B_s° Mixing at the SLD Using Inclusive Semileptonic B Decays

Geordie Zapalac

I. Introduction

This note documents current progress in an ongoing analysis of B° mixing using polarized Z° events at the SLD. W. Atwood originally noted that a mixing measurement at SLD could utilize the forward-backward asymmetry of $b\overline{b}$ quarks with respect to the electron beam polarization to tag b or $\overline{b}$ at the interaction point (IP). This allows the SLD to remain competitive with other experiments that have higher statistics but rely on like-sign dilepton events. The effectiveness of the polarization tag increases with the $b\overline{b}$ asymmetry; the analyzing power to distinguish b jets from $\overline{b}$ jets has the form

$$a = 2A_b\Big(\frac{A_e + P}{1 + PA_e}\Big)\Big(\frac{\cos\theta}{1+\cos^2\theta}\Big)$$

where

$$A_b = \frac{L_b^2 - R_b^2}{L_b^2 + R_b^2}$$

$$A_e = \frac{L_e^2 - R_e^2}{L_e^2 + R_e^2}$$

and the signed electron polarization is given by

$$P = 2f_L - 1.$$

L_b, R_b, L_e, R_e are the left and right handed couplings for $Z^\circ \to b\overline{b}$ and $Z^\circ \to e^+e^-$. f_L is the fraction of left-handed electrons in the polarized beam. $\cos\theta \equiv z$ refers to the angle between the B hadron momentum and the direction of the electron beam.

The $Z^\circ \to b\overline{b}$ event is tagged with the high p, p_t lepton emitted from the semileptonic decay of one of the B hadrons. The sign of the lepton tags the sign of the W boson and hence tags the b or $\overline{b}$ at decay. The proper time of the B hadron between its generation at the IP and its decay is given by

$$t = l/\gamma\beta c,$$

where l is the decay length measured with the vertex chamber and drift chamber (typically 3mm) and $\gamma\beta$ is the boost (typically 6.5) determined by measuring the B hadron energy with the drift chamber and the electromagnetic calorimeter. Knowledge of the $b\overline{b}$ character at the IP and decay point and of the proper time allows a determination of the mixing oscillation frequency $X_s = \Delta M_s/\Gamma_s$, where ΔM_s is the mass difference between the B_s° and $\overline{B_s^\circ}$ mesons.

The measurement of X_s is motivated by the following argument. The unitarity of the CKM matrix constrains the quark couplings in a manner that would force the matrix element V_{td} to be complex for a range of magnitudes of V_{td}. A complex value for V_{td} would imply that time reversal invariance is violated in the standard model. The parameter X_d (determined from B_d° mixing) depends on V_{td}, but the theoretical uncertainty in the calculation of X_d is too large to allow a one to reliably extract V_{td} from X_d. Bjorken has pointed out that the ratio X_d/X_s may be calculated to an accuracy of 15%. The unitarity of the CKM matrix provides the constraint that the magnitudes of V_{ts} and V_{bc} are roughly the same, where V_{bc} is measured from semileptonic B decay. Hence a measurement of B_s° mixing may be combined with a measurement of B_d° mixing to determine the magnitude of V_{td}.

$B^\circ\overline{B^\circ}$ mixing will become an increasingly important component of the SLD physics program as the integrated luminosity continues to accumulate. In the following section the likelihood method is described for extracting the mixing frequency X_s from a sample of inclusive semileptonic B decays with measured proper times. The next section describes the analysis to measure the proper time. The final section describes the procedure used to estimate the reach in X_s after the installation of the upgraded vertex detector.

II. Likelihood Analysis to Extract X_s

Suppose that the proper times are measured for a sample of N hadronic Z° events that have been tagged as B events with a high p, p_t lepton (see Fig. 1). No attempt is made to isolate B_s° events. The ith event is characterized by

q_i = lepton charge,

$z_i = cos\theta$ = angle between the jet direction and the electron beam,

P_i = signed polarization of the electron beam,

T_i = proper time measured for this event,

σ_i^l = error on the decay length measurement,

$\sigma_i^E = \Delta_{\gamma\beta}/\gamma\beta$ = fractional error on the boost measurement.

The probability to observe this event is a function of X_s:

$$\mathcal{P}_i(q_i, z_i, P_i, T_i, \sigma_i^l, \sigma_i^E, X_s)$$

with the normalization

$$\int_{-\infty}^{+\infty} \mathcal{P}_i(q_i, z_i, P_i, T, \sigma_i^l, \sigma_i^E, X_s) dT = 1.$$

We may form the log likelihood for the sample of N events, which tends to a maximum at the correct value of X_s

$$\ln \mathcal{L}(X_s) = \sum_{i=1}^{N} \ln \mathcal{P}_i.$$

We assume for this calculation that X_d is known, as well as the fraction of events that are B_d° and B_s°, and we assume that the lifetimes for all the B hadrons are the same and equal to τ_B. The measurement error on z_i is ignored because of the slow variation of $\mathcal{P}_i$ with z_i.

