B_s° Mixing Using Inclusive Semileptonic B Decays

Geordie Zapalac

I. Introduction

This note presents an analysis method to measure $B_s^\circ \bar{B}_s^\circ$ mixing using $Z^\circ \to b\bar{b}$ events. It was noted by J.D. Bjorken that a measurement of the mixing frequency $X_s = \Delta M_s/\Gamma_s$ for $B_s^\circ \bar{B}_s^\circ$ mixing may be combined with a measurement of X_d from $B_d^\circ \bar{B}_d^\circ$ mixing to determine the magnitude of the CKM matrix element V_{td}. This result may then be combined with measurements of V_{cb} and V_{ub} to test if CP violation is consistent with being a consequence of the three generation standard model.

Rather than isolate B_s° decays, all events tagged as semileptonic B decays with a high p, p_t lepton are used in the analysis; mixed B_s° decays will provide an oscillatory component in the B lifetime distribution from which X_s may be determined. This measurement requires some discrimination between b and $\bar{b}$ at the IP (interaction point) and the charge of the lepton tag to identify a candidate mixed event. The measurement also requires a precise measurement of the B decay time to determine the time evolution of the mixing so that the frequency X_s can be measured.

For the special case of polarized Z°s, W. Atwood originally suggested that a mixing measurement should utilize the forward-backward asymmetry of $b\bar{b}$ quarks with respect to the electron beam polarization to tag the b or $\bar{b}$ quark at the IP. The analyzing power to distinguish b quarks from $\bar{b}$ quarks is given by the standard model:

$$A^{pol} = 2A_b\left(\frac{A_e + P}{1 + PA_e}\right)\left(\frac{\cos\theta}{1+\cos^2\theta}\right), \tag{1}$$

$$\text{where} \quad A_b = \frac{L_b^2 - R_b^2}{L_b^2 + R_b^2}, \qquad A_e = \frac{L_e^2 - R_e^2}{L_e^2 + R_e^2},$$

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 $Z^\circ \to b\bar{b}$ and $Z^\circ \to e^+e^-$. f_L is the fraction of left-handed electrons in the polarized beam. The angle θ is between the b or $\bar{b}$ direction (taken to be the direction of the jet closest to the lepton tag) and the direction of the electron beam. At the SLD detector where $P = 80\%$ and the solid angle coverage of the upgraded vertex detector extends to $|\cos\theta| < 0.90$, the forward-backward asymmetry provides the most effective means of distinguishing b from $\bar{b}$ at the IP, with A^{pol} ranging from zero at $\cos\theta = 0$ to 78% at $\cos\theta = 0.8$.

Unfortunately, the impact parameter resolution of the vertex detector is most degraded by multiple scattering at large $\cos\theta$ where the forward-backward asymmetry is greatest. The momentum-weighted track charge Q^{jet} in the hemisphere opposite to the tagging lepton provides another albeit less effective means of distinguishing b from $\bar{b}$ at the IP with nearly uniform sensitivity over the acceptance of the detector. The thrust axis defines the hemisphere in which the tracks are summed. The momentum-weighted track charge analyzing power A^{jet} depends upon Q^{jet}; it is taken from a Monte Carlo count of events with b and $\bar{b}$ quarks as a function of Q^{jet} in the opposite hemiphere:

$$A^{jet}(Q^{jet}) = \frac{N_b(Q^{jet}) - N_{\bar{b}}(Q^{jet})}{N_b(Q^{jet}) + N_{\bar{b}}(Q^{jet})} \qquad \text{where} \qquad Q^{jet} = \frac{\sum_{i=1}^{n} q_i \vec{p}_i \cdot \hat{t}}{\sum_{i=1}^{n} \vec{p}_i \cdot \hat{t}}. \tag{2}$$

The quantities q_i and $\vec{p}_i$ are the charge and momentum of of track i, $\hat{t}$ is a unit vector along the thrust axis, and n is the total number of tracks in the hemisphere. Note that Q^{jet} is normalized: $-1 < Q^{jet} < 1$. A^{jet} versus Q^{jet} is an odd function well modeled with a third order polynomial fit. On average, $A^{jet} \sim 30\%$.

