In this section, the log-likelihood function is analyzed, and the essential elements in explaining the non-ideal behaviors in the α distribution are presented. In the first part of this section, a plausible but incorrect explanation for the origin of the non-ideal behavior is given. This is followed by a correct version of the explanation. The reason why an incorrect explanation is described is that certain features of the ideal α distribution suggest that the non-ideal behavior can be explained by the n^{th} derivative of the log-likelihood function. The chain of reasoning leading to this incorrect inference is suggestive and plausible, and it can lead one to a dead-end in trying to explain the origin of the non-ideal behavior of

the α_{min} distribution. The section on the incorrect explanation is intended as a "sign post" warning of the dead end.

M.3.1 Incorrect Explanation

Let us start this section by noting the functional form of the ideal α distribution:

$$F(\alpha) = C \cdot \exp\left[-\frac{1}{2}\left.\frac{\partial^2 \mathcal{L}}{\partial \alpha^2}\right|_{\alpha=\alpha_0} \cdot (\alpha - \alpha_0)^2\right] \tag{M.5}$$

The factor C is a normalization constant. By re-defining C, this equation can be re-written as follows:

$$F(\alpha) = C \cdot \exp\left[-\mathcal{L}(\alpha_0) - \frac{1}{2} \cdot \left.\frac{\partial^2 \mathcal{L}}{\partial \alpha^2}\right|_{\alpha=\alpha_0} \cdot (\alpha - \alpha_0)^2\right] \tag{M.6}$$

If the continuum approximation of the log-likelihood function is used, the argument of the exponential can be taken as the first three terms in the Taylor expansion of $\mathcal{L}(\alpha)$:

$$-\mathcal{L}(\alpha) \approx -\left[\mathcal{L}(\alpha_0) + \left.\frac{\partial \mathcal{L}}{\partial \alpha}\right|_{\alpha=\alpha_0} \cdot (\alpha - \alpha_0) + \frac{1}{2} \cdot \left.\frac{\partial^2 \mathcal{L}}{\partial \alpha^2}\right|_{\alpha=\alpha_0} \cdot (\alpha - \alpha_0)^2\right] \tag{M.7}$$

The linear term does not appear in equation M.6 because the first derivative of $\mathcal{L}$ evaluated at $\alpha = \alpha_0$ is zero (this must be true if the most likely value of α is α_0).

Given that the α distribution in the ideal limit is given by the exponential of the first three terms in the Taylor series of $-\mathcal{L}(\alpha)$, it seems reasonable to think that, in general, the α distribution is given by $F(\alpha) = C \cdot \exp[-\mathcal{L}(\alpha)]$. Then it would be natural to assume that the non-ideal behavior originates from the higher-order derivatives of $\mathcal{L}(\alpha)$. This would explain the non-ideal behaviors shown in section M.2:

- As the number of events increases, the range of $(\alpha - \alpha_0)$ gets smaller. Thus the higher-order terms become less and less important with increasing N_{ev}. This would explain the observation that the non-ideal behavior decreases with increasing N_{ev}.

- The bias in the distribution seems explainable by the odd-order derivatives of $\mathcal{L}$. This is consistent with the situation in section M.2, where no bias is observed when

$\alpha_0 = 0$, but it is observed when $\alpha_0 = 0.5$ and $\alpha_0 = 1$. It turns out that all odd-order derivatives of $\mathcal{L}(\alpha)$ vanish when $\alpha_0 = 0$, but increase as α_0 moves away from zero.

- In the pathological situation $\alpha_0 = 1$ where the non-ideal behavior persists for all N_{ev}, 2^{nd} and higher order derivatives of $\mathcal{L}$ are all infinite. Infinite derivatives seem pathological, so it appears that they should somehow cause the pathology in statistical behavior.

Given the suggestive line of reasoning leading to the inference that $F(\alpha) = C \cdot \exp[-\mathcal{L}(\alpha)]$, and given the observations listed above, it may come as a surprise that, in fact, the derivatives of $\mathcal{L}$ do not explain the non-ideal properties of the α distribution. When $F(\alpha) = C \cdot \exp[-\mathcal{L}(\alpha)]$ is compared to the actual α distribution, the details do not match. For instance, the function $F(\alpha) = C \cdot \exp[-\mathcal{L}(\alpha)]$ does not explain the non-gaussian tails seen in the actual distribution. Moreover, the odd-order derivatives of $\mathcal{L}$ do induce bias, but in the wrong sense. In the next section, the correct explanation of the origin of the non-ideal behavior is discussed.

M.3.2 Correct Explanation

M.3.2.1 The Basic Concepts

As a first step in elucidating the origin of the non-ideal behavior, basic concepts relating to the log-likelihood function will be introduced. First, let us make the following assumptions about the component functions $f_0(x)$ and $f_1(x)$:

> Let x_{min} and x_{max} be the minimum and maximum possible value of the observable x.[1] Then, excluding these endpoints, the functions $f_0(x)$ and $f_1(x)$ cross over at exactly one point, x_c.

Figure M.6 illustrates this assumption in a number of typical situations. Given this assumption, the range of observable values $[x_{min}, x_{max}]$ can be divided into two regions, A_+ and A_-. The region A_+ represents the portion of $[x_{min}, x_{max}]$ where $f_1(x) > f_0(x)$, while A_- represents the portion where $f_1(x) < f_0(x)$. See figure M.6.

[1] For the observable E_ℓ, which is unbounded above, x_{max} is taken to be a cut-off value beyond which the probability density functions are negligibly small.

The log-likelihood function is a sum over $\log[f_\alpha(x)]$, where $f_\alpha(x) = \alpha f_1(x) + (1-\alpha)f_0(x)$. For a fixed observable value x, the function $f_\alpha(x)$ is a linear function of α:

$$f_\alpha(x) = f_0(x) + \alpha \cdot [f_1(x) - f_0(x)] \tag{M.8}$$

The logarithm of $f_\alpha(x)$ is defined only for $f_\alpha(x) > 0$. This implies the following:

Case 1: $x \in A_+$

$-\log f_\alpha$ is defined in the region $\alpha > \alpha_-$, where $\alpha_- = f_0(x)/[f_0(x) - f_1(x)]$. At α_-, $-\log f_\alpha$ diverges to $+\infty$, and as α increases from α_-, $-\log f_\alpha$ decreases to $-\infty$. No matter what shape f_0 and f_1 have, $\alpha_- \leq 0$. See figure M.7.

Case 2: $x \in A_-$

$-\log f_\alpha$ is defined in the region $\alpha < \alpha_+$, where $\alpha_+ = f_0(x)/[f_0(x) - f_1(x)]$. At α_+, $-\log f_\alpha$ diverges to $+\infty$, and as α decreases from α_+, $-\log f_\alpha$ decreases to $-\infty$. No matter what shape f_0 and f_1 have, $\alpha_+ \geq 1$. See figure M.7.

In any experiment, the set of measurements $\{x_i\}$ can be divided into two parts, $\{x_i\}_{A_+}$ and $\{x_i\}_{A_-}$, where the set $\{x_i\}_{A_+}$ has all measurements x_i belonging to region A_+, and similarly for $\{x_i\}_{A_-}$. The measurements $\{x_i\}_{A_+}$ contribute terms in the log-likelihood function that: (1) diverge to $+\infty$ in the region $\alpha \leq 0$; and (2) decrease logarithmically to $-\infty$ as $\alpha \to +\infty$. The measurements $\{x_i\}_{A_-}$, on the other hand, contribute terms that: (1) diverge to $+\infty$ in the region $\alpha \geq 1$; and (2) decrease logarithmically to $-\infty$ as $\alpha \to -\infty$. When these two classes of terms are combined, one obtains the total log-likelihood function with the following properties:

- If neither $\{x_i\}_{A_+}$ nor $\{x_i\}_{A_-}$ are empty, then $\mathcal{L}(\alpha)$ diverges to $+\infty$ at α_{low} and α_{upp}, where α_{low} is the largest value of $f_0(x)/[f_0(x) - f_1(x)]$ evaluated over $x \in \{x_i\}_{A_+}$, and α_{upp} is the smallest value of $f_0(x)/[f_0(x) - f_1(x)]$ evaluated over $x \in \{x_i\}_{A_-}$. It should be noted that, no matter what shape f_0 and f_1 have, $\alpha_{low} \leq 0$ and $\alpha_{upp} \geq 1$. $\mathcal{L}(\alpha)$ has exactly one minimum between α_{low} and α_{upp}. See figure M.8.

- If either $\{x_i\}_{A_+}$ or $\{x_i\}_{A_-}$ is an empty set, then the log-likelihood function has no minimum. See figure M.8. If p_+ is the probability that a given measurement from a

parent distribution $f_{\alpha_0}(x)$ is in region A_+ and p_- defined similarly for A_-, then the probability that an experiment with N_{ev} events has a log-likelihood function with no minimum is:

$$P_{no\ min} = p_+^{N_{ev}} + p_-^{N_{ev}} \tag{M.9}$$

This completes the survey of basic concepts regarding the log-likelihood function. Now, these concepts will be used to explain the origins of the non-ideal behavior in the α distribution.

M.3.2.2 Explaining Experiments with No Minimum

This has already been explained in the last section: experiments with no minimum occur when the measurements $\{x_i\}$ all belong to only one of the regions A_+ or A_-. The probability that an experiment with N_{ev} events has no minimum is given by $P_{no\ min} = p_+^{N_{ev}} + p_-^{N_{ev}}$. Thus, this non-ideal behavior decreases exponentially with N_{ev}. This behavior decreases quickest when the probability for x to be in A_+ is equal to that for A_- — *i.e.* $p_+ = p_- = 0.5$. The more unequal the probabilities, the slower the decrease in this non-ideal behavior.

M.3.2.3 Explaining the Non-gaussian Tails

Experiments with α populating the non-gaussian tails can be thought of as generalizations of experiments with no minimum. In experiments with no minimum, all events occupy only one region, A_+ or A_-. In experiments with α in the non-gaussian tails, most events occupy one of the regions, with few events occupying the complementary region. Furthermore, the few events occupying the complementary region are all close the the cross-over point x_c. In other words, these few events in the complementary region are almost, but not quite, in the same region as the majority of events.

As an illustration, let us say that most events occupy the region A_+. The few events occupying A_- are close to the x_c. Events belonging to A_- determine the upper cut-off α_{upp} of the log-likelihood function. The upper cut-off α_{upp} is defined as the smallest value of

$f_0(x)/[f_0(x) - f_1(x)]$ evaluated over events in A_-. Since most of the events are close to x_c, the denominator in $f_0(x)/[f_0(x) - f_1(x)]$ is close to zero for all events in A_-. Thus α_{upp} is very large. Because of this, the minimum of the log-likelihood function is also very large. This is why α in such experiments populate the non-gaussian tails.

Like the experiments with no minimum, experiments with α in the non-gaussian tails become less likely as N_{ev} increases.

M.3.2.4 Explaining Bias

The bias in the distribution is explained by the non-gaussian tails. Bias occurs when $p_+ \neq p_-$. When $p_+ \neq p_-$, the probability that a given experiment gives a non-gaussian tail in the region $\alpha < 0$ is not equal to that for $\alpha > 1$. For instance, if $p_+ > p_-$, then the probability of getting a non-gaussian tail in the region $\alpha > 1$ is greater than that for $\alpha < 0$. Thus the α distribution in this situation tends to be skewed towards large values of α, and the distribution has mean larger than the true value α_0. Since bias is a by-product of the non-gaussian tail, it decreases quickly as N_{ev} increases.

M.3.2.5 Explaining the Pathological Situations

The pathological situation where non-ideal properties persist no matter how large N_{ev} is made occurs when the true parameter value α_0 is equal to the largest possible value of α_{low} or the smallest possible value of α_{upp}. Since $\alpha_{low} \leq 0$ and $\alpha_{upp} \geq 1$, and since $0 \leq \alpha_0 \leq 1$, it is seen that pathological situations occur only when the following conditions are both true:

1. When $\alpha_0 = 0$ or 1 — *i.e.* at the boundary of its allowed value.

2. When the smallest possible value of $\alpha_{upp} = 1$ and/or the largest possible value of $\alpha_{low} = 0$. Whether either or both are possible depends on the shape of $f_0(x)$ and $f_1(x)$.

For example, in the analytic top polarization measurement, $f_0(x) = 1/2$ and $f_1(x) = (1/2) \cdot (1 + x)$. Then $f_0(x)/[f_0(x) - f_1(x)] = -1/x$. Since the range of x is $[-1, 1]$, the

largest possible value of $\alpha_{low} = -1$, and the smallest possible value of $\alpha_{upp} = 1$. Therefore, the pathology shows up only when $\alpha_0 = 1$.

In the case of the analytic W helicity measurement, $f_0(x) = 6 \cdot (x_{max} - x) \cdot (x - x_{min})/(x_{max} - x_{min})^3$, and $f_1(x) = 3 \cdot (x_{max} - x)^2/(x_{max} - x_{min})^3$. Thus $f_0(x)/[f_0(x) - f_1(x)] = 2 \cdot (x - x_{min})/[2 \cdot (x - x_{min}) - (x_{max} - x)]$. Since the range of x in this case is $[x_{min}, x_{max}]$, the largest possible value of $\alpha_{low} = 0$ and the smallest possible value of $\alpha_{upp} = 1$. Thus, in this case, the pathology exists at both $\alpha_0 = 0$ and 1.