Let t be the actual proper time of the ith event. In general $T_i \neq t$ because of the measurement resolution. Define the unsmeared normalized probability η_i to be the probability of observing the ith event if the proper time were measured with no error

$$\eta_i(q_i, z_i, P_i, t, X_s).$$

To calculate $\mathcal{P}_i$ we must sum over all possible times t. Each term in the sum is weighted by the normalized gaussian that depends on the measurement resolution and on the difference between t and the measured T_i:

$$\mathcal{P}_i = \int_0^\infty \frac{e^{\frac{-(T_i - t)^2}{2\sigma_i^2(t)}}}{[2\pi\sigma_i^2(t)]^{\frac{1}{2}}} \eta_i(q_i, z_i, P_i, t, X_s) dt.$$

The measurement resolution depends upon both the boost and decay length errors:

$$\sigma_i^2(t) = \left(\frac{\partial t}{\partial l}\right)^2 (dl)^2 + \left(\frac{\partial t}{\partial \gamma\beta}\right)^2 (d\gamma\beta)^2 = \left(\frac{\sigma_i^l}{\gamma\beta c}\right)^2 + (t\sigma_i^E)^2.$$

The problem is now reduced to finding the expression for η. We first do the calculation for a sample of pure semileptonic B decays and then afterwards extend this calculation to model the contribution from backgrounds. η is the sum of all possible ways that a tagged B event that lives for time t would yield q, z, and P. We let $l^\pm$ denote the tagging lepton and x denote the remainder of the decay products. For $q = -1$ we have:

(1) $b \to B_u^- \to l^- x$

(2) $b \to B_B^- \to l^- x$

(3) $b \to \overline{B_d^\circ} \to l^- x$

(4) $\overline{b} \to B_d^\circ \to \overline{B_d^\circ} \to l^- x$

(5) $b \to \overline{B_s^\circ} \to l^- x$

(6) $\overline{b} \to B_s^\circ \to \overline{B_s^\circ} \to l^- x.$

For $q = +1$ we have:

(1) $\overline{b} \to B_u^+ \to l^+ x$

(2) $\overline{b} \to B_B^+ \to l^+ x$

(3) $\overline{b} \rightarrow B_d^\circ \rightarrow l^+ x$
(4) $b \rightarrow \overline{B_d^\circ} \rightarrow B_d^\circ \rightarrow l^+ x$
(5) $\overline{b} \rightarrow B_s^\circ \rightarrow l^+ x$
(6) $b \rightarrow \overline{B_s^\circ} \rightarrow B_s^\circ \rightarrow l^+ x$.

Each of the above processes constitutes a term in the sum for η. We write the unnormalized sum $\tilde{\eta}$:

$$\tilde{\eta}^- = \tilde{\eta}^-(q = -1, z, P, t, Xs) = (1) + (2) + (3) + (4) + (5) + (6)\Big|_{q=-1},$$

$$\tilde{\eta}^+ = \tilde{\eta}^+(q = +1, z, P, t, Xs) = (1) + (2) + (3) + (4) + (5) + (6)\Big|_{q=+1}.$$

Each term (i) in the sum is a product of probabilities:

$$\Lambda(P, z, b_j)|\langle b_j|\psi(0)\rangle\langle\psi(0)|\psi(t)\rangle\langle\psi(t)|B^f\rangle|^2.$$

The factor $\Lambda(P, z, b_j)$ is the probability that the B event starts as a b_j quark where either $b_j = b$ or $b_j = \overline{b}$. This factor is calculated from the standard model; it is the forward backward asymmetry for $Z^\circ \rightarrow b\overline{b}$ events produced with polarized electrons:

$$\Lambda\Big(P, z, \frac{b}{\overline{b}}\Big) = \frac{1}{2}\Big\{1 \pm 2A_b\Big(\frac{A_e + P}{1 + PA_e}\Big)\Big(\frac{z}{1 + z^2}\Big)\Big\}.$$

The factor $|\langle b_j|\psi(0)\rangle|^2$ is the probability for the quark b_j to fragment into the strong eigenstate hadron $\psi(0)$ at $t = 0$. We denote this factor by the B hadron fractions A_d, A_s, A_u, and A_B for the production of $B_d^\circ, B_s^\circ, B_u$, and B baryons at the IP.

The factor

$$|\langle\psi(0)|\psi(t)\rangle\langle\psi(t)|B^f\rangle|^2 = I(\psi(0), B^f)$$

is the probability that the hadron $\psi(0)$ is projected into a weak eigenstate, propagates for time t, and is projected into the strong eigenstate hadron B^f when it decays at time t. The hadron B^f chooses the sign of the lepton charge that tags the event. Under the assumption that the lifetimes for all the B hadrons are the same (an assumption easily lifted at the cost of a more complicated expression in the end) we have:

$$I(B_u, B_u) = I(B_B, B_B) = e^{-t/\tau_B},$$

$$I(B_{d,s}^\circ, B_{d,s}^\circ) = I(\overline{B_{d,s}^\circ}, \overline{B_{d,s}^\circ}) = \frac{1}{2}e^{-t/\tau_B}(1 + \cos X_{d,s}\frac{t}{\tau_B}),$$

$$I(B^\circ_{d,s}, \overline{B^\circ_{d,s}}) = I(\overline{B^\circ_{d,s}}, B^\circ) = \frac{1}{2} e^{-t/\tau_B} (1 - \cos X_{d,s} \frac{t}{\tau_B}).$$

As an example, the term (5) for $b \to \overline{B^\circ_s} \to l^- x$ is given by:

$$(5) = \frac{1}{4} \Big\{ 1 + 2A_b \Big(\frac{A_e + P}{1 + PA_e} \Big) \Big(\frac{z}{1 + z^2} \Big) \Big\} A_s e^{-t/\tau_B} (1 + \cos X_s \frac{t}{\tau_B}).$$

