Search for the Rare Decays $B \to K \ell^+ \ell^-$ and $B \to K^* \ell^+ \ell^-$

The flavor-changing neutral current decays \( B \rightarrow K\ell^+\ell^- \) and \( B \rightarrow K^*(892)\ell^+\ell^- \), where \( \ell^\pm \) is a charged lepton, are highly suppressed in the Standard Model, with branching fractions predicted to be of order \( 10^{-7} \) –
The Standard Model predictions for $B \to K^{(*)}\ell^+\ell^-$ include three main contributions: the electromagnetic (EM) penguin, the $Z$ penguin, and the $W^+W^-$ box diagram. Evidence for the EM penguin amplitude has been obtained from the observation of $B \to K^{*\gamma}$ and inclusive $B \to X_s\gamma$, where $X_s$ is any hadronic system with strangeness $|3, 4|$. Calculations of decay rates for $B \to K^{(*)}\ell^+\ell^-$ based on the Standard Model have significant uncertainties due to strong interactions. For example, Ali et al. [1] predict $B(B \to K^\ell\ell^-) = (0.57_{-0.17}^{+0.10}) \times 10^{-6}$ for both $e\ell^-$ and $\mu\ell^-$ final states, $B(B \to K^*\ell^+\ell^-) = (2.3_{-0.5}^{+0.7}) \times 10^{-6}$, and $B(B \to K^*\mu^+\mu^-) = (1.9_{-0.4}^{+0.5}) \times 10^{-6}$. The contribution of the EM penguin amplitude to $B \to K^\ell\ell^-$ is particularly strong at low values of $m_{\ell^+\ell^-}$, giving a larger rate for $B \to K^*\ell^+\ell^-$ than for $B \to K^\ell\ell^-$. We search for the following decays: $B^+ \to K^+\ell^+\ell^-$, $B^0 \to K^0_S\ell^+\ell^-$, $B^+ \to K^{*+}\ell^+\ell^-$, and $B^0 \to K^{*0}\ell^+\ell^-$, where $K^{*0} \to K^+\pi^-$, $K^{*+} \to K^0_S\pi^+$, $K^0_S \to \pi^+\pi^-$, and $\ell$ is either an $e$ or $\mu$. We also search for the lepton-family-violating decays $B \to K^{(*)}\ell^+\ell^-$. Throughout this paper, charge-conjugate modes are implied.

The data used in the analysis were collected with the BABAR detector at the PEP-II storage ring at the Stanford Linear Accelerator Center during 1999-2000. We analyzed a 20.7 fb$^{-1}$ data sample taken on the $\Upsilon(4S)$ resonance consisting of $(22.7 \pm 0.4) \times 10^6 \Upsilon(4S) \to B\overline{B}$ events.

This search relies primarily on the charged-particle tracking and particle-identification capabilities of the BABAR detector [5]. Charged particle tracking is provided by a five-layer silicon vertex tracker (SVT) and a 40-layer drift chamber (DCH). The DIRC, a Cherenkov ring-imaging particle-identification system, is used for charged hadron identification. Electrons are identified using the electromagnetic calorimeter (EMC), which comprises 6580 thallium-doped CsI crystals. These systems are mounted inside a 1.5 T solenoidal superconducting magnet. Muons are identified in the instrumented flux return (IFR), in which resistive plate chambers are interleaved with the iron plates of the magnet flux return.

We extract the signal using the kinematic variables $m_{ES} = \sqrt{E_0^2 - \sum(p_i^*)^2}$ and $\Delta E = \sum_i \sqrt{m_i^2 + p_i^*^2} - E_i^*$, where $E_0^*$ is the beam energy in the $e^+e^-$ rest (c.m.) frame, $p_i^*$ is the c.m. momentum of daughter particle $i$ in the $B$ meson candidate, and $m_i$ is the mass of particle $i$. For signal events, $m_{ES}$ peaks at the $B$ meson mass with a resolution of about 2.5 MeV/c$^2$ and $\Delta E$ peaks near zero, indicating that the candidate system of particles has total energy consistent with the beam energy in the c.m. frame. To prevent bias in the analysis, we optimized the event-selection criteria using Monte Carlo samples: we did not look at the data in the signal region or in the sidebands that were used to measure the background until these criteria were fixed. Signal efficiencies were determined using the Ali et al. model [1].