A charged kaon (typically from $B \to D \to K$) or high p, p_t lepton (from $B \to$lepton) in the opposite hemisphere provides an additional analyzing power A^K. A^K may take on five discrete values: $\sim \pm 60\%$ if a lepton is found (at a few percent efficiency), $\sim \pm 40\%$ if a charged kaon is found (at about 45% efficiency), and zero if no charged kaon or lepton is found.

The probability Λ^α that the event began as a b or $\bar{b}$ quark at the IP is given by:

$$\Lambda^\alpha \begin{pmatrix} b \\ \bar{b} \end{pmatrix} = \frac{1}{2}(1 \pm A^\alpha), \tag{3}$$

where A^α is the analyzing power of type α. Although we expect some correlation between A^{jet} and A^K, Monte Carlo studies support combining all analyzing powers as if they were independent to derive a joint probability:

$$\Lambda \begin{pmatrix} b \\ \bar{b} \end{pmatrix} = \frac{(1 \pm A^{pol})(1 \pm A^{jet})(1 \pm A^K)}{2(1 + A^{pol}A^{jet} + A^{pol}A^K + A^{jet}A^K))}, \tag{4}$$

where the normalization is given by requiring $\Lambda(b) + \Lambda(\bar{b}) = 1$.

The X_s measurement must be sensitive to the B_s° events that mix to the antiparticle state and decay to a lepton at proper time t. The probability to decay to a "wrong-sign" lepton at time t is proportional to:

$$A_s \frac{e^{-t/\tau}}{2} \left[1 - \cos(X_s t/\tau)\right]. \tag{5}$$

where A_s is the fraction of B_s° events and τ is the B_s° lifetime. For the remainder of this note we denote the average B hadron lifetime by τ and use this lifetime as an approximation for the lifetime of each of the B hadron species. This approximation has little bearing on the results that follow and is easily removed.

The measured proper time $T = l/\gamma\beta c$ requires a measurement of the B hadron decay length l using vertex detector tracking and a measurement of the B energy $\gamma m_B c^2$ using the entire detector including the electromagnetic calorimeter. If we assume gaussian distributions for the decay length error σ_l and the fractional boost error $\sigma_{\gamma\beta}$, then it follows that the proper time resolution $\sigma(t)$ is given by:

$$\sigma^2(t) = \left(\frac{\sigma_l}{\gamma\beta c}\right)^2 + (t\sigma_{\gamma\beta})^2. \tag{6}$$

The following section describes a maximum likelihood analysis to extract X_s. The last section outlines a fast Monte Carlo study to investigate the efficiency and purity of the signal for the SLD experiment.

II. Likelihood Analysis to Extract X_s

We form the log likelihood for a sample of N events:

$$\ln \mathcal{L}(X_s, ...) = \sum_{i=1}^{N} \ln \mathcal{P}_i(X_s, ..., T_i, q_i, ...). \tag{7}$$

$\mathcal{P}_i(X_s, ..., T_i, q_i, ...)$ is the probability that the event comes from a distribution with mixing parameter X_s, measured proper time T_i, and lepton charge q_i. The notation (...) following X_s refers to all additional parameters common to the entire event sample that are allowed to float in the fit, and the (...) following q_i refers to all additional parameters used to construct $\mathcal{P}_i$ that are not allowed to float in the fit.

We construct a separate normalized probability for the signal ($\mathcal{P}_i^{B\to l}$), the $B \to D \to$lepton background ($\mathcal{P}_i^{B\to D\to l}$), and all remaining backgrounds that are not $B \to D \to$lepton ($\mathcal{P}_i^{bkd}$). If we let c_1 denote the fraction of $B \to D \to$lepton background and c_2 the fraction of remaining backgrounds, then the complete probability distribution is given by:

$$\mathcal{P}_i = (1 - c_1 - c_2)\mathcal{P}_i^{B\to l} + c_1\mathcal{P}_i^{B\to D\to l} + c_2\mathcal{P}_i^{bkd}. \tag{8}$$