Summing the six terms and using

$$A_d + A_s + A_u + A_B = 1,$$

$$a \equiv 2A_b \Big(\frac{A_e + P}{1 + PA_e} \Big) \Big(\frac{z}{1 + z^2} \Big),$$

and normalizing by calculating the integral

$$\mathcal{N}(q, z, P, X_s) = \int_{-\infty}^{+\infty} dT \int_0^{+\infty} \frac{e^{\frac{-(T-t)^2}{2\sigma^2(t)}}}{\sqrt{2\pi\sigma^2(t)}} \tilde{\eta}(t) dt$$

yields the result:

$$\eta = e^{-t/\tau_B} \{ (1 - qa) + qaA_d (1 - \cos X_d \frac{t}{\tau_B}) + qaA_s (1 - \cos X_s \frac{t}{\tau_B}) \} / \mathcal{N},$$

$$\mathcal{N} = \tau_B \Big\{ (1 - qa) + qa(A_d + A_s) - \Big(\frac{qaA_d}{1 + X_d^2} \Big) - \Big(\frac{qaA_s}{1 + X_s^2} \Big) \Big\}.$$

We now extend this result to include backgrounds. Backgrounds are modeled by dividing the tagged sample into three components, each characterized by a single lifetime. The first component is the signal, $B \to l$ characterized by the lifetime τ_B, that has already been described. The second component is the background contribution from $B \to D \to l$, approximately characterized by the longer lifetime τ_{BDl} because of the extra contribution of the D lifetime. This component also contributes six terms to the sum, where $B \to l^\pm x$ is replaced by $B \to D \to l^\mp x$ so that the sign of the tagging lepton is reversed. For $q = -1$ we have:

(1) $\overline{b} \to B^+_u \to D \to l^- x$
(2) $\overline{b} \to B^+_B \to D \to l^- x$
(3) $\overline{b} \to B^\circ_d \to D \to l^- x$
(4) $b \to \overline{B^\circ_d} \to B^\circ_d \to D \to l^- x$
(5) $\overline{b} \to B^\circ_s \to D \to l^- x$

(6) $b \to \overline{B_s^\circ} \to B_s^\circ \to D \to l^- x$.

The third component is modeled as a single term with a simple exponential decay characterized by the lifetime τ_{bkd}. This term includes all remaining backgrounds including lepton misid, $uds \to l$, and prompt $c \to l$. This background is unpolarized in this approximation. In practice the background from lepton misid dominates and the contribution from prompt $c \to l$ is comparatively small.

There is a large and additional background term for the zero lifetime component (mostly due to lepton misid) coming from primary tracks at the IP. This component could conceivably be modeled by a fourth, gaussian component to the background. At present this component is removed by a cut applied to the data to select events above a minimum measured proper time. Because this cut affects the normalization in the likelihood fit, we will discuss it shortly.

We now write η as the sum of these three components:

$$\eta = \{(1-\Theta-\Phi)\eta_1(q,z,P,t,X_s,\tau_B) + \Theta\eta_2(q,z,P,t,X_s,\tau_{BDl}) + \Phi\eta_3(t,\tau_{bkd})\}/\mathcal{N}$$

where Θ is the fraction of tagged events from the $B \to D \to l$ background and Φ is the fraction of events from all remaining backgrounds. These fractions must be estimated from Monte Carlo. The normalization $\mathcal{N}$ must also be recalculated. Proceeding exactly as before we find:

$$\eta_1 = e^{-t/\tau_B}\{(1-qa) + qaA_d(1-\cos X_d\frac{t}{\tau_B}) + qaA_s(1-\cos X_s\frac{t}{\tau_B})\},$$

$$\eta_2 = e^{-t/\tau_{BDl}}\{(1+qa) - qaA_d(1-\cos X_d\frac{t}{\tau_{BDl}}) - qaA_s(1-\cos X_s\frac{t}{\tau_{BDl}})\},$$

$$\eta_3 = e^{-t/\tau_{bkd}},$$

and for the normalization:

$$\mathcal{N} = (1-\Theta-\Phi)\mathcal{N}_1 + \Theta\mathcal{N}_2 + \Phi\mathcal{N}_3,$$

$$\mathcal{N}_1 = \tau_B\Big\{(1-qa) + qa(A_d+A_s) - \Big(\frac{qaA_d}{1+X_d^2}\Big) - \Big(\frac{qaA_s}{1+X_s^2}\Big)\Big\},$$

$$\mathcal{N}_2 = \tau_{BDl}\Big\{(1+qa) - qa(A_d+A_s) + \Big(\frac{qaA_d}{1+X_d^2}\Big) + \Big(\frac{qaA_s}{1+X_s^2}\Big)\Big\},$$

$$\mathcal{N}_3 = \tau_{bkd}.$$

As it is now constructed, the probability distribution

$$\mathcal{P} = \int_0^\infty \frac{e^{\frac{-(T-t)^2}{2\sigma^2(t)}}}{[2\pi\sigma^2(t)]^{\frac{1}{2}}}\eta(t)dt$$

is normalized:

$$\int_{-\infty}^{+\infty} \mathcal{P}(T)dT = 1.$$

If a cut is made to remove all events with a measured proper time less than T_0, then the probability distribution is no longer properly normalized. The normalized distribution may be written as:

$$\mathcal{P} = \frac{1}{\aleph}\int_0^\infty \frac{e^{\frac{-(T-t)^2}{2\sigma^2(t)}}}{[2\pi\sigma^2(t)]^{\frac{1}{2}}}\eta(t)dt,$$

where $\aleph$ is the normalization:

$$\aleph = \int_{T_0}^\infty \mathcal{P}(T)dT = \int_0^\infty \eta(t)\Big\{\frac{1}{2}erfc\Big[\frac{T_0 - t}{\sqrt{2}\sigma(t)}\Big]\Big\}dt.$$