As a final example, consider the case of the top polarization measurement at the output level — *i.e.* after $t\bar{t}$ event reconstruction. The component functions $f_0(x)$ and $f_1(x)$ are shown in figure M.6(c). In this case, the largest possible value of α_{low} and the smallest possible value of α_{upp} must be obtained numerically; they are:

$$\max[\alpha_{low}] = -2.00 \tag{M.10}$$

$$\min[\alpha_{upp}] = 1.90 \tag{M.11}$$

This shows that no pathologies exist in this case. This is typical of observables that are smeared by processes such as event reconstruction. In fact, the only way that pathologies can occur is if either or both component functions become zero at some values of x. Smeared observable distributions often do not have any zeros.

So far, the necessary conditions for the existence of the pathology has been shown. Now, let us examine why the pathology exists when the conditions are met. For the sake of concreteness, let us assume that $\alpha_0 = 1$ and $\min[\alpha_{upp}] = 1$. The following is a chain of reasoning explaining why the pathology exists:

1. Since $\alpha_0 = \min[\alpha_{upp}]$, in any given experiment, α_{upp} can be arbitrarily close to α_0. In contrast, since $\max[\alpha_{low}] \leq 0$, there is always a comfortable margin between α_0 and α_{low}.

2. In any given experiment, the log-likelihood function diverges to $+\infty$ at α_{low} and α_{upp}. The point α_{min} where the log-likelihood function is minimum lies somewhere between α_{low} and α_{upp}.

3. The closer α_{min} is to the boundary points α_{low} or α_{upp}, the more influence the boundary points have on the location of α_{min}.

4. The value of α_{upp} in a particular experiment is determined by a single measured value in $\{x_i\}_{A\_}$.

5. Since α_{upp} can get arbitrarily close to α_0 in any experiment, the location of the minimum in an experiment is heavily influenced by a single data point in region $A\_$. In other words, a single data point has an inordinate amount of influence in determining the location of α_{min}. This results in the partial loss of information available from all other events in the experiment. This is the source of the pathology.

6. Since $\alpha_0 = \min[\alpha_{upp}]$, α_{upp} can get arbitrarily close to α_0, no matter how large N_{ev} is made. Thus the pathology persists for all N_{ev}. This is in contrast to situations where some margin exists between α_0 and α_{upp} and α_{low}; in these situations, by increasing N_{ev} sufficiently, the probability that α_{min} is close to α_{upp} and α_{low} can be made arbitrarily small.

M.4 The Ideality of the Observables in the Experimental Data

In this section, the fact that the observables in the experimental data satisfy ideal statistical behavior is demonstrated. The pseudo-experiment method is used for the demonstration. Before going into the demonstration, let us review the following facts about the observables and the experimental data:

The Observables

The observables consist of $\cos U_i$ $(i = 1, 2, 3)$ and E_ℓ. For $\cos U_i$, the standard model predicts that the top quark is unpolarized. This means that the true parameter value is $\alpha_0 = 0$. For E_ℓ, the standard model predicts that the soft transverse helicity fraction r_T is 0.30. After acceptance correction, this is $r_T = 0.25$. This is taken as the true parameter value α_0 for E_ℓ.

The Polarization Observables

The polarization observables $\cos U_i$ are directional, so some care must be taken to deal with this fact. The '1' direction is forward-backward asymmetric, so that polarization that produces asymmetry *along* the '1' direction must be treated separately from that which produces asymmetry *against* this direction. The former case is denoted as $U1+$, the latter case as $U1-$. The '2' and '3' directions, on the other hand, are forward-backward symmetric, so that the statistical behavior for polarization in the forward direction is identical to that in the backward direction. In these directions, therefore, the forward direction alone is examined in the pseudo-experiments. These observables are referred to as $U2$ and $U3$.

The Data Sample

The data sample is divided into eight parts. First, the events are categorized as 3.5-jet (NJ3.5) and 4-jet (NJ4) events. Second, both categories are divided according to the type of b-tags: SVX only (XO), SLT only (TO), SVX and SLT (XT), and no tag (NT). Each subsample has its own component functions $f_0(x)$ and $f_1(x)$. Also, each subsample has different background fraction β and background observable distribution $f_b(x)$. The total log-likelihood function is taken as the sum over the log-likelihood function in each subsample. See section 8.3 for more details.

Analysis Optimization

The analysis is optimized by: (1) applying an H_T cut to certain subsamples to improve $S^2/(S+B)$; and (2) applying a top-mass constraint to the $t\bar{t}$ reconstruction process in order to reduce smearing in the observable distributions. The first optimization will be referred to as HT CUT, and the second one will be referred to as MTCON. Before HT CUT, the total number of events is 159; afterwards, it is 114. See section 8.3.3 for details.

Figure M.9 shows the distribution of α_{min} from 1000 pseudo-experiments for each of the observables in the unoptimized analysis — *i.e.* no H_T cut and no top mass constraint. Figures M.10 through M.12 are the corresponding plots for (HT CUT, MTCON) = (YES, NO), (NO, YES), and (YES, YES), respectively. It is seen that, in each case, a χ^2 fit to a gaussian

OBSERVABLE	(HT CUT, MTCON)			
	NO, NO	YES, NO	NO, YES	YES, YES
$U1+$	-2.36	0.94	0.05	-1.45
$U1-$	-2.41	0.83	1.28	1.55
$U2$	-0.10	0.18	0.60	1.00
$U3$	-0.84	-1.10	-0.69	0.30
E_ℓ	0.46	-0.96	0.51	-1.39

Table M.1: The bias in the α_{min} distribution for each observable, in the analysis configurations (HT CUT, MTCON) = (NO, NO), (YES, NO), (NO, YES), and (YES, YES). Each entry is obtained from the α_{min} distribution obtained from a pseudo-experiment with $N_{exp} = 1000$.

gives a good result. Table M.1 shows the bias for each observable in four different analysis configurations, where the bias b is defined as:

$$b = \frac{mean - \alpha_0}{\sigma_{mean}} \tag{M.12}$$

$$\sigma_{mean} = \frac{rms}{\sqrt{N_{exp}}} \tag{M.13}$$

There is no indication of bias beyond that explainable by statistical fluctuation.

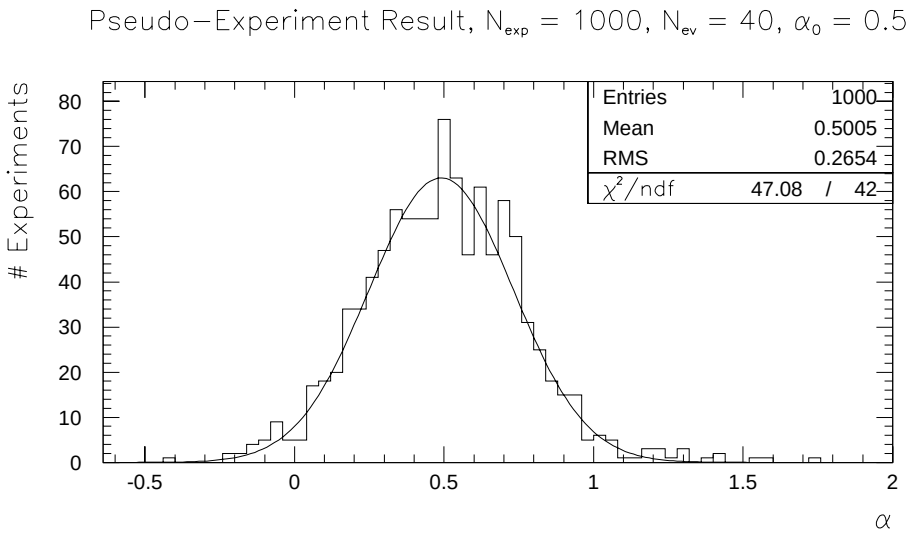


Figure M.4: The distribution of α for $N_{exp} = 1000$, $N_{ev} = 40$, $\alpha_0 = 0.5$.

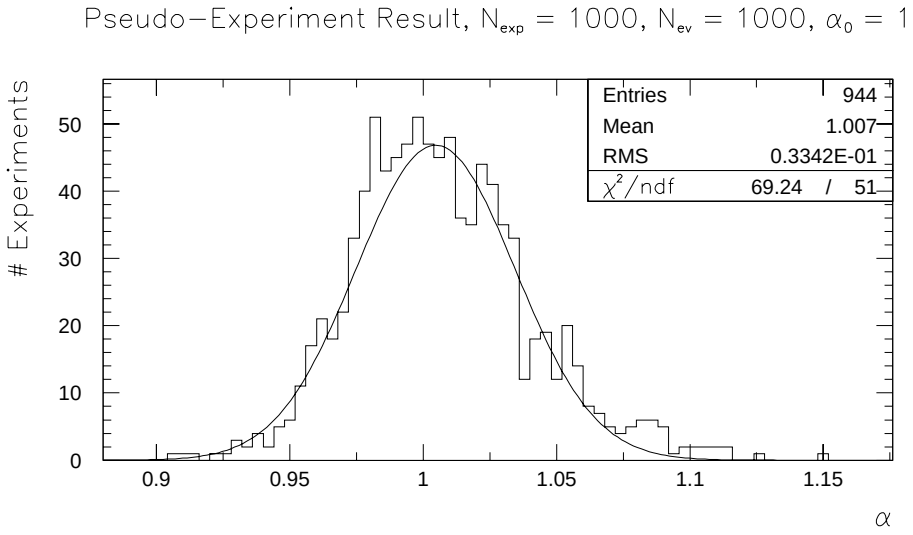


Figure M.5: The distribution of α for $N_{exp} = 1000$, $N_{ev} = 1000$, $\alpha_0 = 1$.

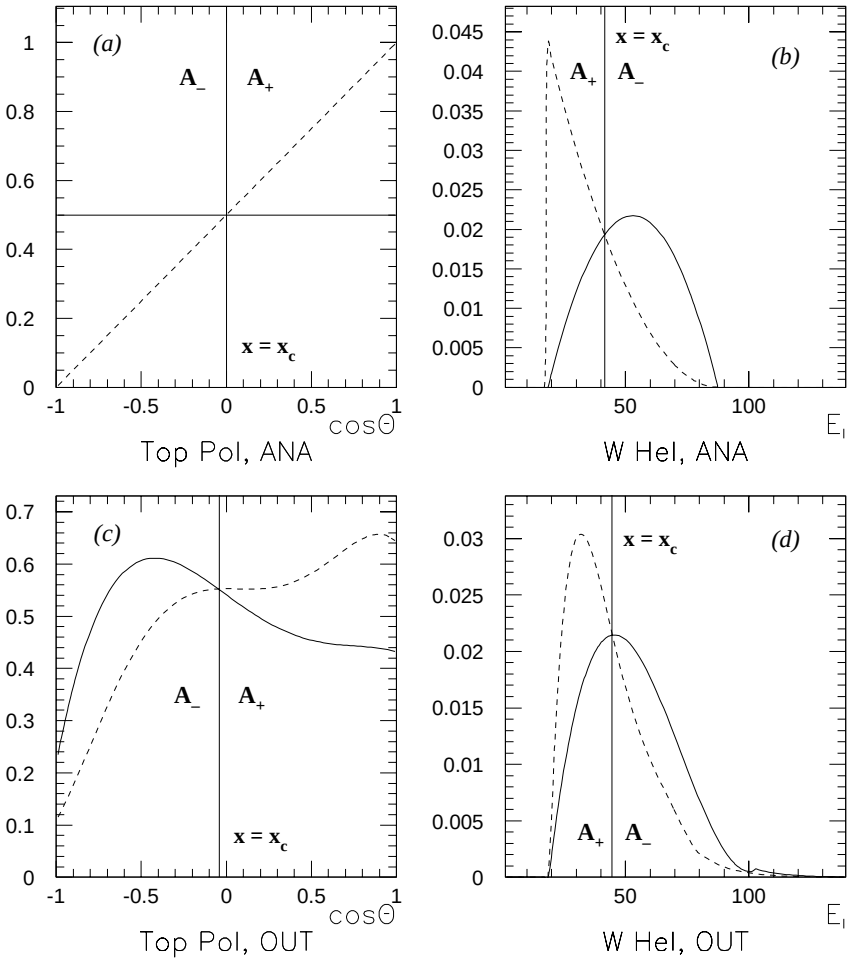


Figure M.6: The component functions $f_0(x)$ and $f_1(x)$ for: (a) top polarization measurement, analytic level; (b) W helicity measurement, analytic level; (c) top polarization measurement, output level (*i.e.* after $t\bar{t}$ event reconstruction); (d) W helicity measurement, output level. The solid curve in each figure is $f_0(x)$, and the dashed curve is $f_1(x)$. The vertical line indicates the cross-over point, x_c. The region where $f_1(x) > f_0(x)$ is labeled A_+, and the region where $f_1(x) < f_0(x)$ is labeled A_-.