To construct $\mathcal{P}_i^{B\to l}$ we define an "unsmeared" probability $\eta(X_s, ..., t, q, ...)$ which depends on the true proper time t of the event, unsmeared by the detector resolution:

$$\mathcal{P}_i^{B\to l}(X_s, ..., T_i, ...) = \frac{1}{\mathcal{N}_i}\int_0^\infty \frac{\exp\left[\frac{-(T_i - t)^2}{2\sigma_i^2(t)}\right]}{\sqrt{2\pi\sigma_i^2(t)}}\ \eta_i(X_s, ..., t, q_i, ...)dt. \tag{9}$$

η_i is a sum of the probability for each B species (B_u, B_d, B_s, and B_B =B baryon) in the $Z^\circ \to b\bar{b}$ event sample to survive for time t and then decay to a lepton with charge q_i. $\mathcal{N}$ is a normalization factor.

Let $l^\pm$ denote the tagging lepton and x denote the remainder of the decay products. For $q = \pm 1$ we list the decay channels contributing to η:

$\underline{q = -1}$	$\underline{q = +1}$
(1) $b \to B_u^- \to l^- x$	(1) $\overline{b} \to B_u^+ \to l^+ x$
(2) $b \to B_B^- \to l^- x$	(2) $\overline{b} \to B_B^+ \to l^+ x$
(3) $b \to \overline{B_d^\circ} \to l^- x$	(3) $\overline{b} \to B_d^\circ \to l^+ x$
(4) $b \to \overline{B_s^\circ} \to l^- x$	(4) $\overline{b} \to B_s^\circ \to l^+ x$
(5) $\overline{b} \to B_d^\circ \to \overline{B_d^\circ} \to l^- x$	(5) $b \to \overline{B_d^\circ} \to B_d^\circ \to l^+ x$
(6) $\overline{b} \to B_s^\circ \to \overline{B_s^\circ} \to l^- x$	(6) $b \to \overline{B_s^\circ} \to B_s^\circ \to l^+ x$

Each of the above processes constitutes a term in the sum for η:

$$\eta^{\pm} = \eta(X_s, ..., t, q = \pm 1, ...) = (1) + (2) + (3) + (4) + (5) + (6)\Big|_{q=\pm 1}. \tag{12}$$

Each term (j) in the sum for $\eta^{\pm}$ is a product of probabilities:

$$\Lambda\begin{pmatrix} b \\ \bar{b} \end{pmatrix} \left|\langle b_j|\psi(0)\rangle\langle\psi(0)|\psi(t)\rangle\langle\psi(t)|B^f\rangle\right|^2. \tag{13}$$

The factor Λ is the probability from eq.(4) that the B event originates at the IP from a b or $\bar{b}$ —whichever quark type is relevant for term (j) and charge q.

The factor $\left|\langle b_j|\psi(0)\rangle\right|^2$ is the probability for the b or $\bar{b}$ quark 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

$$\left|\langle\psi(0)|\psi(t)\rangle\langle\psi(t)|B^f\rangle\right|^2 = I(\psi(0), B^f) \tag{14}$$

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 we have:

$$I(B_u, B_u) = I(B_B, B_B) = e^{-t/\tau}, \tag{15}$$

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

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

As an example, term (6) for $q = -1$ is given by:

$$(6)_{q=-1} = \frac{1}{2}\Lambda(\bar{b})A_s e^{-t/\tau}(1 - \cos X_s\frac{t}{\tau}).$$

Summing the six terms and using the relation $A_d + A_s + A_u + A_B = 1$ with the definition $a = (\Lambda(b) - \Lambda(\bar{b}))/2$ yields:

$$\eta = \frac{1}{2}e^{-t/\tau}\{(1 - qa) + qaA_d(1 - \cos X_d\frac{t}{\tau}) + qaA_s(1 - \cos X_s\frac{t}{\tau})\}. \tag{16}$$