We select events that have at least four charged tracks, the ratio $R_2$ of the second and zeroth Fox-Wolfram moments [6] less than 0.5, and two oppositely charged leptons with momentum $p > 0.5$ (1.0) GeV/c for $e (\mu)$ candidates. Electron-positron pairs consistent with photon conversions in the detector material are vetoed. We require charged kaon candidates to be identified as kaons and the charged pion in $K^\ast \to K\pi$ not to be identified as a kaon. For $B \to K^{*\ell^+\ell^-}$, we require the mass of the $K^*$ candidate to be within 75 MeV/c$^2$ of the mean $K^*$ mass. $K^0_S$ candidates are reconstructed from two oppositely charged tracks that form a good vertex displaced from the primary vertex by at least 1 mm.

The decays $B \to J/\psi(\to \ell^+\ell^-)K^{(*)}$ and $B \to \psi(2S)(\to \ell^+\ell^-)K^{(*)}$ have identical topologies to signal events. These backgrounds are suppressed by applying a veto in the $\Delta E$ vs. $m_{\ell^+\ell^-}$ plane (Fig. 1). This veto removes charmonium events not only with reconstructed $m_{\ell^+\ell^-}$ values near the nominal charmonium masses, but also events that lie further away in $m_{\ell^+\ell^-}$ due to photon radiation (more pronounced in electron channels) or track mismeasurement. Removing all of these events simplifies the description of the background shape. Charmonium events can, however, pass this veto if one of the leptons (typically a muon) and the kaon are misidentified as each other. If reassignment of particle types results in a dilepton mass consistent with the $J/\psi$ or $\psi(2S)$ mass, the candidate is vetoed. There is also significant feed-up from $B \to J/\psi K$ and $B \to \psi(2S)K$ into $B \to K^{*\ell^+\ell^-}$, since energy lost due to bremsstrahlung in $B \to J/\psi K$ can be compensated for by including a random pion. If the $K^{*\ell^+\ell^-}$ system in a $B \to K^{*\ell^+\ell^-}$ candidate is kinematically consistent with $B \to J/\psi(\to \ell^+\ell^-\gamma)K$, assuming the photon (which is not directly observed) was radiated along the direction of either lepton, then the candidate is vetoed. Apart from the charmonium vetoes, we analyze the full $m_{\ell^+\ell^-}$ range.

Continuum background from non-resonant $e^+e^- \to q\bar{q}$ production is suppressed using a Fisher discriminant [7], a linear combination of the input variables with optimized coefficients. The variables are $R_2$: $\cos \theta_B$, the cosine of the angle between the $B$ candidate and the beam axis in the c.m. frame; $\cos \theta_T$, the cosine of the angle between the thrust axis of the candidate $B$ meson daughter particles and that of the rest of the particles in the c.m. frame; and $m_{K\ell\ell}$, the invariant mass of the $K$-lepton system, where the lepton is selected according
to its charge relative to the strangeness of the $K^{(*)}$. The variable $m_{K\ell}$ helps discriminate against background from semileptonic $D$ decays, for which $m_{K\ell} < m_D$.

Combinatorial background from $B\bar{B}$ events is suppressed using a signal-to-$B\bar{B}$ likelihood ratio that combines candidate $B$ and dilepton vertex probabilities; the significance of the dilepton separation along the beam direction; $\cos \theta_B$; and the missing energy, $E_{\text{miss}}$, of the event in the c.m. frame. The variable $E_{\text{miss}}$ provides the strongest discrimination against $B\bar{B}$ background, since events with semileptonic decays usually have significant unobserved energy due to neutrinos. For each final state, we select at most one combination of particles per event as a $B$ signal candidate. If multiple candidates occur, we select the candidate with the greatest number of drift chamber and SVT hits on the charged tracks.

We use the known charmonium decays $B \to J/\psi K^{(*)}$ and $B \to \psi(2S)K^{(*)}$ to check the efficiency of our analysis cuts. Figure 2 compares the $\Delta E$ distributions (absolutely normalized) of these charmonium samples in Monte Carlo with data. We find good agreement in both the normalization and the shape.