An X_s signal will appear as a dip in the likelihood plot of $\mathcal{L}(X_s)$ versus X_s. This plot will generally have additional structure so that interpreting the statistical significance of the X_s signal from this plot alone may be confusing, particularly for signals with low significance. The significance of the signal may be unambiguously estimated from the significance of the amplitude A_s, the B_s° fraction. One approach would be to fit to both X_s and A_s. The present practice has instead been to replot the likelihood versus A_s with X_s fixed to the best-fit value. If we let $\mathcal{A}_s$ be the fraction of B_s° events assumed in determining X_s and allow A_s to vary, then the fractions of the other species B_i in the expression for the likelihood probability $\mathcal{P}$ are given by:

$$\frac{A_i}{1-\mathcal{A}_s}(1 - A_s).$$

This plot is parabolic with no additional structure. The significance is taken as the difference in likelihood between the minimum of the plot and the point on the plot where $A_s = 0$, which represents the hypothesis of no B_s° mixing:

$$\sigma(X_s) = \sqrt{2(\ln\mathcal{L}(X_s, A_s = 0) - \ln\mathcal{L}^{min})}.$$

III. Measurement of the Proper Time

A measurement of the B hadron proper time requires a measurement of both the B decay length and the B energy. The decay length is measured by first isolating a two prong vertex from the cascade D decay. The vertex position and the resultant of the momenta of the prongs from the D decay define a "virtual D track". This track is intersected with the tagging lepton track to find the B vertex; the B vertex is placed on the lepton track at the distance of closest approach to the virtual D track. The B energy is estimated by summing the fragmentation energy in the hemisphere defined by the calorimeter thrust axis. The fragmentation energy is subtracted from half of the Z° energy to estimate the B energy. This procedure circumvents the problem of estimating the energy from the missing neutrino. Each B event that survives the cuts to be assigned a decay length is also assigned a B energy.

Decay Length

The large number of fragmentation tracks from the IP makes the decay length determination difficult, particularly if a cut in the B hadron decay length is avoided to prevent biasing the proper time distribution. Selection of D vertices based on one variable such as the χ^2 of the vertex fit is inadequate for effectively separating true vertices from "fakes", where at least one of the tracks in the fit does not originate from the B hadron cascade. The strategy outlined here is to apply a large number of vertex cuts designed to exploit the topology of the semileptonic B decay. This approach may be extended to isolate semileptonic D decays that originate from B hadrons, and perhaps also extended to the problem of vertexing hadronic B decays without a lepton track. These extensions of the algorithm are under study and not discussed further in this note. The vertex efficiency for semileptonic B decays is typically about 25% for the present vertex detector (VXD2) and about 40% for the planned upgrade (VXD3).

It should be emphasized that the algorithm described below is a snapshot of an ongoing study. Work is now in progress, for example, to search for and correct biases in the B lifetime that are introduced by the vertex cuts. All descriptions refer to VXD2 unless stated otherwise.

The first step in the analysis is to select all of the "good quality" tracks in the event that have a link to the vertex detector. Tracks with a total momentum less that 0.5 GeV/c are rejected because the rate of mislinking for these tracks is greater than a few percent. Tracks must also satisfy $|\cos\theta_{beam}| < 0.76$, where θ_{beam} is the angle between the track and the electron beam; the electron beam lies on the Z-axis. This cut restricts the track to the fiducial region of the vertex detector. The distance of closest approach of the track to the IP must be less than 1 cm in the XY plane and less than 1 cm in the Z. Further cuts are applied to select tracks well measured in the CDC: tracks must have at least 40 CDC hits with a momentum in the XY plane greater than 200 MeV/c. There is also a cut requiring the reduced χ^2 for the CDC track fit to be less than 5, and a very loose cut requiring the reduced χ^2 for the combined CDC/vertex detector track to be less than 10.

For the case of VXD3 the cuts are loosened to take advantage of the superior vertex detector tracking and the increased solid angle coverage. A track must satisfy $|\cos\theta_{beam}| < 0.9$, have a total momentum greater than 300 MeV/c, and have at least 20 hits in the CDC. The remaining track quality cuts are identical to those for VXD2.

Semileptonic B decays are selected by identifying an electron or muon and applying the following elliptical cut introduced by the A_b group:

$$\left(\frac{p(GeV/c)}{16}\right)^2 + \left(\frac{p_t(GeV/c)}{1.2}\right)^2 \geq 1,$$

where p is the total momentum of the lepton and p_t is the transverse momentum with respect to the closest calorimeter jet. The lepton must also be a good quality track with $p > 2$ GeV/c. Jets are found using a ycut of 0.005; an event typically has three or four jets with this ycut. The jet includes the lepton track. This occurs automatically for electrons because jets are found using calorimeter clusters. For muons the jet direction must be recalculated after summing the momentum of the muon and the jet.