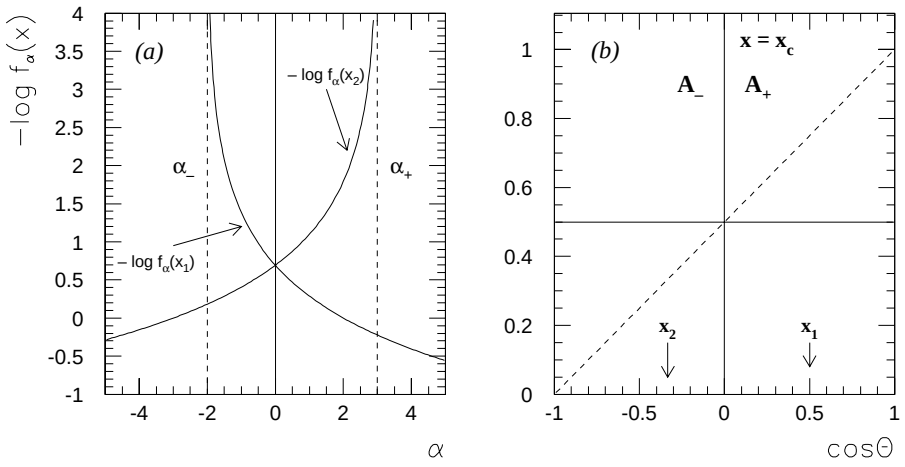


Figure M.7: (a) Examples of typical terms in the log-likelihood function. The measurement is taken as top polarization at the analytic level. The curves are $-\log f_\alpha(x)$. The vertical dashed lines are the values of α at which the log-likelihood functions diverge. The divergent values of α are $\alpha_- = -2.0$ and $\alpha_+ = 3.0$. The two log-likelihood functions are for $x_1 = 1/2$ and $x_2 = -1/3$. (b) The location of x_1 and x_2 in the observable distribution.

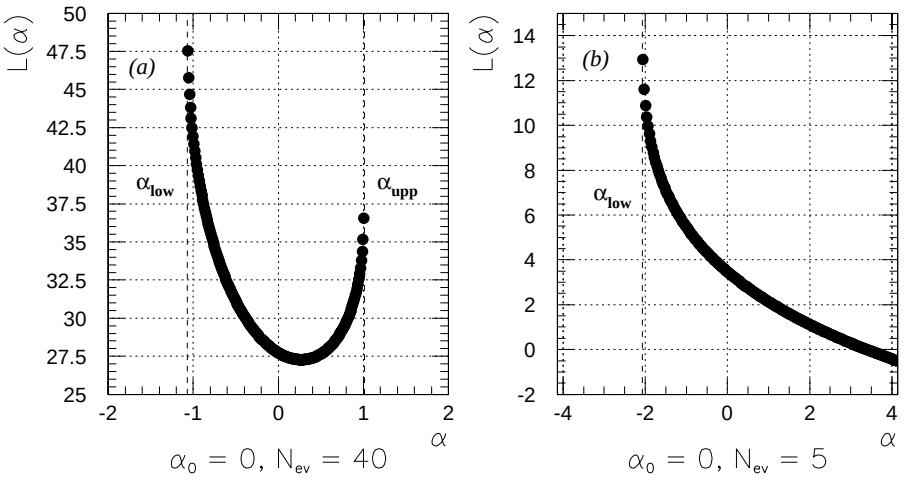


Figure M.8: (a) Log-likelihood function from a typical experiment with $N_{ev} = 40$. The function diverges to $+\infty$ at α_{low} and α_{upp}, and has exactly one minimum between these points. (b) An example of a log-likelihood function without a minimum. In this experiment, $N_{ev} = 5$, and all five points lie in the region A_+. In both (a) and (b), the parent distribution is the from the unpolarized top quark at the analytic level.

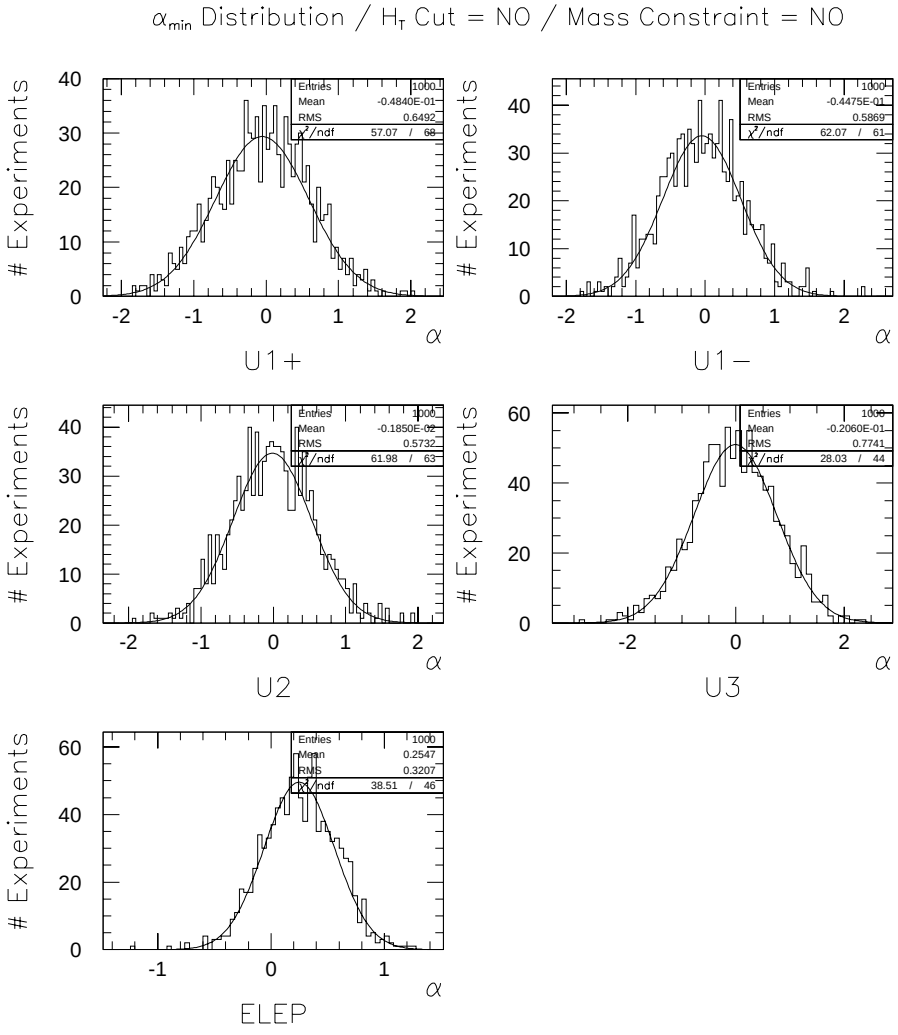


Figure M.9: The distribution of α_{min} from 1000 pseudo-experiments for the observables $U1+$, $U1-$, $U2$, $U3$, and E_ℓ. The analysis is unoptimized — *i.e.* no H_T cut and no top mass constraint.

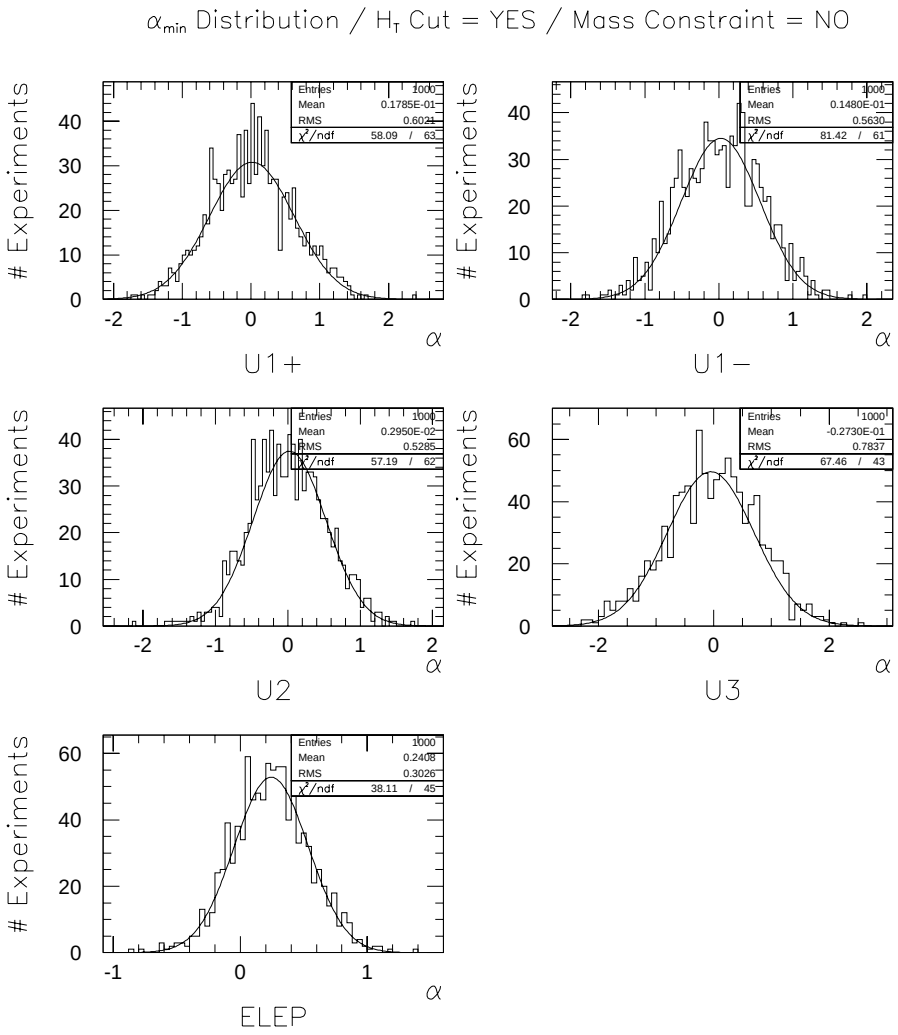


Figure M.10: The distribution of α_{min} from 1000 pseudo-experiments for the observables $U1+$, $U1-$, $U2$, $U3$, and E_ℓ. H_T cut is applied, but no top mass constraint is applied.

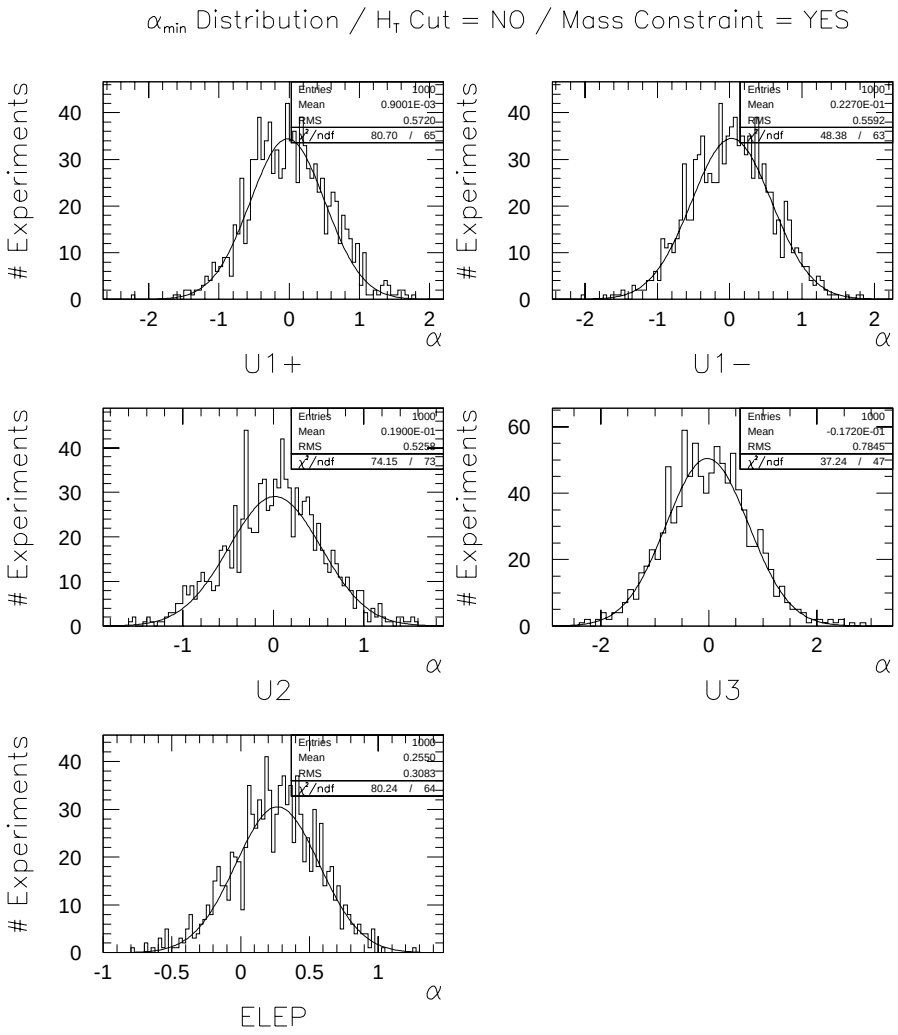


Figure M.11: The distribution of α_{min} from 1000 pseudo-experiments for the observables $U1+$, $U1-$, $U2$, $U3$, and E_ℓ. H_T cut is not applied, but top mass constraint is applied.