We extract the signal and background yields in each channel using a two-dimensional extended unbinned maximum likelihood fit in the region defined by $m_{\text{ES}} > 5.2$ GeV/$c^2$ and $|\Delta E| < 0.25$ GeV. The signal shapes, including the effects of radiation on the $\Delta E$ distribution and the correlation between $m_{\text{ES}}$ and $\Delta E$, are obtained by parametrizing the GEANT3 Monte Carlo [8] simulation of the signal. The background is described by a function [9] with two parameters that are determined in our fits to the data. Backgrounds from $B\bar{B}$ that peak in the signal region are suppressed to less than 0.2 events in each mode. Although we allow the signal yield to be negative, we have imposed a lower cut-off such that the total fit function is positive. The fit results are shown in Fig. 3 and summarized in Table I. We observe no significant signals.

To determine 90% C.L. upper limits on the signal yields, we generate and fit a series of toy Monte Carlo samples in which the background probability density function is taken from our fit to the data, but the mean number of signal events is varied. We generate ten thousand samples for each mean value, increasing the mean until 90% of the fits to a set of samples give a signal yield greater than that obtained by fitting the data. To give a measure of the sensitivity of the analysis we list in Table I an effective background yield. This quantity is defined as the square of the error on the signal yield from a fit to a toy Monte Carlo sample drawn from the background probability function, with no signal contribution.

Table I lists the systematic uncertainties from the fit, $(\Delta B/B)^{\text{fit}}$, expressed according to their effect on the limits. The sensitivity of the limits to the values used for signal-shape parameters is determined by performing alternative fits using parameters from the $B \to J/\psi K^{(*)}$ control samples. For modes with electrons, we also varied the fraction of signal events in the tail of the $\Delta E$ distribution. To determine whether a more general background shape would lead to different results, we introduced additional parameters and allowed for a correlation between $m_{\text{ES}}$ and $\Delta E$. This procedure shifted the upper limits by 2% to 5%, depending on the mode. Most of the uncertainty associated with the background shape is in-

![FIG. 1: Charmonium veto in the $\Delta E$ vs. $m_{e^+e^-}$ plane for (a) $B \to K^{(*)}e^+e^-$ and (b) $B \to K^{(*)}\mu^+\mu^-$. Hatched regions are vetoed. The dots correspond to a Monte Carlo simulation of $B \to J/\psi(\to e^+e^-)K$ and $B \to \psi(2S)(\to \ell^+\ell^-)K$. Most signal events would lie in the $\Delta E$ region between the horizontal lines.](image1)

![FIG. 2: Comparison of event yields and $\Delta E$ shapes between data and Monte Carlo for the charmonium control samples. The points with error bars show the data, and the solid histograms show the prediction of the charmonium Monte Carlo. All of the analysis selection criteria have been applied except for the charmonium veto, which is reversed. The large tails in the $e^+e^-$ modes are due to photon radiation. Small shifts between data and Monte Carlo are taken into account as systematic uncertainties on the signal yields.](image2)
TABLE I: Results from the fits to \( B \rightarrow K^{(*)}\ell^+\ell^- \) and \( B \rightarrow K^{(*)}\mu^+\mu^- \) modes. The columns from left to right are fitted signal yield \([10]\); upper limit on the signal yield; the contribution of the background to the error on the signal yield, expressed as an effective background yield (see text); the signal efficiency, \( \epsilon \) (not including the branching fractions for \( K^*, K^0, \) and \( K_S^0 \) decays); the systematic error on the selection efficiency, \( (\Delta B/B)_S \); the systematic error from the fit, \( (\Delta B/B)_F \); the branching fraction central value (\( B \)); and the upper limit on the branching fraction, including systematic errors.