Each event is divided into two hemispheres using the calorimeter thrust axis. If multiple leptons are selected within the same hemisphere, a single lepton is chosen. In the case of two or more muons, the highest quality muons are first selected based on the muon identification cuts. Leptons are then selected for the highest p_t. If two leptons share the same jet but occur in opposite hemispheres, the highest p_t lepton is also selected.

If a hemisphere contains a tagging lepton, the remaining good quality tracks are further screened to collect "vertex quality" tracks which will be used to search for the D vertex. Each vertex quality track must have a normalized impact paramter in the XY plane ("bnorm"$=b_{xy}/\sigma$) greater than 1. This cut reduces the multiplicity of fragmentation tracks in the hemisphere to about 1.3. The distance of closest approach of the track to the IP must be less than 3 mm to select against K_s° and hyperon decays. A new version of the vertexing algorithm that does not include these potentially biasing cuts is now under study.

The vertex quality tracks are combined in pairs to create a collection of two prong vertices. The vertex fitting routine (ZXFIT from the SLD ZXFIND package) is called separately for each pair of tracks; it returns a vertex with its associated χ^2 and error matrix. The vertex cuts described below are then applied to these vertices to select true D vertices.

Vertices are rejected if the lepton and both vertex tracks all have the same charge. These events would imply at least four D prongs for B_d° decays and five D prongs for B_s° decays; it is more likely that such a vertex would be a fake or originate from a charged B decay.

The two vertex prongs are combined as pions to form an invariant mass m_D. This mass cannot be greater than the D mass, and some background rejection is achieved with a lower mass cut as well. We require $0.5 < m_D < 1.9$GeV ($0.3 < m_D < 1.9$GeV for VXD3). The two prongs are also combined as pions with the lepton to form the invariant "B" mass m_B. Monte Carlo indicates that essentially no real vertices are lost by the cut $m_B < 4.5$ GeV.

The χ^2 from the vertex fit is required to be less than 5. Because a badly measured track may yield a good χ^2, it is also required that the distance of closest approach between the two prongs be less than $120\mu m$.

The D vertex is expected to be displaced from the IP because both the B and D hadrons have significant lifetimes. The vector V^{IP} connecting the IP to the D vertex is projected along the direction $\hat{J}$ of the B jet to obtain the signed distance $\hat{J} \cdot V^{IP} = V_{||}^{jet}$. This distance reflects how far downstream the D hadron is displaced from the IP along the B direction. We require $V_{||}^{jet} > 500\mu m$ ($250\mu m$ for VXD3). This distance may also be normalized to the error in determining the D vertex position. The error in the D vertex position along the jet direction may be calculated using the error matrix M returned by the vertex fit:

$$(\delta V_{||}^{jet})^2 = \sum_{i,j} \left(\frac{\partial V_{||}}{\partial x_i}\right)\left(\frac{\partial V_{||}}{\partial x_j}\right) M_{ij} = \hat{J}_i \hat{J}_j M_{ij}.$$

The normalized distance $V_{||}^{jet}/\delta V_{||}^{jet}$ is required to be greater than 1.

The B vertex is calculated from the distance of closest approach between the lepton and the D "track" as described at the beginning of this section. The distance of closest approach between the virtual D track and the lepton must be less than $120\mu m$. The D vertex is then required to be downstream of the B vertex. The vector V^B connecting the B vertex to the D vertex is projected along the direction $\hat{D}$ of the virtual D track (the two prong momentum resultant) to obtain the signed distance $\hat{D} \cdot V^B = V_{||}^D$. This distance is normalized to the error in the D vertex position projected along the direction of the virtual D track:

$$(\delta V_{||}^D)^2 = \hat{D}_i \hat{D}_j M_{ij}.$$

The normalized distance $V_{||}^D/\delta V_{||}^D$ is required to be greater than 1.

After the vertex cuts are applied, about 17% of the selected vertices are fakes. Of these fakes, about half contain one D track and one B track. This class of fakes still yields a good measurement of the B vertex; the error in the vertex position normal to the D direction is still small. The remainder of the fakes (less than 10%) contain at least one fragmentation track. (As of the time of writing, these numbers have not been rechecked for the RECON 10 Monte Carlo, although the remainder of the results do refer to the RECON 10 Monte Carlo.) The vertex efficiency for VXD2 is $27 \pm 1.4\%$. The vertex cuts are very effective in reducing background in the high p, p_t lepton sample: the events remaining after the vertex cuts are applied are 86% $B \to$lepton events. This is to be compared with the purity of about 70% for the high p, p_t lepton sample before the vertex cuts are applied. The fractions of background events are $4.4 \pm 0.9\%$ from $B \to D \to$lepton, $0.7 \pm 0.4\%$ from prompt $D \to$lepton,

$6.2 \pm 1.1\%$ from lepton misid (including punchthrough in the WIC), and $2.4 \pm 0.7\%$ from all remaining sources. The fraction of high p, p_t lepton events (including all backgrounds) that vertex in the Monte Carlo is $20.1 \pm 0.9\%$. The corresponding fraction for the 1993 data is $17.5 \pm 1.5\%$. This is a 1.5σ discrepancy. At this time this discrepancy is believed to be a statistical fluctuation; distributions of all the variables that enter the vertex cuts have been overlaid for data and Monte Carlo and show good agreement.