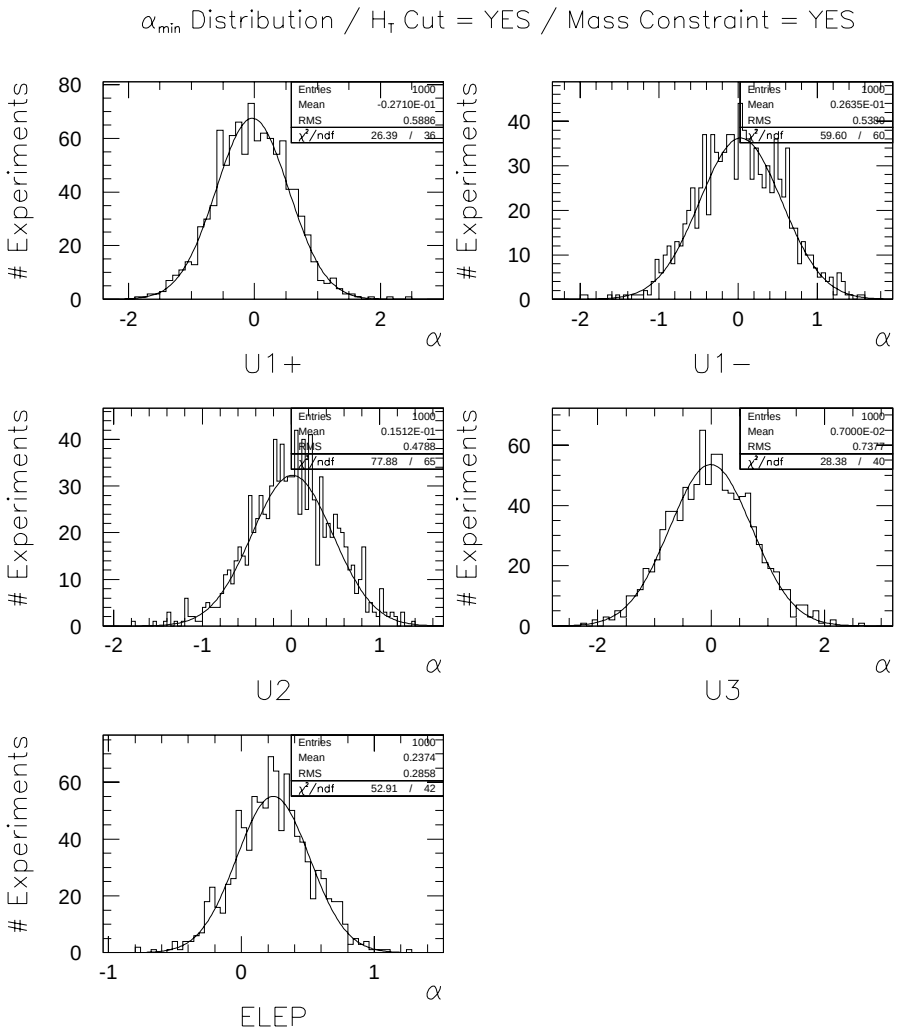


Figure M.12: The distribution of α_{min} from 1000 pseudo-experiments for the observables $U1+$, $U1-$, $U2$, $U3$, and E_ℓ. Both H_T cut and top mass constraint are applied.

Appendix N

Generalization of the W Helicity Measurement

N.1 Introduction

The W helicity measurement described in chapter 8 fixes the hard transverse helicity fraction r'_T to zero, and uses the soft helicity fraction r_T as a fit parameter to compare the E_ℓ distribution in the experimental data with that from monte carlo. This restriction on r'_T allows one to obtain an experimental measurement of r_T with considerably smaller statistical error than is obtainable when both r_T and r'_T are used as fit parameters. In this appendix, the W helicity measurement is generalized so that both r_T and r'_T are used as fit parameters. The objectives of this appendix are:

- Examine the log-likelihood function in a 3-component, 2-parameter fit and elucidate some of its statistical properties.
- Obtain an estimate of the statistical error on r_T, r'_T, and the longitudinal helicity fraction $r_L = 1 - r_T - r'_T$.

N.2 The Log-likelihood Function

N.2.1 The Definition of the Log-likelihood Function

Since the full definition of the log-likelihood function is complicated, it will be defined in three steps, as was done in section 8.2 in the discussion of 2-component fits. In the first step, let us assume that: (1) no background events contribute to the data sample; and (2)

the data sample is not divided into subsamples. Then the log-likelihood function is given by:

$$\mathcal{L}(\alpha,\beta) = \sum_{i=1}^{N_{ev}} -\log\left[\alpha\cdot f_1(x_i) + \beta\cdot \tilde{f}_1(x_i) + (1-\alpha-\beta)\cdot f_0(x_i)\right] \tag{N.1}$$

The fit parameters α and β stand for r_T and r'_T, respectively, the component functions f_1, $\tilde{f}_1$, and f_0 stand for g_s, g_h, and g_m, respectively, and x stands for E_ℓ. The sum is taken over the events in the data, where N_{ev} is the number of events.

For the second step, let us relax the assumption about the background by allowing a non-zero background fraction B. The log-likelihood function in this case is:

$$\mathcal{L}(\alpha,\beta) = \sum_{i=1}^{N_{ev}} -\log\left[B\cdot f_b(x_i) + (1-B)\cdot\left\{\alpha\cdot f_1(x_i) + \beta\cdot \tilde{f}_1(x_i) + (1-\alpha-\beta)\cdot f_0(x_i)\right\}\right] \tag{N.2}$$

The function $f_b(x)$ is the distribution of x in background processes.

For the final step, let us relax the assumption about subdividing the data. Instead of one, monolithic data sample, the data are subdivided according to features such as the number of tight jets and the number and type of b-tags. Then the generalization of equation N.2 to this situation is the following:

$$\mathcal{L}(\alpha,\beta) = \sum_{\lambda} \mathcal{L}_\lambda(\alpha,\beta) \tag{N.3}$$

$$\mathcal{L}_\lambda(\alpha,\beta) = \sum_{i=1}^{N_\lambda} -\log\left[B_\lambda\cdot f_b^\lambda(x_i) + (1-B_\lambda)\cdot\left\{\alpha\cdot f_1^\lambda(x_i) + \beta\cdot \tilde{f}_1^\lambda(x_i) + (1-\alpha-\beta)\cdot f_0^\lambda(x_i)\right\}\right] \tag{N.4}$$

Each subsample of the data is represented by the index λ. The λ in subscripts and superscripts indicate a given quantity or function for the subsample λ. In this thesis, the data are subdivided into eight categories labeled (C_1, C_2), where C_1 = NJ3.5, NJ4 and C_2 = XO, TO, XT, and NT. See section 8.2.4 for details.

N.2.2 The Statistical Properties of the Log-likelihood Function

The statistical properties of $\mathcal{L}(\alpha, \beta)$ can be described by the pseudo-experiment method described in section 8.2.2. Let us imagine performing N_{exp} pseudo experiments with N_{ev} events in each experiment. Let (α_i, β_i) be the parameter values that minimize the log-likelihood function in experiment i. Then, in ideal situations, the distribution of α_i and β_i have the following properties:

- α_i is distributed as a gaussian centered at the true parameter value α_0. β_i is a gaussian centered at the true parameter value β_0.
- The width of the gaussian for α_i and for β_i are determined by: (1) N_{ev}; and (2) the second derivatives of $\mathcal{L}$. This is discussed further in the next section.
- The longitudinal helicity fraction $r_L = 1 - \alpha - \beta$ is distributed as a gaussian centered at the true parameter value $1 - \alpha_0 - \beta_0$.
- The distributions α_i and β_i are correlated. Thus the error on r_L is not given simply by the quadrature of the width of α_i and β_i, but by a formula that takes account of the correlation. This is discussed in detail in the next section. The correlated width is $>$ the uncorrelated width.

Just as in the 1-parameter case discussed in section 8.2, these ideal properties are valid only when: (1) N_{ev} is sufficiently large; and (2) no pathologies exist. For this appendix, it will simply be assumed that the ideal properties are valid.

N.2.3 The Statistical Error of the Parameters

The statistical error of the parameters α and β, and the derived quantity $r_L = 1 - \alpha - \beta$, can be estimated by analyzing the error contour of $\mathcal{L}(\alpha, \beta)$. The error contour is simply the Taylor expansion of $\mathcal{L}(\alpha, \beta) - \mathcal{L}(\alpha_0, \beta_0)$ up to the quadratic term, and is given as follows:

$$T(\alpha, \beta) \equiv \mathcal{L}(\alpha, \beta) - \mathcal{L}(\alpha_0, \beta_0)$$

$$= \frac{1}{2}A(\alpha-\alpha_0)^2 + \frac{1}{2}B(\beta-\beta_0)^2 + C(\alpha-\alpha_0)(\beta-\beta_0) \tag{N.5}$$

$$A \equiv \left.\frac{\partial^2\mathcal{L}}{\partial\alpha^2}\right|_{\alpha=\alpha_0,\beta=\beta_0} \tag{N.6}$$

$$B \equiv \left.\frac{\partial^2\mathcal{L}}{\partial\beta^2}\right|_{\alpha=\alpha_0,\beta=\beta_0} \tag{N.7}$$

$$C \equiv \left.\frac{\partial^2\mathcal{L}}{\partial\alpha\partial\beta}\right|_{\alpha=\alpha_0,\beta=\beta_0} \tag{N.8}$$

The width σ_α of the distribution of α alone is obtained by the following rule:

1. Draw the contour $T(\alpha,\beta) = 1/2$. This is a tilted ellipse in the α-β plain, and corresponds to the set of all points where $\mathcal{L}$ is larger than the minimum value by $1/2$. See figure N.1(a).

2. The distance from α_0 to the largest possible value of α on this ellipse is equal to the corresponding distance to the smallest possible value of α on the ellipse. This distance is equal to σ_α. This is shown graphically in figure N.1(b).

The width σ_β for the distribution of β alone is obtained in a similar manner (figure N.1(c)). For the distribution of r_L, one evaluates $r_L = 1-\alpha-\beta$ over the ellipse. There is one point r_L^{min} at which r_L is minimum, and one point r_L^{max} at which r_L is maximum. The width of the r_L distribution is given by $\sigma_L = r_L^0 - r_L^{min} = r_L^{max} - r_L^0$. See figure N.1(d).

The widths σ_α, σ_β, and σ_L can be expressed in terms of the second derivatives of $\mathcal{L}$ evaluated at $(\alpha,\beta) = (\alpha_0,\beta_0)$. They are:

$$\sigma_\alpha = \sqrt{\frac{B}{Q}} \tag{N.9}$$

$$\sigma_\beta = \sqrt{\frac{A}{Q}} \tag{N.10}$$

$$\sigma_L = \sqrt{\frac{A+B-2C}{Q}} \tag{N.11}$$

$$Q = A\cdot B - C^2 \tag{N.12}$$

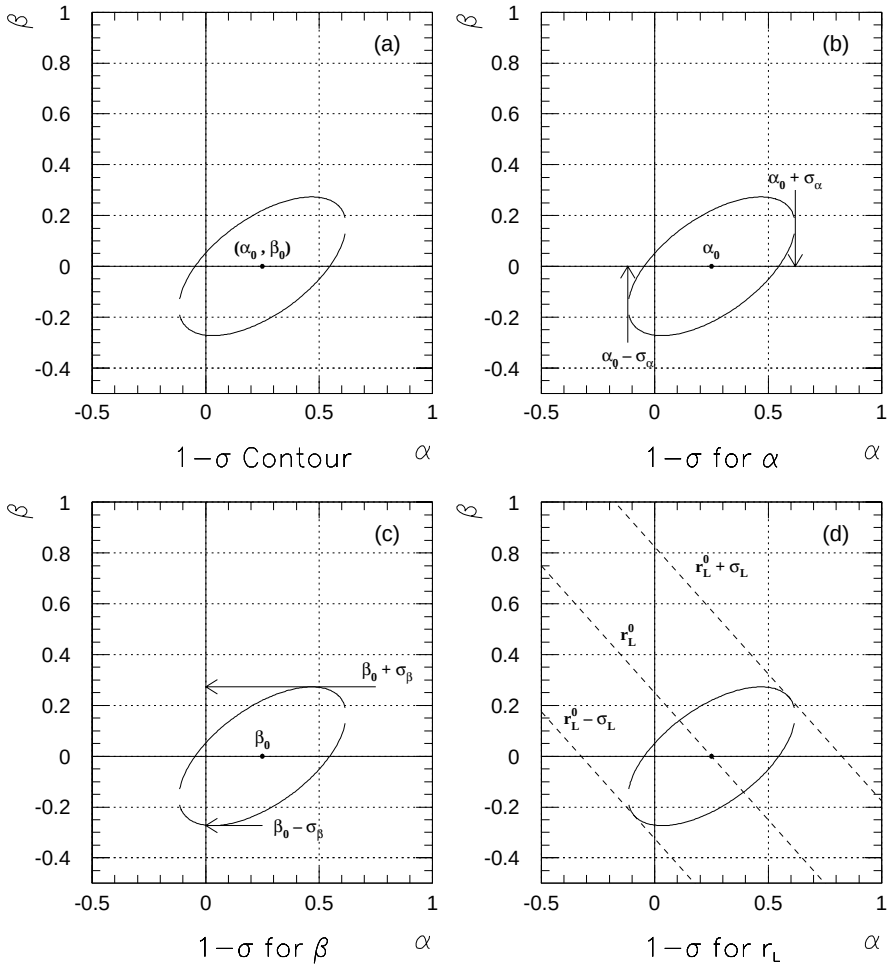


Figure N.1: (a) 1-σ contour for the 2-parameter log-likelihood function for the observable E_ℓ. This is obtained from the experimental data, before any optimization is performed. There are 159 events in data, with expected signal contribution of 55 events. (b) Obtaining the 1-σ value for the α distribution. (c) Obtaining the 1-σ value for the β distribution. (d) Obtaining the 1-σ value for the r_L distribution. The dashed lines are lines of constant r_L.