<table>
<thead>
<tr>
<th>Mode</th>
<th>Signal yield</th>
<th>90% C.L.</th>
<th>Effective</th>
<th>( \epsilon )</th>
<th>( (\Delta B/B)_S )</th>
<th>( (\Delta B/B)_F )</th>
<th>( B )</th>
<th>( B/10^{-6} )</th>
<th>90% C.L.</th>
</tr>
</thead>
<tbody>
<tr>
<td>( B^+ \rightarrow K^+\mu^+\mu^- )</td>
<td>-0.2^{+1.0}_{-0.3}</td>
<td>3.1</td>
<td>0.7</td>
<td>17.5</td>
<td>±7.6</td>
<td>±4.0</td>
<td>0.0^{+0.9}_{-0.9}</td>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td>( B^+ \rightarrow K^+\mu^+\mu^- )</td>
<td>-0.3^{+0.9}_{-0.0}</td>
<td>2.6</td>
<td>0.6</td>
<td>10.5</td>
<td>±7.5</td>
<td>±4.0</td>
<td>-0.1^{+0.9}_{-0.9}</td>
<td>1.2</td>
<td></td>
</tr>
<tr>
<td>( B^0 \rightarrow K^0\mu^+\mu^- )</td>
<td>3.8^{+2.1}_{-3.8}</td>
<td>8.8</td>
<td>1.4</td>
<td>10.2</td>
<td>±8.8</td>
<td>±11.9</td>
<td>2.5^{+2.5}_{-2.1}</td>
<td>6.6</td>
<td></td>
</tr>
<tr>
<td>( B^0 \rightarrow K^0\mu^+\mu^- )</td>
<td>-0.3^{+1.7}_{-0.0}</td>
<td>3.5</td>
<td>0.7</td>
<td>8.0</td>
<td>±10.8</td>
<td>±3.0</td>
<td>-0.2^{+1.4}_{-0.9}</td>
<td>3.2</td>
<td></td>
</tr>
<tr>
<td>( B^0 \rightarrow K^0\mu^+\mu^- )</td>
<td>1.1^{+2.7}_{-1.1}</td>
<td>4.2</td>
<td>0.2</td>
<td>15.7</td>
<td>±8.8</td>
<td>±9.5</td>
<td>0.9^{+2.2}_{-0.8}</td>
<td>3.9</td>
<td></td>
</tr>
<tr>
<td>( B^0 \rightarrow K^0\mu^+\mu^- )</td>
<td>0.0^{+1.2}_{-0.0}</td>
<td>2.5</td>
<td>0.1</td>
<td>9.6</td>
<td>±8.8</td>
<td>±3.0</td>
<td>0.0^{+1.6}_{-0.9}</td>
<td>3.7</td>
<td></td>
</tr>
<tr>
<td>( B^+ \rightarrow K^+\mu^+\mu^- )</td>
<td>-0.4^{+1.9}_{-0.0}</td>
<td>3.8</td>
<td>1.6</td>
<td>8.5</td>
<td>±11.0</td>
<td>±5.0</td>
<td>-0.8^{+1.3}_{-0.9}</td>
<td>9.6</td>
<td></td>
</tr>
<tr>
<td>( B^+ \rightarrow K^+\mu^+\mu^- )</td>
<td>1.2^{+2.4}_{-1.2}</td>
<td>4.5</td>
<td>0.3</td>
<td>5.8</td>
<td>±13.0</td>
<td>±7.6</td>
<td>3.9^{+3.3}_{-1.2}</td>
<td>17.3</td>
<td></td>
</tr>
<tr>
<td>( B^+ \rightarrow K^+\mu^+\mu^- )</td>
<td>-0.4^{+1.4}_{-0.0}</td>
<td>2.9</td>
<td>1.3</td>
<td>16.8</td>
<td>±5.7</td>
<td>±4.0</td>
<td>-0.1^{+0.7}_{-0.4}</td>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td>( B^0 \rightarrow K^0\mu^+\mu^- )</td>
<td>1.1^{+3.3}_{-1.1}</td>
<td>5.3</td>
<td>2.7</td>
<td>11.9</td>
<td>±7.1</td>
<td>±10.4</td>
<td>0.6^{+1.8}_{-0.9}</td>
<td>3.3</td>
<td></td>
</tr>
<tr>
<td>( B^0 \rightarrow K^0\mu^+\mu^- )</td>
<td>1.1^{+2.1}_{-0.9}</td>
<td>4.1</td>
<td>0.5</td>
<td>14.6</td>
<td>±7.3</td>
<td>±11.2</td>
<td>0.9^{+1.9}_{-0.8}</td>
<td>4.1</td>
<td></td>
</tr>
<tr>
<td>( B^+ \rightarrow K^+\mu^+\mu^- )</td>
<td>-0.4^{+1.8}_{-0.0}</td>
<td>3.5</td>
<td>1.1</td>
<td>9.3</td>
<td>±9.6</td>
<td>±3.0</td>
<td>-0.8^{+2.8}_{-0.0}</td>
<td>8.0</td>
<td></td>
</tr>
</tbody>
</table>

FIG. 3: Projections from individual maximum likelihood fits onto \( m_{ES} \) for the \( \Delta E \) signal regions: -0.11 < \( \Delta E < 0.05 \) GeV (electrons) and -0.07 < \( \Delta E < 0.05 \) GeV (muons). The dotted lines show the background component, and the solid lines show the sum of background and signal components.

corporated in the statistical error on the yield because the background shape is determined from the fit.