For the case of VXD3, the cut on the normalized distance for the D vertex - IP is not applied. The cut on the D vertex - B vertex is made on the absolute quantity rather than the normalized distance; we require $V_{||}^{D} > 250\mu m$. The remaining cuts are identical to those for VXD2 except where indicated in the above discussion. About 26% of the selected vertices are fakes, of which about 66% contain one D track and one B track. The remainder of the fakes (again less than 10%) contain at least one fragmentation track. The vertex efficiency for VXD3 is 41%. The background from $B \rightarrow D \rightarrow$lepton after vertex cuts is reduced by about a factor of two over the background in the high p, p_t sample. The other backgrounds are still being studied for VXD3.

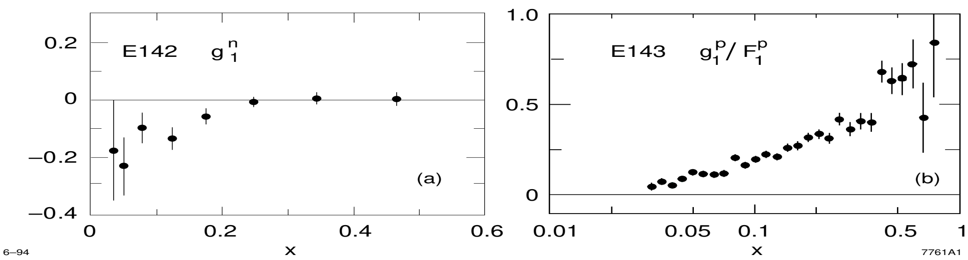

Fig. 1. Electroweak data in the S,T plane.

It is possible that multiple vertices will pass the selection cuts. Because each track has a 90% probability to come from the B cascade, all selected vertices are regarded as originating from the B cascade. One approach would be to combine all of the independent tracks to form a single vertex. This is not done in the present analysis. The two prong vertex closest to the B vertex is taken to define a position on the virtual D track, and the independent tracks from all of the selected vertices are summed to recalculate the direction of the D. The same set of independent tracks will enter later into the boost calculation.

The decay length is taken as the distance between the B vertex and the beam position calculated for that event. The direction of the B jet is used to sign the decay length. The decay length residual $l^{measured} - l^{generated}$ is well approximated by a gaussian with $\sigma = 200\mu m$ (Fig. 2). The decay length error Δl depends upon the opening angle θ_D between the lepton and the virtual D track, the error $\Delta\theta_D$ in the direction of the virtual D track, the distance l_D between the D vertex and the B vertex, and the error $\delta V_\perp^D$ in the placement of the vertex normal to the direction of the D:

$$(\Delta l)^2 = \left(\frac{l_D \Delta\theta_D}{\sin\theta_D}\right)^2 + \left(\frac{\delta V_\perp^D}{\sin\theta_D}\right)^2.$$

The $\sim 10\mu m$ error in the decay length contributed by the beam spot is ignored. The error $\Delta\theta_D$ in the direction of the D depends upon neutrals and other charged tracks from the D that are not included in the momentum resultant. This error is estimated from Monte Carlo to be 3°. The error in the placement of the vertex is calculated from the error matrix in the manner outlined in the above discussion:

$$(V_\perp^D)^2 = |V^B|^2 - (\hat{D}\cdot V^B)^2,$$

$$(\delta V_\perp^D)^2 = \left(\frac{1}{V_\perp^D}\right)^2 \sum_{ij} [V_i^B - \hat{D}_i(\hat{D}\cdot V^B)][V_j^B - \hat{D}_j(\hat{D}\cdot V^B)]M_{ij}.$$

In practice the error in the direction of the D dominates the total error.

Boost

The boost algorithm is designed to avoid estimating the energy of the unseen neutrino while measuring as much as possible about the B energy with the full SLD detector. The sum of the energies of all the tracks assigned to the B cascade (there are at least three tracks) is taken to be the minimum possible energy of the B hadron E_B^{min}. The D tracks are taken to be pions for this sum. The B energy E_B is then calculated by estimating the total fragmentation energy in the hemisphere defined by the calorimeter thrust axis:

$$E_B = \frac{M_{Z^\circ}}{2} - E_f^{charged} - E_f^{neutral},$$

where $E_f^{charged}$ and $E_f^{neutral}$ are estimations of the charged and neutral fragmentation energy in the hemisphere. If it is found that $E_B < E_B^{min}$, then we set $E_B = E_B^{min}$.

The charged fragmentation energy $E_f^{charged}$ is estimated by summing the energy of the charged tracks found by the CDC that have not been assigned to the B cascade. All charged tracks are assumed to be pions. No link to the vertex detector is required for these tracks, and some of the track quality cuts are looser than those required for tracks used in the vertexing: the distance of closest approach of the track to the IP is required to be less than 2 cm in the XY plane and in the Z direction, and we now require $|\cos\theta_{beam}| > 0.80$ with $p_t > 150$ MeV/c.

The neutral fragmentation energy $E_f^{neutral}$ is estimated by summing electromagnetic clusters in the calorimeter that are not associated with a charged track. A calorimeter cluster is regarded as electromagnetic if at least half of the energy in the first four layers of the LAC is localized in the first two layers.

The boost is calculated from:

$$boost = \gamma\beta = E_B/M_B\sqrt{1-(M_B/E_B)^2}.$$

The fractional residual in the boost $(\gamma\beta^{measured} - \gamma\beta^{generated})/\gamma\beta^{generated}$ is roughly described by a gaussian with σ =9% (Figure 2). The distribution appears reasonably centered, but the predominantly positive tails offset the entire distribution by about 3%. Currently there is no existing algorithm to estimate the error on the boost on an event by event basis.