N.2.4 Continuum Approximation of the Statistical Errors

The continuum approximation of the log-likelihood function for the case of 1-parameter fits is discussed in section 8.2. The generalization to 2-parameter fits is the following:

$$\mathcal{L}_\lambda(\alpha,\beta) = -N_{ev} r_\lambda \int dx \, \log\Big[\quad B_\lambda \cdot f_b^\lambda(x) + (1-B_\lambda)\cdot\Big\{\alpha\cdot f_1^\lambda(x) + \beta\cdot \tilde{f}_1^\lambda(x) + (1-\alpha-\beta)\cdot f_0^\lambda(x)\Big\}\Big] \tag{N.13}$$

This expression is for the subsample λ; the total log-likelihood function is obtained by summing over all the subsamples.

In terms of the continuum approximation of $\mathcal{L}_\lambda$, the second derivatives for the subsample λ are:

$$\begin{aligned} A_\lambda &\equiv \left.\frac{\partial^2 \mathcal{L}_\lambda}{\partial\alpha^2}\right|_{\alpha=\alpha_0,\beta=\beta_0} \\ &= N_{ev} r_\lambda (1-B_\lambda)^2 \int dx \, \frac{[f_1(x)-f_0(x)]^2}{f_{exp}(x)} \end{aligned} \tag{N.14}$$

$$\begin{aligned} B_\lambda &\equiv \left.\frac{\partial^2 \mathcal{L}_\lambda}{\partial\beta^2}\right|_{\alpha=\alpha_0,\beta=\beta_0} \\ &= N_{ev} r_\lambda (1-B_\lambda)^2 \int dx \, \frac{\left[\tilde{f}_1(x)-f_0(x)\right]^2}{f_{exp}(x)} \end{aligned} \tag{N.15}$$

$$\begin{aligned} C_\lambda &\equiv \left.\frac{\partial^2 \mathcal{L}_\lambda}{\partial\alpha\partial\beta}\right|_{\alpha=\alpha_0,\beta=\beta_0} \\ &= N_{ev} r_\lambda (1-B_\lambda)^2 \int dx \, \frac{[f_1(x)-f_0(x)]\left[\tilde{f}_1(x)-f_0(x)\right]}{f_{exp}(x)} \end{aligned} \tag{N.16}$$

$$f_{exp}(x) = B_\lambda \cdot f_b(x) + (1-B_\lambda)\left[\alpha_0\cdot f_1(x) + \beta_0\cdot \tilde{f}_1(x) + (1-\alpha_0-\beta_0)\cdot f_0(x)\right] \tag{N.17}$$

Let us denote the integrals appearing in the above equations as follows:

$$(IA)_\lambda = \int dx \, \frac{[f_1(x) - f_0(x)]^2}{f_{exp}(x)} \tag{N.18}$$

$$(IB)_\lambda = \int dx \, \frac{\left[\tilde{f}_1(x) - f_0(x)\right]^2}{f_{exp}(x)} \tag{N.19}$$

$$(IC)_\lambda = \int dx \, \frac{[f_1(x) - f_0(x)]\left[\tilde{f}_1(x) - f_0(x)\right]}{f_{exp}(x)} \tag{N.20}$$

Then A_λ, B_λ, and C_λ can be written as follows:

$$A_\lambda = N_{ev} r_\lambda (1 - B_\lambda)^2 (IA)_\lambda \tag{N.21}$$

$$B_\lambda = N_{ev} r_\lambda (1 - B_\lambda)^2 (IB)_\lambda \tag{N.22}$$

$$C_\lambda = N_{ev} r_\lambda (1 - B_\lambda)^2 (IC)_\lambda \tag{N.23}$$

Let a_λ, b_λ, and c_λ be the λ-dependent part of A_λ, B_λ, and C_λ. Then the above expressions can be written as:

$$A_\lambda = N_{ev} a_\lambda \tag{N.24}$$

$$B_\lambda = N_{ev} b_\lambda \tag{N.25}$$

$$C_\lambda = N_{ev} c_\lambda \tag{N.26}$$

Let us denote the sum of A_λ, B_λ, and C_λ over λ as A, B, and C, respectively. Similarly, let us denote the sum of a_λ, b_λ, and c_λ as a, b, and c, respectively. Then the continuum approximation of the statistical errors σ_α, σ_β, and σ_L is given as follows:

$$\sigma_\alpha = \sqrt{\frac{B}{Q}} = \sqrt{\frac{b}{q}} \cdot \frac{1}{\sqrt{N_{ev}}}$$

$$\sigma_\beta = \sqrt{\frac{A}{Q}} = \sqrt{\frac{a}{q}} \cdot \frac{1}{\sqrt{N_{ev}}} \tag{N.27}$$

$$\sigma_L = \sqrt{\frac{A+B-2C}{Q}} = \sqrt{\frac{a+b-2c}{q}} \cdot \frac{1}{\sqrt{N_{ev}}}$$

$$Q = A \cdot B - C^2 \tag{N.28}$$

$$q = a \cdot b - c^2 \tag{N.29}$$

The errors, as expressed in equation N.27, underestimate the true errors, but not by much. The advantage of using the continuum approximation is that the errors can be factored into a statistical part ($1/\sqrt{N_{ev}}$) and non-statistical part (*e.g.* $\sqrt{b/q}$ for σ_α). The non-statistical part depends on the background fractions B_λ and the shape of the functions $f_0(x)$, $f_1(x)$, $\tilde{f}_1(x)$, and $f_b(x)$.

N.3 Estimating the Statistical Error of the Parameters in the Experimental Data

In this section, the continuum approximation of the log-likelihood function is used to obtain an estimate of the statistical error of the soft and hard transverse helicity fraction (α and β), and the longitudinal helicity fraction ($r_L = 1 - \alpha - \beta$). Following that is a discussion of the results.

N.3.1 Estimated Statistical Error

Table N.1 shows the numbers needed to obtain σ_α, σ_β, and σ_L. From those numbers, one obtains the following for the statistical errors:

$$\sigma_\alpha = \frac{4.65}{\sqrt{N_{ev}}} = 0.37 \tag{N.30}$$

λ	r_λ	B_λ	$(IA)_\lambda$	$(IB)_\lambda$	$(IC)_\lambda$	a_λ	b_λ	c_λ
(NJ3.5, XO)	0.0252	0.29	0.326	0.467	−0.255	4.14	5.93	−3.24
(NJ3.5, TO)	0.0314	0.66	0.452	0.824	−0.306	1.64	2.99	−1.11
(NJ3.5, XT)	0.0252	0.18	0.312	1.020	−0.263	5.29	17.28	−4.46
(NJ3.5, NT)	0.4843	0.90	0.343	0.599	−0.302	1.66	2.90	−1.46
(NJ4, XO)	0.0692	0.07	0.267	0.726	−0.264	15.98	43.45	−15.80
(NJ4, TO)	0.0566	0.25	0.422	0.462	−0.283	13.44	14.71	−9.01
(NJ4, XT)	0.0252	0.04	0.330	0.843	−0.312	7.66	19.58	−7.25
(NJ4, NT)	0.2830	0.58	0.431	0.464	−0.295	21.52	23.16	−14.73
					TOTAL:	71.33	130.00	−57.06

Table N.1: The ingredients for calculating σ_α, σ_β, and σ_L. The numbers under a_λ, b_λ, and c_λ are in units of 10^{-3}.

$$\sigma_\beta = \frac{3.44}{\sqrt{N_{ev}}} = 0.27 \tag{N.31}$$

$$\sigma_L = \frac{7.24}{\sqrt{N_{ev}}} = 0.57 \tag{N.32}$$

The total number of events N_{ev} is 159. To put these numbers in perspective, let us assume that a measurement is performed, and the result is the true parameter values $(\alpha_0, \beta_0, r_L^0) = (0.25, 0.00, 0.75)$. Then, the result of the measurement can be expressed as follows:

$$\begin{aligned} \alpha &= 0.25 \pm 0.37 \\ \beta &= 0.00 \pm 0.27 \\ r_L &= 0.75 \pm 0.57 \end{aligned} \tag{N.33}$$

N.3.2 Discussion of the Results

The following are some observations on the results obtained in the last section:

Comparison with 1-parameter Fit

In the 1-parameter fit described in chapter 8, the statistical error for the soft transverse helicity fraction is equal to that for the longitudinal fraction: $\sigma_{1p} = 0.31$. Thus, in going from a 1-parameter fit to a 2-parameter fit, the statistical error of the soft transverse helicity fraction increases by 19%, while that for the longitudinal fraction increases by 84%.

Why $\sigma_L \approx 2 \times \sigma_\alpha$ and σ_β

The results of the last section shows that $\sigma_L/\sigma_\alpha = 1.54$ and $\sigma_L/\sigma_\beta = 2.11$ — in other words, longitudinal error is roughly twice as large as the transverse errors. This is due to the fact that, throughout most of the range of x, the longitudinal distribution $f_0(x)$ lies in between the transverse distributions $f_1(x)$ and $\tilde{f}_1(x)$. See figure N.2. In order to see how this explains $\sigma_L \approx 2 \times \sigma_\alpha$ and σ_β, consider the function $f(x) = \alpha \cdot f_1(x) + \beta \cdot \tilde{f}_1(x) + \gamma \cdot f_0(x)$, where $\gamma = r_L$, and the parameters α, β, and γ are constrained by $\alpha + \beta + \gamma = 1$. Suppose the parameters are changed by $\delta\alpha$, $\delta\beta$, and $\delta\gamma$, consistent with constraint on the parameters. The constraint imposes the condition $\delta\alpha + \delta\beta + \delta\gamma = 0$. The change in $f(x)$ is then $\delta f(x) = \delta\alpha \cdot f_1(x) + \delta\beta \cdot \tilde{f}_1(x) + \delta\gamma \cdot f_0(x)$. Suppose, for argument's sake, that, for all x, $f_0(x)$ lies approximately half way between $f_1(x)$ and $\tilde{f}_1(x)$. Then, if $\delta\alpha$ and $\delta\beta$ are both increased or decreased by the same amount Δ, and $\delta\gamma$ is changed in the opposite sense by $2 \times \Delta$, the net change in $f(x)$ is approximately zero. In other words, the log-likelihood function does not change much when the correlated change in parameter described above is made. In particular, near the minimum of the log-likelihood function, the error contour has the smallest curvature along the line $(\alpha - \alpha_0) = (\beta - \beta_0)$, and largest curvature along the line perpendicular to it — *i.e.* the ellipse has major axis along the line $(\alpha - \alpha_0) = (\beta - \beta_0)$, and minor axis along the line perpendicular to it. In the limit minor axis $\ll$ major axis, $\sigma_\alpha = \sigma_\beta$, and $\sigma_L = 2 \times \sigma_\alpha$ and σ_β. The fact that $\sigma_L = 2 \times \sigma_\alpha$ and σ_β is a result of the major axis of the ellipse coinciding with the line along which γ changes most rapidly ($\gamma = 1 - \alpha - \beta$). In reality, $f_0(x)$ is only roughly half way between the two other functions, and there is even a small region in x where $f_0(x)$ is not sandwiched by the other functions. Because of this, and other complicating factors, the scale factors relating the longitudinal error to the transverse errors is only roughly 2.

Data Optimization

The results were obtained using the unoptimized analysis — *i.e.* no H_T cut and no top mass constraint. With full optimization, the estimated errors should improve moderately, just as was the case in the 1-parameter fit discussed in chapter 8.

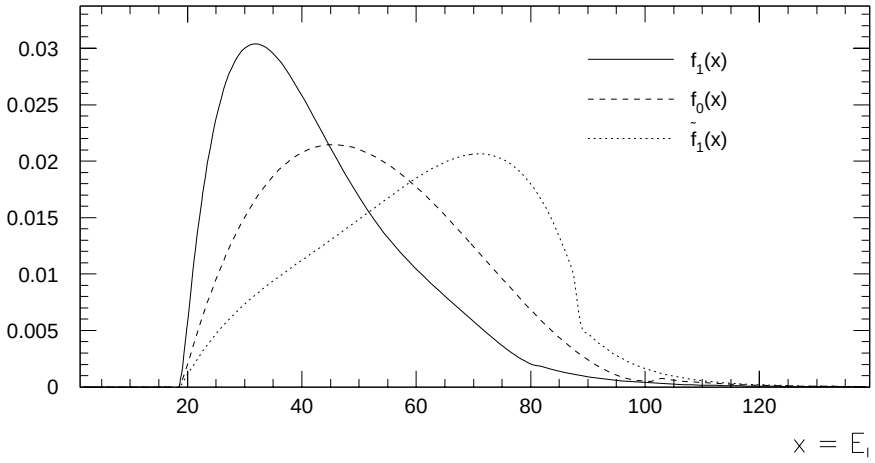


Figure N.2: The longitudinal ($f_0(x)$) and transverse ($f_1(x)$ and $\tilde{f}_1(x)$) E_ℓ distributions. $f_0(x)$ is sandwiched between the two other functions throughout most of the range of x.