The systematic uncertainties on the efficiency, \( (\Delta B/B)_S \), are listed in Table I and arise from charged-particle tracking (±1.2%/lepton, ±2.0%/track for kaons and pions), the continuum suppression cut (±2.0%), the \( B \bar{B} \) suppression cut (±3.0%), \( K_S^0 \) selection (±4.0%), Monte Carlo signal statistics (±3.0% to ±5.0%), the theoretical model dependence of the efficiency (±4.0% to ±7.0%, depending on the mode), and the number of \( B \bar{B} \) events (±1.6%). The uncertainties on the efficiencies due to model-dependence of form factors are taken to be the full range of variation obtained from different theoretical models [1]. In setting an upper limit, the systematic uncertainties from the efficiency, \( (\Delta B/B)_S \), and from the fit, \( (\Delta B/B)_F \), are added in quadrature, and the limit is increased by this factor.

Table I also includes the results for the lepton-family-violating decays \( B \rightarrow K^{(*)}\ell\mu \), where the signal efficiencies were determined from phase-space Monte Carlo simulations. We observe no evidence for these decays.

We determine the branching fractions \( B(B \rightarrow K\ell^+\ell^-) \) and \( B(B \rightarrow K^*\ell^+\ell^-) \) averaged over both \( B \) meson charge and lepton type (\( e^+e^- \) and \( \mu^+\mu^- \)) by performing a simultaneous maximum likelihood fit to the four contributing channels in each case. In combining the \( B \rightarrow K^*\ell^+\ell^- \) modes, the ratio of branching fractions \( B(B \rightarrow K^*e^+e^-)/B(B \rightarrow K^*\mu^+\mu^-) = 1.2 \) from the model of Ali et al. [1] is used to weight the yield in the muon channel relative to that in the electron channel. The extracted yield corresponds to the electron mode. The combined fits give

\[
B(B \rightarrow K\ell^+\ell^-) = (-0.06^{+0.24}_{-0.08} \pm 0.03) \times 10^{-6},
\]

\[
B(B \rightarrow K^*\ell^+\ell^-) = (0.9^{+1.3}_{-0.9} \pm 0.1) \times 10^{-6},
\]

where the first error is statistical and the second is systematic. We evaluate the upper limits on these combined
modes and obtain
\[
B(B \rightarrow K\ell^+\ell^-) < 0.50 \times 10^{-6} \text{ at 90\% C.L.}
\]
\[
B(B \rightarrow K^*\ell^+\ell^-) < 2.9 \times 10^{-6} \text{ at 90\% C.L.}
\]

These limits represent an improvement over previously published results from CDF [11] and CLEO [12]. The Belle [13] experiment has also recently obtained results on these modes. We see no evidence for a signal, and our limits are close to many of the predictions based on the Standard Model. With the rapidly increasing size of our data sample, we expect to have significantly better sensitivity to these modes in the future.

We are grateful for the excellent luminosity and machine conditions provided by our PEP-II colleagues. The collaborating institutions wish to thank SLAC for its support and kind hospitality. This work is supported by DOE and NSF (USA), NSERC (Canada), IHEP (China), CEA and CNRS-IN2P3 (France), BMBF (Germany), INFN (Italy), NFR (Norway), MIST (Russia), and PPARC (United Kingdom). Individuals have received support from the Swiss NSF, A. P. Sloan Foundation, Research Corporation, and Alexander von Humboldt Foundation.

---

* Also with Università di Perugia, Perugia, Italy
† Also with Università della Basilicata, Potenza, Italy


[9] We parametrize the background shape using
\[
f(m_{ES},\Delta E) = N e^{-\frac{m_{ES}^{2}}{2s}} \sqrt{1 - \frac{m_{ES}^{2}}{s_{b}} e^{-\xi \left(1 - \frac{m_{ES}^{2}}{s_{b}}\right)}}
\]
where $N$ is a normalization factor and $s$ and $\xi$ are free parameters determined from the fit to the data.

[10] Whenever possible, we report two-sided 68\% central confidence intervals. For channels constrained by the requirement that the total fit function be non-negative, we quote a single-sided 68\% confidence interval and set the lower statistical error to zero.