We return to eq.(9) and calculate $\mathcal{N}$ to obtain a normalized expression for $\mathcal{P}^{B\to l}$. Formally, we must integrate over all the variables that are not common to the entire event sample. This includes q,P, $z = \cos\theta$, Q^{jet}, A^K, and T. We restrict our event sample to the interval $T_0 < T < T_1$. The lower bound T_0 helps reject background from the IP; it is typically set to 0. The upper bound helps reject outlying events with anomalously large proper times; it is typically set to 10 psec.

$$\mathcal{N}_i = \sum_{q=\pm 1} \sum_{P} \sum_{A^K} \int_{-1}^{1} A(z)dz \int_{-1}^{1} dQ^{jet} \int_{T_0}^{T_1} dT \int_{0}^{+\infty} \frac{\exp\left[\frac{-(T-t)^2}{2\sigma_i^2(t)}\right]}{\sqrt{2\pi\sigma_i^2(t)}}\, \eta_i dt \tag{17}$$

The acceptance $A(z)$ is taken to be a normalized, even function of z. From eq.(16) we see by inspection that the sum over $q = \pm 1$ leaves an expression that depends only on T:

$$\mathcal{N}_i = \text{const} \times \int_{T_0}^{T_1} dT \int_{0}^{\infty} \frac{\exp\left[\frac{-(T-t)^2}{2\sigma_i^2(t)}\right]}{\sqrt{2\pi\sigma_i^2(t)}}\; e^{-t/\tau} \tag{18}$$

$$= \text{const} \times \frac{1}{2} \int_{0}^{\infty} e^{-t/\tau} \left\{ \text{erfc}\left[\frac{T_0 - t}{\sqrt{2}\sigma_i(t)}\right] - \text{erfc}\left[\frac{T_1 - t}{\sqrt{2}\sigma_i(t)}\right] \right\} dt.$$

To construct the $B \to D \to$lepton background distribution $\mathcal{P}^{B\to D\to l}$ we define η^{BDl} as a sum over $B \to D \to l^{\pm}x$ for different B species in exactly the manner used to define η for the signal events. The expression for η^{BDl} is identical to η derived for the signal with q replaced by $-q$. Unlike the $B \to$lepton signal, the proper time residual for $B \to D \to$lepton is poorly described by a gaussian because the lepton vertex is displaced from the B vertex by the D decay length. Monte Carlo studies show that the proper time residual distribution $R^{BDl}(T-t)$ is well modeled by the sum of a gaussian and an exponential:

$$R^{BDl}(T-t) = F_{BDl} \frac{\exp\left[\frac{-(T-t-X_{BDl})^2}{2\sigma_{BDl}^2}\right]}{\sqrt{2\pi\sigma_{BDl}^2}} + (1 - F_{BDl}) \frac{e^{-(T-t)/\tau_{BDl}}}{\tau_{BDl}}, \tag{19}$$

where the parameters F_{BDl}, X_{BDl}, σ_{BDl}, and τ_{BDl} may be estimated from Monte Carlo. The distribution $\mathcal{P}^{B\to D\to l}$ is then given by:

$$\mathcal{P}^{B\to D\to l} = \frac{1}{\mathcal{N}^{BDl}} \int_{0}^{\infty} R^{BDl}(T-t)\; \eta^{BDl} dt, \tag{20}$$

with the normalization factor:

$$\mathcal{N}^{BDl} = \int_{0}^{\infty} dt \left\{ \frac{F_{BDl}}{2} e^{-t/\tau} \left(\text{erfc}\left[\frac{T_0 - t - X_{BDl}}{\sqrt{2}\sigma_{BDl}}\right] - \text{erfc}\left[\frac{T_1 - t - X_{BDl}}{\sqrt{2}\sigma_{BDl}}\right] \right) \right. \tag{21}$$

$$\left. + (1 - F_{BDl}) \exp\left(-t \left[\frac{\tau_{BDl} - \tau}{\tau \cdot \tau_{BDl}}\right] \right) \left(e^{-T_0/\tau} - e^{-T_1/\tau} \right) \right\}.$$