Non-ideal Behaviors

The types of non-ideal statistical behavior in the 2-parameter log-likelihood method is the same as in the 1-parameter case — bias, non-gaussian tail, likelihood functions without minimum, and pathologies. These non-idealities in 2-parameters is usually considerably worse than in 1-parameter. For example, in both 1-parameter and 2-parameter fits, non-ideal behaviors generally decrease with increasing statistics; in 2-parameter fits, however, the threshold between non-ideal and ideal behavior is generally larger than in 1-parameter fits. As another example, the bias in 2-parameter fits tends to be considerably larger than in 1-parameter fits. A detailed examination of the non-ideal behaviors in 2-parameter fits is beyond the scope of this appendix. In the present context, it will simply be noted that the results may be somewhat inaccurate because of possible non-idealities.

Appendix O

The Dependence of the K-factor on N_{ev}

O.1 Introduction

The continuum approximation of the log-likelihood function is used throughout this thesis in order to express the statistical error in modular form — *i.e.* σ is factorizable into a statistical term $1/\sqrt{N_{ev}}$, a background term $1/(1-\beta)^2$, and a geometric term K:

$$\sigma = K \cdot \frac{1}{(1-\beta)^2} \cdot \frac{1}{\sqrt{N_{ev}}} \tag{O.1}$$

The geometric term is given by the following integral:

$$\frac{1}{K^2} = \int dx \, \frac{[f_1(x) - f_0(x)]^2}{f_{exp}(x)} \tag{O.2}$$

The functions $f_1(x)$ and $f_0(x)$ are the component functions, while $f_{exp}(x)$ is the theoretically expected distribution of the observable x. According to equation O.2, K is independent of N_{ev}. In reality, however, K does depend somewhat on N_{ev}: K starts out large, and decreases monotonically to an asymptotic value given by the integral in equation O.2. The objective of this appendix is to show the degree to which K depends on N_{ev}, and establish that the integral approximation of K is close to the true value for $N_{ev} = 159$, which is the number of events in the experimental data.

O.2 The K-factor as a Function of N_{ev}

Before going into the dependence of K on N_{ev}, a discussion on some technical matters is in order. The expression for the statistical error shown in equation O.1 is valid only when the data sample is treated as a single unit. It is shown in section 8.2.4 that the measurement resolution can be improved by dividing the data into subsamples. When the data are subdivided in this manner, the continuum approximation of σ must be generalized as follows:

$$\frac{1}{\sigma^2} = \sum_{\lambda=1}^{8} \frac{1}{\sigma_\lambda^2} \tag{O.3}$$

$$\frac{1}{\sigma_\lambda^2} = \frac{(1-\beta_\lambda)^2 \, r_\lambda \cdot N_{ev}}{K_\lambda^2} \tag{O.4}$$

The summation is over the eight subsamples. The quantities β_λ and K_λ are the background fraction and K-factor for subsample λ, while r_λ is the fraction of events belonging to subsample λ. Equation O.3 makes it clear that the total error σ cannot be factored into a statistical part, background part, and geometric part, as is possible when the data are not subdivided. However, σ can still be factored into a statistical part and a non-statistical part, as follows:

$$\sigma = \frac{\tilde{K}}{\sqrt{N_{ev}}} \tag{O.5}$$

$$\frac{1}{\tilde{K}^2} = \sum_{\lambda=1}^{8} \frac{(1-\beta_\lambda)^2 \, r_\lambda \cdot}{K_\lambda^2} \tag{O.6}$$

Henceforth, the term "K-factor" will refer to $\tilde{K}$ given by this equation.

Figure O.1 shows the K-factor as a function of N_{ev} for all of the observables. The true value of K (the points in the figure) is estimated using the pseudo-experiment technique described in page 147. These plots confirm the form of the N_{ev} dependence of K described earlier. Also, given that $N_{ev} = 159$, it is clear that the continuum value of K approximates the true value very well.

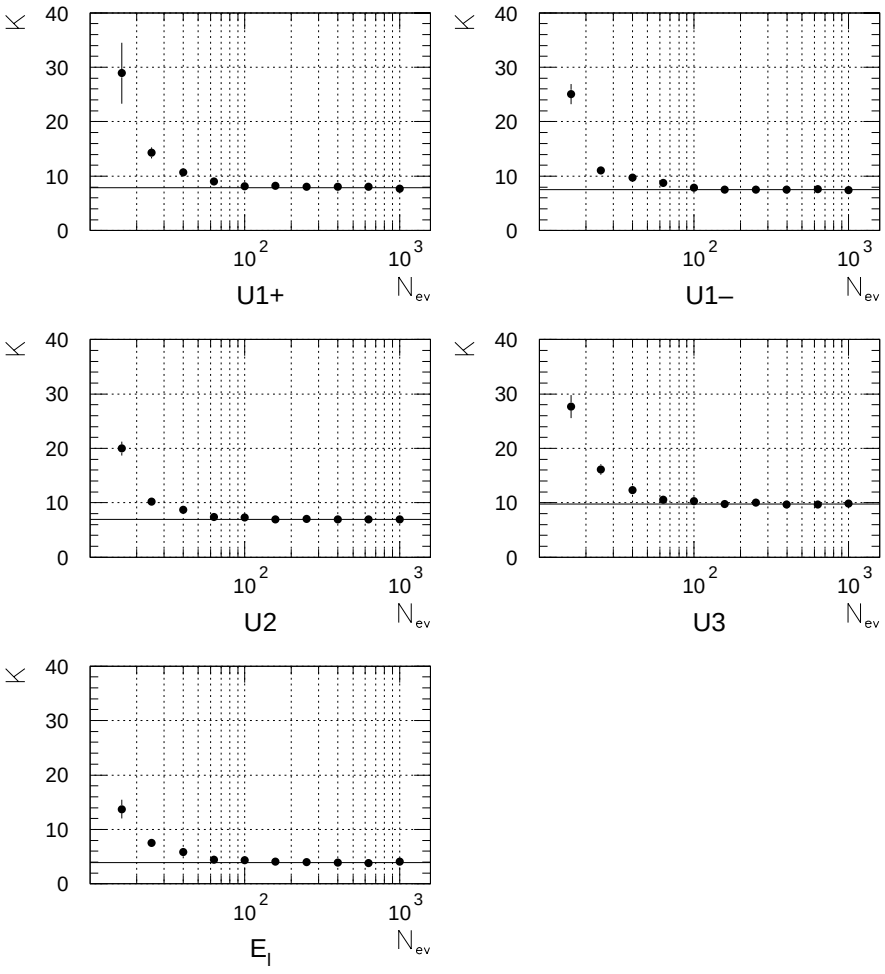


Figure O.1: K-factor *versus* N_{ev} for the observables U1+, U1−, U2, U3, and E_ℓ. The horizontal line in each plot shows the value of K from the continuum approximation.

Appendix P

Analyzing the Apparent Anomaly in the Measurement of r_T

P.1 Introduction

In chapter 10, the results of the measurement of the soft transverse W helicity fraction (r_T) were presented. In the optimization configuration (HTCUT,MTCON) = (YES,YES), r_T was measured to be 1.064, which is off by 2.91 standard deviations from the standard model prediction $r_T = 0.25$. If the modeling of the signal and background were correct, this would be strong evidence for non-standard model physics. There are indications, however, that the modeling is, in fact, inadequate. The main purpose of this appendix is to demonstrate that the modeling is inadequate. A secondary objective is to speculate on the source(s) of this inadequacy.

P.2 Demonstrating the Inadequacy in Modeling the Signal and Background

Evidence for the inadequacy in the modeling of the signal and background comes from the change in the r_T measurement going from the unoptimized to fully optimized analysis configuration. Before optimization, $r_T = 0.675$, which is 1.33 standard deviations from the standard model prediction $r_T = 0.250$. After optimization, $r_T = 1.064$, which is 2.91 standard deviations from prediction. The optimization process involves: (1) application

of H_T cut on certain subsamples of data; (2) application of top mass constraint in the $t\bar{t}$ reconstruction process. These optimizations alter the E_ℓ distribution, so a certain degree of change in the measured value of r_T is expected. An important question in this context is the following:

> Suppose, before the application of H_T cuts, that the experimental data originate from signal and background events in the proportions given in table 10.1. Suppose that the signal events originate from the standard model top quark, so that $r_T = 0.25$. Then, given that r_T is measured to be 0.675 before optimization, how likely is it that r_T is measured to be 1.064 after optimization? In other words, how likely is it that r_T changes by $1.064 - 0.675 = 0.389$?

In order to answer this question, the following pseudo-experiments have been performed:

1. In a given pseudo-experiment, randomly select 159 values of E_ℓ from background and standard model signal events in the proportions given in table 10.1. Note that 159 is the number of events in the experimental data before the application of H_T cuts.

2. Using the log-likelihood method, extract r_T from this set of events.

3. For each of the 159 values of E_ℓ chosen above, there corresponds a different value of E_ℓ obtained by applying a top mass constraint in the $t\bar{t}$ reconstruction process. Switch the original values of E_ℓ to the mass-constrained ones. Also, from among the 159 events, remove those that fail the H_T cuts. This new set of values of E_ℓ gives a fully-optimized version of the E_ℓ distribution.

4. Extract r_T from this set of events.

5. Repeat the above pseudo-experiment $N_{exp} = 10{,}000$ times.

Figure P.1(a) through (d) show the result of the pseudo-experiments. Figure P.1(a) is a contour plot of $r_T(y,y)$ vs. $r_T(n,n)$, where (y,y) and (n,n) stand for the optimization configurations (YES,YES) and (NO,NO). The dashed diagonal line represents $r_T(y,y) = r_T(n,n)$. The two dashed vertical lines correspond to $r_T(n,n) = 0.575$ and $r_T(n,n) = 0.775$ — the region between these lines indicates a window of $\pm$ 0.10 within the experimentally

measured value of $r_T(n,n)$ of 0.675. Figure P.1(b) shows a profile plot of $r_T(y,y)$ vs. $r_T(n,n)$ — the points show the mean of $r_T(y,y)$ in each $r_T(n,n)$ bin, while the vertical bars show the dispersion of $r_T(y,y)$ about the mean in each bin. Figure P.1(c) shows the distribution of $r_T(y,y) - r_T(n,n)$ for all events, while Figure P.1(d) shows the same distribution for experiments in the window $0.575 < r_T(n,n) < 0.775$.

Figure P.1(c) indicates that, on average, $r_T(y,y) = r_T(n,n)$. Furthermore, a typical value of the difference between $r_T(y,y)$ and $r_T(n,n)$ is 0.236 (the r.m.s. of the distribution). This, however, is the situation when averaged over all values of $r_T(n,n)$. In the experimental data, $r_T(n,n)$ is measured to be 0.675. If this is taken into account, then figure P.1(d) shows that, on average, $r_T(y,y) = r_T(n,n) - 0.159$. The spread in the value of $r_T(y,y)$ is 0.206. The fact that $r_T(y,y) < r_T(n,n)$ in the window $0.575 < r_T(n,n) < 0.775$ is expected because the optimization decreases the statistical error in the measurement. In other words, if $r_T(n,n)$ is measured to be $>$ the theoretical value 0.250, then, on average, $r_T(y,y)$ should be closer to the theoretical value because the optimizations decrease the uncertainty in the measurement. Thus, on average, $r_T(y,y)$ should be $< r_T(n,n)$.

Let us examine, in light of the above observations, the change in the r_T measurement going from (NO,NO) to (YES,YES) in the experimental data. Figure P.1(d) shows that, given that $r_T(n,n) = 0.675$, $r_T(y,y) - r_T(n,n)$ is distributed as a gaussian centered at -0.159, with a width of 0.206. In the experimental data, one has $r_T(y,y) - r_T(n,n) = 0.389$. This corresponds to a standard deviation of $(0.389 + 0.159)/0.206 = 2.66$. This shows that the change in r_T observed in the experimental data is rather unlikely to be a statistical fluctuation. In other words, the observed change in r_T is more likely to be due to inadequacies in the modeling of the signal and background than due to statistical fluctuation.