Combining the decay length and boost measurements, we may plot the residual of the proper time $(l/\gamma\beta c - t^{generated})/\tau_B$ (Figure 2). This distribution is roughly estimated by assigning half the events to a gaussian with σ =7% and the remaining half to a gaussian with σ =25%. The corresponding residual plots for VXD3 are shown in Figure 3, overlaid with the residuals from VXD2.

An analysis of the 1993 data for the average B lifetime and the B_d° mixing is currently in progress. We mention here that the measured lifetime is in good agreement with the world average, and that the statistics for this data sample probably do not allow a measurement of the B_d° mixing.

IV. Estimate of the Reach in X_s for the Upgraded Vertex Detector

The estimate of the reach in X_s outlined here refers only to the analysis of inclusive semileptonic B decays described in the previous three sections. The actual analysis for VXD3 will combine this analysis with other techniques that are beyond the scope of this report. As with the other sections in this report, this section represents a progress report rather than a completed analysis.

In section II we described a procedure to estimate the significance of a signal ("bump") in the likelihood plot for X_s. If we knew to look for X_s only in the region where the bump occured, we could interpret the significance in the usual manner. However, we do not know at what X_s the bump will appear. Hence the significance does not carry the same weight as one would first guess. The safest procedure to estimate the reach in X_s is to use the likelihood significance only as a yardstick, and quote the efficiency and purity of the signal passing a certain significance cut for each X_s. The efficiency for an input value of X_s and a given significance cut is the probability that an SLD experiment will yield a signal passing the significance cut and reproducing the correct value of X_s. Given a signal that passes the significance cut, we define the purity to be the probability that the true value of X_s is the value we observe. For the remainder of this discussion we will refer to the fake rate= 1−purity.

Ideally one would want to run a large number of SLD Monte Carlo experiments to estimate the efficiency and fake rate. This would require sufficient CPU time to generate several thousand experiments, each with one million Z^0 events. In reality only a single Monte Carlo experiment with $\sim 10,000$ semileptonic B events was available. To overcome this computing obstacle it was decided to write a fast "standalone" Monte Carlo program. Each SLD experiment became a single statistically independent run of the standalone program. The full SLD Monte Carlo was used to determine the boost residual, decay length residual, and vertex efficiency versus $\cos\theta$ used by the standalone program.

The efficiency was determined by running N identical but independent SLD Monte Carlo experiments at a given X_s and counting the number N_{pass} of experiments that passed the significance cut (typically 3 to 4 σ) with a signal "close" to the input value of X_s. The efficiency is then $\epsilon(X_s, cut) = N_{pass}(X_s, cut)/N$. "Close" to X_s is defined here to be within 2σ of the input value of X_s, where σ refers to the width of the measured X_s distribution for the N experiments. This width is typically 10% to 20%, and increases with increasing X_s (see Figure 4 as an example of this distribution for $X_s = 10$). The fake rate was studied by running 10000 experiments with X_s values distributed in a "line source" within the range $7 < X_s < 51$. This is the range allowed by taking the theoretical central value of all unknown quantities in the box diagram for ΔM_s with the exception of the top quark mass, which is allowed to vary over its full range of uncertainty (before the announcement from CDF). A subset of these runs was chosen with input values of X_s at least 3σ away from the particular X_s under study. Let $f(X_s, cut)$ be the fraction of experiments in this subset that both passed the significance cut and yielded a signal within 2σ of the X_s under study. The fake rate is then $fr = f(X_s, cut)/\epsilon(X_s, cut)$. It is important to keep the following distinction in mind: the fake rate as defined here is not the probability that one will measure a spurious signal at X_s, it is instead the probability that a signal *already measured* at X_s is spurious. It should be noted that the distribution in X_s of fakes is quite flat for a significance cut $> 2.5\sigma$ independent of the X_s under study. The property was not expected *a priori* but supports our approach to determining the fake rate.

Standalone Program and Lookup Table

A standalone experiment assigns the B species, lepton charge, $z = \cos\theta$, signed polarization, smeared boost, and smeared decay length to each of n semileptonic B events. The quantity z for the direction of the B jet is selected randomly from the distribution $1 + z^2$. The $|z|$ distribution is divided into nine bins of width $\Delta z = 0.1$ for comparison with the full SLD Monte Carlo acceptance. The standalone program then randomly throws away events as a function of z bin to reproduce the falloff in acceptance for vertexed events observed in the full SLD Monte Carlo (recall that at least three tracks are required to vertex the event). The number of events remaining at this stage reflects the average vertexing efficiency observed in the full SLD Monte Carlo. A B species (B_d, B_s, B_u, or B baryon) is randomly assigned to each event so that the final fraction for each species reflects the input to the program. The signed polarization P is then assigned with equal numbers of left and right-handed events. Once the P and z are assigned, it is possible to randomly assign the flavor b or $\bar{b}$ to the event at the IP using the probability from Section II. The lepton charge is now decided for each event by ignoring the mixing at this stage. The proper time t of the B is then selected randomly from the distribution $\exp(-t/\tau_B)$. At this point the neutral B_d and B_s species are allowed to mix by changing the sign of the lepton charge according to the probability $0.5(1 - \cos(Xt/\tau_B))$, where $X = X_d$ for B_d and $X = X_s$ for B_s.