The distribution $\mathcal{P}^{bkd}$ from background events that are not $B \to D \to$lepton is modeled using a fit to the proper time distribution taken from Monte Carlo. Most of these events are due to lepton misidentification. The proper time distribution is peaked at the

IP with a long positive tail; it is well modeled by the sum of a truncated gaussian and an exponential:

$$\mathcal{P}^{bkd} = \frac{F_{bkd} \exp\left[\frac{-(T-X_{bkd})^2}{2\sigma_{bkd}^2}\right]}{\sqrt{\pi/2}\sigma_{bkd}\left[\mathrm{erfc}\left(\frac{T_0 - X_{bkd}}{\sqrt{2}\sigma_{bkd}}\right) - \mathrm{erfc}\left(\frac{T_1 - X_{bkd}}{\sqrt{2}\sigma_{bkd}}\right)\right]} + \frac{(1-F_{bkd})e^{-T/\tau_{bkd}}}{\left\{\tau_{bkd}\left(e^{-T_0/\tau_{bkd}} - e^{-T_1/\tau_{bkd}}\right)\right\}}, \tag{22}$$

where, as in the case of the $B \to D \to$lepton background, the parameters F_{bkd}, X_{bkd}, σ_{bkd}, and τ_{bkd} are estimated from Monte Carlo.

We have completed constructing the probability distribution of eq.(8) in terms of integrals $\int_0^\infty dt$ over the true B decay time t. These integrals are calculated numerically. Because of the prohibitive CPU time required to calculate the integrals for each event, lookup tables with the values of the integrals for a grid of parameters are computed in advance. Four lookup tables are generated:

(1) The integral:

$$\int_0^\infty \frac{\exp\left[\frac{-(T-t)^2}{2\sigma_i^2(t)}\right]}{\sqrt{2\pi\sigma_i^2(t)}} \, e^{-t/\tau} \, \cos\left(X\frac{t}{\tau}\right) \, dt$$

is computed on a 3-dimensional grid of T, X, and $\sigma_l/(\gamma\beta c)$. This table is used to evaluate $\mathcal{N} \cdot \mathcal{P}^{B\to l}$.

(2) The integral from eq.(18) is computed on a 2-dimensional grid of T_0 and $\sigma_l/(\gamma\beta c)$ to provide a table for computing the normalization $\mathcal{N}$ for $\mathcal{P}^{B\to l}$.

(3) The integral:

$$\int_0^\infty R^{BDl}(t-T) \, e^{-t/\tau} \, \cos\left(X\frac{t}{\tau}\right) \, dt$$

is computed on a 2-dimensional grid of T and X.

This table is used to evaluate $\mathcal{N}^{BDl} \cdot \mathcal{P}^{B\to D\to l}$ for the $B \to D \to$lepton background.

(4) The integral from eq.(21) is computed on a 1-dimensional grid of T_0 to provide a table for computing the normalization of the $B \to D \to$lepton background.

The likelihood expression from eq.(7) may now be constructed and minimized with respect to all the parameters, including X_s, that are allowed to vary in the fit. We denote the values of these other parameters at the minimum by the vector $\vec{\alpha}^{min}$. In the limit of infinite statistics, the significance of the fit Σ is given by:

$$\frac{\Sigma^2}{2} = \ln\mathcal{L}(X_s = X_s^{min}, \vec{\alpha}^{min}) - \ln\mathcal{L}(X_s = \infty, \vec{\alpha}^{min}). \tag{23}$$

If X_s^{min} is suitably large ($X_s > 5$), the apparent dilution of the analyzing power due to the time integrated B_s° mixing is nearly independent of X_s and Σ may be interpreted as the significance for observing the time dependent oscillation from B_s° mixing.

Any real experiment within the next several years will have limited sensitivity to much of the range of X_s allowed by a complex component of the CKM matrix element V_{td} ($10 < X_s < 25$). For such an experiment the quantity Σ may no longer be reliably interpreted as the significance of the oscillation signal; the fast oscillatory nature of the signal makes the probability distribution of eq.(8) a complicated function prone to erratic behavior as the statistical power of the data set becomes marginal. It is then no longer possible to determine from the shape of the likelihood function alone whether an observed signal is real. In particular, one must still quantify the probability that an observed signal characterized by Σ could originate from a parent distribution with X_s well outside of the range where the experiment is sensitive. This "fake rate" probability may only be estimated by generating a large ensemble of Monte Carlo experiments under conditions identical to the data set.