P.3 Speculating on the Source(s) of the Inadequacy in the Modeling of the Signal and Background

The fact that the measured value of r_T increases significantly from (NO,NO) $\rightarrow$

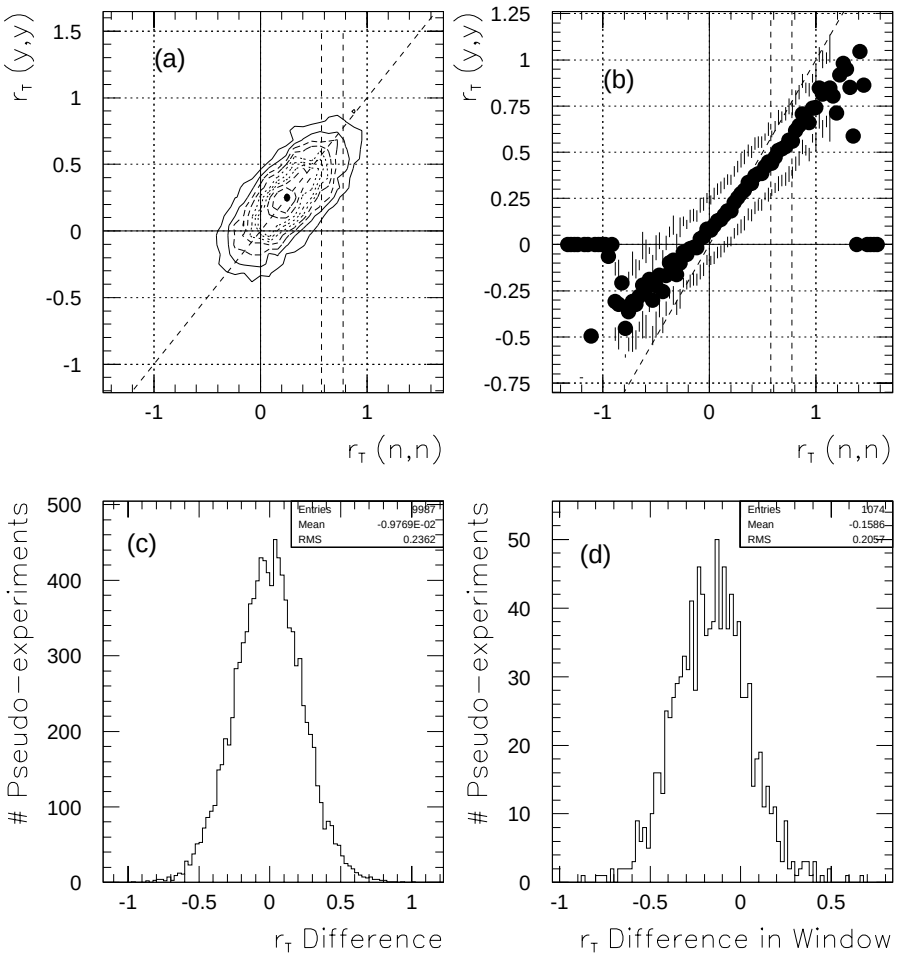


Figure P.1: The correlation between $r_T(n,n)$ and $r_T(y,y)$, as obtained from 10,000 pseudo-experiments. (a) Contour plot showing $r_T(y,y)$ vs. $r_T(n,n)$. (b) Dispersion plot of $r_T(y,y)$ vs. $r_T(n,n)$. (c) Distribution of $r_T(y,y) - r_T(n,n)$, all experiments. (d) Distribution of $r_T(y,y) - r_T(n,n)$ for experiments in the window $0.575 < r_T(n,n) < 0.775$. In (a) and (b), the dashed diagonal line represents $r_T(y,y) = r_T(n,n)$, while the two dashed vertical lines indicate a window of $\pm$ 0.10 about the experimentally measured value of $r_T(n,n)$ of 0.675.

	$< E_\ell >$		
SAMPLE	(NO,NO)	(YES,YES)	$\Delta < E_\ell >$
SIG(SOFT)	43.6	42.9	-0.7
SIG(MED)	52.3	52.7	$+0.4$
BACK	49.7	55.4	$+5.7$
$s+b\,(r_T=0.00)$	50.6	54.1	$+3.5$
$s+b\,(r_T=0.25)$	49.8	52.9	$+3.1$
$s+b\,(r_T=1.00)$	47.6	49.4	$+1.8$
DATA	47.2	47.5	$+0.3$

Table P.1: Average E_ℓ for the soft component of the signal (SIG(SOFT)), the medium component of the signal (SIG(MED)), the background (BACK), signal + background with $r_T = 0$, $r_T = 0.25$, and $r_T = 1$ ($s+b\,(r_T = \cdots)$), and experimental data (DATA). All entries in GeV.

(YES,YES) implies that the experimental E_ℓ distribution is becoming softer relative to the monte carlo signal and background distributions. This relative softening can happen by: (1) the experimental E_ℓ distribution getting softer, while the monte carlo distribution stays fixed; (2) the monte carlo distribution gets harder, while the experimental distribution stays fixed; or (3) both of the above. Table P.1 indicates that (2) describes the situation best. According to the numbers in the table, the hardness/softness of the two components of the signal distribution change little under optimization, whereas the background distribution hardens significantly. Since more than half of the events in the experimental data are expected to originate from background, one would expect significant hardening of the experimental E_ℓ distribution. Instead, it is seen that the experimental distribution stays nearly fixed.

Three plausible scenarios exist that can explain the observed behavior of the experimental E_ℓ distribution. They are: (1) some, or all, of the background fractions are overestimated; (2) the background E_ℓ distribution after optimization does not harden as much as predicted; (3) the signal E_ℓ distribution softens after optimizations. These items are described more fully below:

Background Fractions Overestimated

Table P.1 shows that the two components of the signal E_ℓ distribution stays approxi-

mately fixed under optimizations, while the background distribution hardens significantly. This implies that, the smaller the background content, the smaller the change in the E_ℓ distribution. This suggests that some or all of the background fractions are overestimated. Lowering the background fractions would cause a decrease in the measured value of r_T. Since the measured values of r_T are higher than expected, the lowering of the background fractions would have the side-effect of causing the r_T measurement to be more consistent with the background + standard model signal.

Background Distribution Does Not Harden as Much as Predicted

As was mentioned in chapter 5, the background monte carlo sample was generated using W + 3 jets QCD matrix elements. Although this process is expected to account for the bulk of the background events in data, other processes also make significant contributions (see chapter 4). In the top mass measurement by the CDF collaboration, it was found that the top mass distribution in the W + 3 jets monte carlo is similar to the mass distribution in other background processes. Thus the W + 3 jets monte carlo was used to represent all of the background processes [44]. This similarity in the observable distribution between W + 3 jets and other background processes, however, does not carry over to other observables. For instance, in the W helicity measurement using the lab frame transverse momentum of ℓ [63], there are indications that the $P_T(\ell)$ distribution from the W + 3 jets monte carlo is considerably harder than that from the "non-W" background. Since the "non-W" background accounts for about 10% and 25% of the total background in 3.5-jet and 4-jet events, respectively, the neglect of "non-W" events causes the measured value of r_T to be larger than expected. Although the results in [63] do not propagate directly to the present analysis, the neglect of the "non-W" background in this analysis may well account for the significant shift in r_T going from (NO,NO) $\rightarrow$ (YES,YES). More specifically, the inclusion of non-W background may cause the total background distribution to not harden as much as predicted when W + 3 jets background alone is taken into account. A study on the non-W background has not been performed in this thesis because of time constraints. A future study of the r_T measurement using E_ℓ ought to take this into account.

Signal Distribution Softens

Table P.1 shows that the two components of the signal E_ℓ distribution do not harden/soften by much after optimization. This, however, assumes that the top quark mass is 175 GeV. Currently, the top quark mass, as reported by the CDF collaboration, is measured to be 176.0 GeV $\pm$ 6.5 GeV [35]. Thus it is not unlikely that the true top quark mass is actually 165 GeV or 185 GeV. If the true top quark mass were off from 175 GeV by a rather large margin, then it is not clear that the signal E_ℓ distribution would not change significantly after optimization is applied. In other words, one of the optimizations involves constraining the top quark mass to m_{top} = 175 GeV in the $t\bar{t}$ reconstruction process. Application of this mass cut on a parent sample where m_{top} = 165 GeV or 185 GeV may cause the E_ℓ distribution to harden or soften. Because of time constraints, this issue is not explored in this thesis. A future study ought to take this into account, since it does not seem implausible that the signal E_ℓ distribution might soften because of such an effect, and this softening could account for the observed shift in the r_T measurement from (NO,NO) $\rightarrow$ (YES,YES).

Bibliography

[1] Vernon D. Barger and Roger J. N. Phillips, *Collider Physics*, 1st ed., Addison-Wesley (1987).

[2] Peter Renton, *Electroweak Interactions*, 1st ed., Cambridge University Press (1990).

[3] Gordon Kane, *Modern Elementary Particle Physics*, 1st ed., Addison-Wesley (1987).

[4] Richard Fernow, *Introduction to Experimental Particle Physics*, Cambridge University Press (1992).

[5] F. Abe *et al.*, *Observation of Hadronic W Decays in $t\bar{t}$ Events with the Collider Detector at Fermilab*, Phys. Rev. Lett **80**, 5720 (1998).

[6] F. Abe *et al.*, *Measurement of the Top Quark Mass and $t\bar{t}$ Production Cross Section from Dilepton Events at the Collider Detector at Fermilab*, Phys. Rev. Lett **80**, 2779 (1998).

[7] F. Abe *et al.*, *Measurement of the $t\bar{t}$ Production Cross Section in $p\overline{p}$ Collisions at $\sqrt{s}$ = 1.8 TeV*, Phys. Rev. Lett **80**, 2773 (1998).

[8] F. Abe *et al.*, *Measurement of the Top Quark Mass*, Phys. Rev. Lett **80**, 2767 (1998).

[9] F. Abe *et al.*, *The μ-τ and e-τ Decays of Top Quark Pairs Produced in $p\overline{p}$ Collisions at $\sqrt{s}$ = 1.8 TeV*, Phys. Rev. Lett **79**, 3585 (1997).

[10] F. Abe *et al.*, *First Observation of the All-Hadronic Decay of $t\bar{t}$ Pairs*, Phys. Rev. Lett **79**, 1992 (1997).

[11] F. Abe *et al.*, *Study of $t\bar{t}$ Production in $p\bar{p}$ Collisions Using Total Transverse Energy*, Phys. Rev. Lett **75**, 3997 (1995).

[12] F. Abe *et al.*, *Identification of Top Quarks Using Kinematic Variables*, Phys. Rev. D. **52**, 2605 (1995).

[13] F. Abe *et al.*, *Kinematic Evidence for Top Quark Pair Production in W + Multijet Events in $p\bar{p}$ Collisions at $\sqrt{s}$ = 1.8 TeV*, Phys. Rev. D. **51**, 4623 (1995).

[14] F. Abe *et al.*, *Observation of Top Quark Production in $p\bar{p}$ Collisions with the Collider Detector at Fermilab*, Phys. Rev. Lett **74**, 2626 (1995).

[15] S. Abachi *et al.*, *Observation of the Top Quark*, Phys. Rev. Lett **74**, 2632 (1995).

[16] F. Abe *et al.*, *Evidence for Top Quark Production in $p\bar{p}$ Collisions at $\sqrt{s}$ = 1.8 TeV*, Phys. Rev. D. **50**, 2966 (1994).

[17] F. Abe *et al.*, *Evidence for Top Quark Production in $p\bar{p}$ Collisions at $\sqrt{s}$ = 1.8 TeV*, Phys. Rev. Lett **73**, 225 (1994).

[18] F Abe *et al.*, *Kinematics of $t\bar{t}$ Events at CDF*, to be published in Phys. Rev. D.

[19] P. C. Bhat, H. B. Prosper, and S. S. Snyder, *Top Quark Physics at the Tevatron*, FERMILAB-PUB-98/236 (1998).

[20] M. L. Perl *et al.*, *Evidence for Anomalous Lepton Production in e^+e^- Annihilation*, Phys. Rev. Lett **35**, 1489 (1975).

[21] S. W. Herb *et al.*, *Observation of a Dimuon Resonance at 9.5 GeV in 400 GeV Proton Nucleus Collisions*, Phys. Rev. Lett **39**, 252 (1977).

[22] Robert N. Cahn and Gerson Goldhaber, *The Experimental Foundations of Particle Physics*, 1st ed., Cambridge University Press, (1991).

[23] J. Ellis, *Current Issues in the Phenomenology of Particle Physics*, Int. J. Mod. Phys. A **12**, 5531 (1997).

[24] G. Altarelli, R. Barbieri, and F. Caravaglios, *Electroweak Precision Tests: A Concise Review*, Int. J. Mod. Phys. A **13**, 1031 (1998).

[25] LEP Collaborations, LEP Electroweak Group, and SLD Heavy Flavour Group, *A Combination of Preliminary Electroweak Measurements and Constraints on the Standard Model*, Report No. CERN-PPE/97-154 (1997).

[26] D. Schaile and P. M. Zerwas, *Measuring the Weak Isospin of b Quarks*, Phys. Rev. D. **45**, 3262 (1992).

[27] T. Stelzer, Z. Sullivan, and S. Willenbrock, *Single-top-quark Production at Hadron Colliders*, HEP-PH/9807340 (1998).

[28] S. Abachi *et al.*, *Search for the Top Quark in $p\overline{p}$ Collisions at $\sqrt{s}$ = 1.8 TeV*, Phys. Rev. Lett **72**, 2138 (1994).

[29] V. Barger, J. Ohnemus, and R. J. N. Phillips, *Spin Correlation Effects in the Hadroproduction and Decay of Very Heavy Top Quark Pairs*, Int. J. Mod. Phys. A **4**, No. 3, 617-625 (1989).

[30] G. L. Kane, G. A. Ladinsky, C.-P. Yuan, *Using the Top Quark for Testing Standard-Model Polarization and CP Predictions*, Phys. Rev. D. **45**, No. 1 (1992).