Having generated the inclusive semileptonic B sample, it is now necessary to introduce the measurement errors for the proper time. The boost $\gamma\beta$ is crudely generated by selecting randomly from a gaussian distribution centered at 6.5 with $\sigma = 1.0$. The decay length is then given by $l = \gamma\beta ct$. This quantity is smeared by one of three gaussians to reflect the decay length residual observed in the full SLD Monte Carlo. The full Monte Carlo residual is modeled separately in the standalone program for $z < 0.5$ and $z > 0.5$ to account for degradation in the decay length measurement due to increased multiple scattering at large z. The boost is similarly smeared to reflect the boost residual from the full Monte Carlo, and the small offset observed in the boost residual is also introduced (as of this writing, only the older less precise results for the boost have been used in the standalone analysis). The measured proper time is then given by $T = l^{smeared}/(\gamma\beta c)^{smeared}$. Because we have the information to calculate the decay length error σ^l, we assign to σ^l the σ of the gaussian chosen to smear the decay length. The boost error σ^E is assigned the value 15% for every event.

This analysis was done before the backgrounds had been carefully studied. As a consequence, the $B \to D \to$lepton background only was crudely introduced by flipping the sign of the lepton charge at random for 6% of the events. This is now believed to be a conservative estimate: the lepton originates from a D with significant lifetime and therefore does not exactly cancel the oscillations from the proper time distribution of B decays (where the lepton originates from the B). The unsmeared probability distribution η used in the likelihood fit reflects this approximation to the background:

$$\eta = e^{-t/\tau_B}\{[1 - qa(1-2\Theta)] + qa(1-2\Theta)[A_d(1-\cos X_d\frac{t}{\tau_B}) + A_s(1-\cos X_s\frac{t}{\tau_B})]\}/\mathcal{N},$$

where Θ is the fraction of $B \to D \to$ lepton background and the normalization $\mathcal{N}$ is given by:

$$\mathcal{N}_1 = \tau_B\Big\{[1-qa(1-2\Theta)]+qa(1-2\Theta)(A_d+A_s)-\Big(\frac{qaA_d(1-2\Theta)}{1+X_d^2}\Big)-\Big(\frac{qaA_s(1-2\Theta)}{1+X_s^2}\Big)\Big\}.$$

A serious computational obstacle remains: the CPU time required to numerically perform the integrals for the likelihood fit. To generate the likelihood plot versus X_s requires a large fraction of an hour on a DEC Alpha machine. This time was reduced to about six seconds by creating a lookup table for the numerical integrals.

The expression for the likelihood probability may be reduced to evaluating three integrals:

$$\mathcal{I}_1 = \int_0^\infty \frac{e^{\frac{-(T-t)^2}{2\sigma^2(t)}}}{\sigma(t)} e^{t/\tau_B} dt,$$

$$\mathcal{I}_2 = \int_0^\infty \frac{e^{\frac{-(T-t)^2}{2\sigma^2(t)}}}{\sigma(t)} e^{t/\tau_B} \cos X_d \frac{t}{\tau_B} dt,$$

$$\mathcal{I}_3 = \int_0^\infty \frac{e^{\frac{-(T-t)^2}{2\sigma^2(t)}}}{\sigma(t)} e^{t/\tau_B} \cos X_s \frac{t}{\tau_B} dt.$$

These integrals are evaluated on a three dimenional grid in the proper time t (801 grid points from 0 to $8\tau_B$), the component of proper time error from the decay length measurement $\sigma^l/\gamma\beta c$ (30 grid points from 0.01 to 0.30), and X_s (41 grid points from $X_s = 0$ to $X_s = 40$).

Assumptions and Results

Following the procedure outlined above, the following assumptions were made when running the standalone SLD experiments for VXD3:

1) One million hadronic Z° events,

2) 7.5% efficiency to tag semileptonic B decays,

3) 75% electron beam polarization,

4) the following fractions: 15% B_s, 38% B_d, 38% B_u, and 9% B baryons (treated exactly like B_u in this analysis),

5) 100% electron identification to $|z| = 0.77$ with no electrons identified beyond $|z| = 0.77$ because of the increased radiation lengths for material blocking the LAC,

6) 100% muon identification to $z = 0.9$,

7) 41% average vertexing efficiency,

8) 100% tracking efficiency to $|z| = 0.8$ with the SLD Monte Carlo tracking efficiency beyond $|z| = 0.8$ (this is nearly equivalent to using the tracking efficiency for the SLD Monte Carlo for all values of z).

Figure 5 shows the likelihood plots for twenty randomly selected experiments using the input value $X_s = 10$. As of this writing, errors were found in the study performed in the 1993 fall study for the EPAC presentation which tended to underestimate the efficiency for measuring X_s by 10 to 15%. The study has since been repeated for VXD3 only at X_s =10, 12.5, and 15 at significance cuts of 3σ ($\sim$ 5% fake rate), 3.5σ ($\sim$ 3% fake rate), and 4σ ($\sim$ 2% fake rate). These results are shown in Figure 6. We see that the efficiency has dropped to 55% for $X_s = 15$ at a 5% fake rate. For VXD2 the same efficiency occurs at about $X_s = 5$. These efficiencies drop steeply with increasing X_s; a reliable measurement in VXD3 beyond $X_s = 20$ is probably impossible.