III. A Sample Study: One Million Z° Events in the SLD Detector

We have a run fast Monte Carlo to generate an ensemble of experiments under conditions likely to be encountered by the SLD experiment with the upgraded vertex detector. The algorithm to find the proper time is described in Appendix A. We first run the full SLD Monte Carlo to generate distributions for the the vertex efficiency as a function of $\cos\theta$, the boost and decay length residuals for $B \rightarrow$lepton events, the proper time residual for $B \rightarrow D \rightarrow$lepton background events, the proper time distribution for all remaining background events, and the distributions of Q^{jet} (see eq.(2)) for b and $\bar{b}$ events. In this particular study we make no use of any additional analyzing power from charged kaons identified in the opposite hemisphere. Muon identification to $|\cos\theta| < 0.9$ and electron identification to $|\cos\theta| < 0.77$ is assumed for calculating the vertex efficiency as a function of $\cos\theta$. The average vertex efficiency is found to be 41%. If we assume a lepton identification efficiency of 7.5% for $Z^\circ \rightarrow b\bar{b}$ events, then a tagged semileptonic B decay sample corresponding to one million hadronic Z° events contains 13530 events. Although SLD is only scheduled to accumulate 500,000 hadronic Z° events with the upgraded vertex detector, it is believed at the time of writing that further improvements to the vertexing efficiency from other algorithms besides the one described in the Appendix A will improve the statistical power of the measurement by at least a factor of two.

Background fractions and distributions at the time of writing are only available from a study of Monte Carlo for the previous vertex detector (VXD2); they are used here as a conservative estimate of the backgrounds one expects in the data from the upgraded detector. The background fractions used in the standalone are 3% $B \rightarrow D \rightarrow$lepton and 5% from all remaining backgrounds.

1000 fast Monte Carlo experiments using the distributions from the full Monte Carlo are generated with input X_s values of 6, 8, 10, 12, 14, 16, 18, 20, and 100. The B_s° fraction is taken to be 12%. Figure (1) shows a set of log likelihood curves for nine random experiments generated by the fast Monte Carlo with an input value of 10 for X_s.

For this study all parameters save X_s were held fixed during the likelihood fit; the value of X_s returned from the fit is the minimum of the plot of $-\ln\mathcal{L}(X_s)$ versus X_s for

$0 < X_s < 29$. Figures (2) and (3) show distributions of experiments versus X_s(found) for eight different input values of X_s. All of the experiments histogrammed in Figures (2) and (3) were required to pass a significance cut of $\Sigma > 3$ (see eq.(23)).

Figure (4) is a plot of efficiency versus X_s. To be counted as efficient, an experiment must satisfy $\Sigma > 3$ within 20% on the input X_s value.

Figs (5) and (6) are plots of experiments versus *input* X_s for eight different bins of X_s(found). All of the experiments in these plots were also required to satisfy $\Sigma > 3$. Each bin is two units of X_s wide and centered at the value of X_s(found) given in the title of the plot. Notice that the plots are normalized to the highest point. Qualitatively, these plots peak at the correct input value of X_s and have a long (presumably infinite) low tail at $X_s \gg X_s$(input). The width of these peaks provides an estimate of the error in the X_s measurement; recall the discussion from the previous section where it is asserted that the shape of the log likelihood distribution of the data is not expected to provide a reliable estimate of the error on X_s if the statistics are marginal.

To estimate the probability that the signal is a statistical fluctuation, we generate 1000 experiments at a value of X_s ($X_s = 100$) well beyond the sensitive range of the experiment (taken to be $0 < X_s < 25$) and determine the fraction of these experiments that yield a signal within the sensitive region as a function of the significance (Σ) cut. Fig. (7) shows a plot of this fraction; note that a cut of $\Sigma > 3$ corresponds to a 1.5% probability that the signal is a statistical fluctuation with the true value of X_s beyond the reach of the experiment.