[31] Glenn A. Ladinsky, *Top-Quark Polarization at the Fermilab Tevatron*, Phys. Rev. D. **46**, No. 9 (1992).

[32] T. Arens and L.M. Sehgal. *Azimuthal Correlation of Charged Leptons Produced in $p\overline{p}$ $\rightarrow t\bar{t}$ + ...* Phys. Lett. B. **302**, 501-506 (1993).

[33] Y.S. Tsai. *Decay Correlations of Heavy Leptons in $e^+ + e^- \rightarrow \ell^+ + \ell^-$*, Phys. Rev. D. **4** (1971) 2821-2837. See section IV. Page 2833 is especially instructive.

[34] Y. Hara, *Angular Correlations of Charged Leptons from $t\bar{t}$ Produced in the Gluon Fusion near Threshold*, Prog. Theor. Phys. **86**, No. 4, (1991).

[35] W.-M. Yao, *Top Quark Mass from CDF*, CDF internal note 4762 (1998).

[36] *The CDF Detector for Tevatron Run II — Technical Design Report*, FERMILAB-PUB-96/390-E.

[37] F. Abe, *et al.*, *Measurement of* $\sigma B(W \rightarrow e\nu)$ *and* $\sigma B(Z^0 \rightarrow e^+e^-)$ *in* $p\overline{p}$ *Collisions at* $\sqrt{s} = 1800\ GeV$, Phys. Rev. D. **44**, 29 (1991).

[38] F. Abe, *et al.*, *Measurement of the W Boson Mass*, Phys. Rev. D. **52**, 4784 (1995).

[39] Nathan Eddy, *Study of Z Vertex on Lepton Plus Four Jet Mass Sample*, CDF internal note 3661 (1996).

[40] F. Abe, *et al.*, *Topology of Three-jet Events in* $p\overline{p}$ *Collisions at* $\sqrt{s} = 1.8\ TeV$, Phys. Rev. D. **45**, 1448 (1992).

[41] J. Alitti, *et al.*, *A Measurement of Two-jet Decays of the W and Z Bosons at the CERN* $p\overline{p}$ *Collider*, Z. Phys. C49, 17-28 (1991).

[42] F. Abe, *et al.*, *Study of Four-jet Events and Evidence for Double Parton Interactions in* $p\overline{p}$ *Collisions at* $\sqrt{s} = 1800\ GeV$, Phys. Rev. D. **47**, 4857 (1993).

[43] A. Caner, L. Galtieri, J. Lys, W. Yao, and A. Yagil, *Jet Corrections for Top Mass Fitting*, CDF Internal Note 2469 (1994).

[44] F. Abe, *et al.*, *Measurement of the Top Quark Mass at CDF*, to be published in Phys. Rev. D.

[45] F. Abe, *et al.*, *Measurement of the W-boson Mass in 1.8 TeV* $p\overline{p}$ *Collisions*, Phys. Rev. D. **43**, 2070 (1991).

[46] J. Lys and A. B. Galtieri, *Background Estimate in the Lepton + Jets Top Mass Sample*, CDF internal note 4075 (1997).

[47] G. Marchesini and B. R. Webber, *Monte Carlo Simulation of General Hard Processes with Coherent QCD Radiation*, Nucl. Phys. **B310**, 461 (1988).

[48] G. Marchesini, *et al.*, *HERWIG: A Monte Carlo Event Generator for Simulating Hadron Emission Reactions with Interfering Gluons*, Comp. Phys. Comm. **67**, 465, (1992).

[49] F. Paige and S. D. Protopopescu, BNL Report No. 38034, 1986 (unpublished).

[50] P. Avery, K. Read, and G. Tranhern, Cornell Internal Note CSN-212, March 25, 1985 (unpublished).

[51] CLEO Collaboration World Wide Web Page, `http://www.lns.cornell.edu/public/CLEO/soft/QQ/index.html`

[52] F. A. Berends, W. T. Giele, H. Kuijf, and B. Tausk, *On the Production of a W and Jets at Hadron Colliders*, Nucl. Phys. **B357**, 32 (1991).

[53] J. Benlloch, *Proceedings of the 1992 DPF Meeting*, 10-14 Nov., 1992, Batavia, IL., ed. C. H. Albright *et al.*, World Scientific, 1091 (1993).

[54] N. Eddy and S. Vejcik, *Background Studies for Top Mass Reconstruction*, CDF Internal Note 3920 (1996).

[55] D. Winn and D. Amidei, *Measurement of W Helicity Fractions in Top Decay*, CDF Internal Note 4464 (1998).

[56] J. R. Dittmann, *Studies of Multiple Interactions and Jet Corrections*, CDF Internal Note 4001 (1996).

[57] F. Abe, *et al.*, *Properties of Jets in W Boson Events from 1.8 TeV $p\overline{p}$ Collisions*, Phys. Rev. Lett. **79**, 4670 (1997).

[58] Hisanori Sam Kambara and Xin Wu, *Study of the $t\bar{t}$ Kinematic Fitter Using Events with More than Four Jets*, CDF Internal Note 3789 (1996).

[59] Minuit Reference Manual, CERN Program LIbrary Entry **D506**, 1994.

[60] Review of Particle Properties, Phys. Rev. D. **45** (1992).

[61] Gene Guillian, Myron Campbell, and Dan Amidei, *A Modified Version of HERWIG for Studying the Spin Properties of the Top Quark*, CDF Internal Document 4261, (1997).

[62] J. H. Kühn, *et al.*, *Top Quark Physics — Theoretical Aspects*, 1991 Workshops on Future e^+e^- Colliders, DESY, Hamburg and Saariselkä, Finnland, September 1991.

[63] F. Abe, *et al.*, *Measurement of the Helicity of W Bosons in Top Quark Decays at CDF*, to be published in Phys. Rev. Lett. (1999).

[64] David Gerdes, *Prospects for Observing Single Top Production at the Tevatron*, CDF Internal Document 2865 (1994).

[65] S. Dawson, *The Effective W Approximation*, Nucl. Phys. **B249**, 42 (1985).

[66] S. Dawson and S. Willenbrock, *Heavy Fermion Production in the Effective W Approximation*, Nucl. Phys. **B284**, 449 (1987).

[67] S. Willenbrock and D. A. Dicus, *Production of Heavy Quarks from W Gluon Fusion*, Phys. Rev. D. **34**, 155 (1986).

[68] C. P. Yuan, *A New Method to Detect a Heavy Top Quark at the Tevatron*, Phys. Rev. D. **41**, 42 (1990).

[69] D. O. Carlson and C. P. Yuan, *Studying the Top Quark via the W-gluon Fusion Process*, Phys. Rev. Lett **B306**, 386 (1993).

[70] R. K. Ellis and S. Parke, *Top Quark Production by W Gluon Fusion*, Phys. Rev. D. **46**, 3785 (1992).

[71] A. Caner, *CDFSIM + QFL Simulation of the CDF Detector*, CDF Internal Document 2177 (1993).

[72] M. Shapiro, A. Bhatti, J. Benlloch, R. Harris, T. Rodrigo, P. Sphicas, T. Westhusing, *A User's Guide to QFL*, CDF Internal Document 1810 (1992).

[73] J. Benlloch and T. Rodrigo, *Tracking in QFL*, CDF Internal Document 1634 (1991).

[74] T. Hessing and B. Winer, *A Brief Note Comparing Jet Data to QFL*, CDF Internal Document 1347 (1991).

[75] A. Roodman, *Update on Central Calorimeter Response to Pions and Tuning of QFL*, CDF Internal Document 1344 (1991).

[76] R. Harris, S. Kuhlmann, M. Shapiro, and B. Wicklund, *Simulation of CES Showers and χ^2 in QFL*, CDF Internal Document 1222 (1990).

[77] M. Shapiro, D. Brown, S. Kannappan, *QFL Version 2.0: Improviements and Checks*, CDF Internal Document 753 (1988).

[78] F. Abe *et al.*, *The CDF Detector: An Overview*, Nucl. Instrum. Methods **A271**, 387 (1988).

[79] D. Amidei *et al.*, *The Silicon Vertex Detector of the Collider Detector at Fermilab*, Nucl. Instrum. Methods **A350**, 73 (1994)

[80] F. Bedeschi *et al.*, *Design and Construction of the CDF Central Tracking Chamber*, Nucl. Instrum. Methods **A268**, 50 (1988).

[81] F. Snider *et al.*, *The CDF Vertex Time Projection Chamber System*, Nucl. Inst. and Methods **A268**, 75 (1988). This reference describes the older model of the VTX. The newer model has more modules with shorter drift lengths, but is otherwise similar to the old one.

[82] C. Haber *et al.*, *The CDF SVX: A Silicon Vertex Detector for a Hadron Collider*, Nucl. Inst. and Methods, **A289**, 388 (1990).

[83] D. Amidei *et al.*, *Electrical Performance of the CDF Silicon Vertex Detector*, Nucl. Inst. and Methods, **A342**, 251-259 (1994).

[84] D. Amidei *et al.*, *The Silicon Vertex Detector of the Collider Detector at Fermilab*, Nucl. Inst. and Methods, **A350**, 73-130 (1994).

[85] P. Azzi *et al.*, *The SVX': The New CDF Silicon Vertex Detector*, Nucl. Inst. and Methods, **A360**, 137-140 (1995).

[86] L. Balka *et al.*, *The CDF Central Electromagnetic Calorimeter*, Nucl. Inst. and Methods, **A267**, 272 (1988).

[87] S. R. Hahn *et al.*, *Calibration Systems for the CDF Central Electromagnetic Calorimeter*, Nucl. Inst. and Methods, **A267**, 351 (1988).

[88] K. Yasuoka *et al.*, *Response Maps of the CDF Central Electromagnetic Calorimeter with Electrons*, Nucl. Inst. and Methods, **A267**, 315 (1988).

[89] R. G. Wagner *et al.*, *Cosmic Ray Test of the CDF Central Calorimeters*, Nucl. Inst. and Methods, **A267**, 330 (1988).

[90] T. Devlin *et al.*, *Phototube Testing for CDF*, Nucl. Inst. and Methods, **A268**, 24 (1988).

[91] S. Bertolucci *et al.*, *The CDF Central and Endwall Hadron Calorimeter*, Nucl. Inst. and Methods, **A267**, 301 (1988).

[92] Y. Fukui *et al.*, *CDF End Plug Electromagnetic Calorimeter Using Conductive Plastic Proportional Tubes*, Nucl. Inst. and Methods, **A267**, 280 (1988).

[93] W. C. Carithers *et al.*, *Proceedings of the Gas Sampling Calorimetry Workshop II*, Batavia, Illinois, 1985 (unpublished).

[94] S. Cihangir *et al.*, *The CDF Forward/Backward Hadron Calorimeter*, Nucl. Inst. and Methods, **A267**, 249 (1988).

[95] G. Brandenburg *et al.*, *An Electromagnetic Calorimeter for the Small Angle Regions of the Collider Detector at Fermilab*, Nucl. Inst. and Methods, **A267**, 257 (1988).

[96] G. W. Foster *et al.*, *A Fast Hardware Track-finder for the CDF Central Tracking Chamber*, Nucl. Inst. and Methods, **A269**, 93 (1988).

[97] L. Nodulman and K. Byrum, *Tuning up CEM Using Inclusive Electrons*, CDF internal note 2291, (1993).

[98] L. Nodulman and K. Byrum, *CEM Tower Calibration and Map Trim*, CDF internal note 2487, (1994).

[99] L. Nodulman, *Preliminary CEM Calibration for 1b to the Feb. Shutdown*, CDF internal note 3048, (1995).

[100] L. Nodulman and K. Byrum, *CEM Calibration for 1b (but not 1c)*, CDF internal note 3477, (1996).

[101] Lou Keeble and Brenna Flaugher, *QDJTMC Jet Resolution Study*, CDF internall note 1656 (1992).

[102] Leon M. Lederman, *The Tevatron*, Sci. Am. March 1991 (pp. 48-55).

[103] N. Gelfand, *The Performance of the Tevatron Collider at Fermilab*, FERMILAB-CONF-91/282 (1991).

[104] Robert J. Noble, *The Fermilab Linac Upgrade*, FERMILAB-CONF-91/39, (1991).

[105] The Fermi III World Wide Web site, `http://www-fermi3.fnal.gov`.

[106] The Official Fermilab World Wide Web site, *Click N' Tour the Accelerators*, `http://www.fnal.gov/pub/accel_tour.html`.

[107] The Fermilab Accelerator Division/Proton Source Department, Linac World Wide Web site, `http://www-linac.fnal.gov/linac_tour.html`.

[108] The Fermilab Accelerator Division/Proton Source Department, Booster World Wide Web site, `http://adwww.fnal.gov/proton/booster/booster.html`.

[109] The Fermilab Beams Division/Operations Department World Wide Web site, "Operations Rookie Book", `http://www-bd.fnal.gov/operations/rookie_books/rbooks.html`.

[110] The Fermilab Accelerator Division/Antiproton Group World Wide Web site, "Rookie Book", `http://adwww.fnal.gov/operations/rookie_book/contents.html`.

[111] The CDF World Wide Web site, `http://www-cdf/experiment/experiment.html`.