IV. Appendix A: One Method to Measure 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.

IV.1 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 previous vertex detector (VXD2) and about 40% for the planned upgrade (VXD3). All analysis cuts below refer to VXD2 unless otherwise stated.

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:

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

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.

The vertex quality tracks are combined in pairs to create a collection of two prong vertices. The vertex fitting routine 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 $\mathbf{V}^{IPD}$ connecting the IP to the D vertex is projected along the direction $\hat{J}$ of the B jet to obtain the signed distance $\hat{J}\cdot\mathbf{V}^{IPD} = V_\parallel^{jet}$. This distance reflects how far downstream the D hadron is displaced from the IP along the B direction. We require $V_\parallel^{jet} > 500\mu m$ ($250\mu m$ 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 $\mathbf{M}$ returned by the vertex fit:

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

The normalized distance $V_\parallel^{jet}/\delta V_\parallel^{jet}$ 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 $\mathbf{V}^B$ 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 $\hat{D}\cdot\mathbf{V}^B = V_\parallel^D$. This distance may be normalized to the error in the D vertex position projected along the direction of the virtual D track:

$$(\delta V_\parallel^D)^2 = \hat{D}_i\hat{D}_j\mathbf{M}_{ij}.$$

For the present study it is only required that $V_\parallel^D > 0$.

After the vertex cuts are applied to a pure sample of $B \rightarrow$lepton events, about 12% 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 remains small because the D direction is still well determined. The remainder of the fakes contain at least one fragmentation track. The vertex efficiency of VXD2 for $B \rightarrow$lepton events for the Monte Carlo event sample under study (RECON 11 hadronic Z° events) was $26 \pm 1\%$.

The vertex cuts are effective in reducing background in the high p, p_t lepton sample. After vertex cuts were applied event sample purity for $B \rightarrow$lepton events was $91.9 \pm 4.6\%$. This is to be compared with a purity of 70% for the high p, p_t lepton sample before the vertex cuts were applied. Background fractions in the vertexed sample were $2.4 \pm 0.5\%$ $B \rightarrow D \rightarrow$lepton, leaving $5.7 \pm 0.9\%$ of the backgrounds from other sources. These other sources are dominated by lepton misidentification (about 70% of the total) so that to good approximation the non-$B \rightarrow D \rightarrow$lepton background has no forward-backward asymmetry.

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. 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 is the difference $l^{meas} - l^{gen}$ between the decay length measured from the vertexing algorithm and the Monte Carlo generated decay length: is well approximated by a gaussian with $\sigma = 200\mu m$ (Fig. (8)). 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:

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

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

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

IV.2 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:

$$E_B = \frac{M_{Z^\circ}}{2} - E_f^{\pm} - E_f^0,$$

where $E_f^{\pm}$ and E_f^0 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 $E_f^{\pm}$ 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^0 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 $\gamma\beta = E_B/M_B\sqrt{1-(M_B/E_B)^2}$. The fractional residual in the boost $[(\gamma\beta)^{meas} - (\gamma\beta)^{gen}]/(\gamma\beta)^{gen}$ is roughly described by a gaussian with $\sigma = 9\%$ (Figure (8)). The tails of the distribution are strongly correlated with events where the measured B energy E_B is small. A cut requiring $E_B > 30$ GeV/c removes 22% of the events and removes most of these tails, leaving a boost residual well described by a single gaussian with $\sigma = 10\%$.

Combining the decay length and boost measurements, we may plot the residual of the proper time $(l/\gamma\beta c - t^{gen})/\tau$ (Figure (8)). This distribution is roughly estimated by assigning half the events to a gaussian with $\sigma = 7\%$ and the remaining half to a gaussian with $\sigma = 25\%$.

Acknowledgements

This study has greatly benefited from numerous discussions with John Jaros, Art Snyder, Morris Schwartz, Su Dong, and Rafe Schindler.