Search for Decays of $B^0 \to e^+e^-$, $B^0 \to \mu^+\mu^-$, $B^0 \to e^+\mu^-$


10 JUNE 2005

PRL 94, 221803 (2005) PHYSICAL REVIEW LETTERS week ending 10 JUNE 2005

(DABAR Collaboration)
We present a search for the decays $B^0 \to e^+e^-$, $B^0 \to \mu^+\mu^-$, and $B^0 \to e^+e^-$ in data collected at the $Y(4S)$ resonance with the BABAR detector at the SLAC $B$ Factory. Using a data set of $111$ fb$^{-1}$, we find no evidence for a signal in any of the three channels investigated and set the following branching fraction

$PRL\ 94, \ 221803 \ (2005) \ PHYSICAL\ REVIEW\ LETTERS$
In the standard model (SM), rare $B$ decays such as $B^0 \rightarrow \ell^+ \ell^-$, where $\ell$ refers to $e$ or $\mu$, are expected to proceed through diagrams such as those shown in Fig. 1 (charge conjugate processes are included implicitly throughout). These decays are highly suppressed since they involve a $b \rightarrow d$ transition and require an internal quark annihilation within the $B$ meson. In addition, the decays are helicity suppressed by factors of $(m_{\ell}/m_B)^2$, where $m_{\ell}$ is the mass of the lepton $\ell$ and $m_B$ is the mass of the $B$ meson. $B^0$ decays to leptons of two different flavors violate lepton flavor conservation and are therefore forbidden in the SM, although permitted in extensions to the SM with nonzero neutrino mass [1]. The SM expectations are given in Table I.

Since these processes are highly suppressed in the SM, they are potentially sensitive probes of physics beyond the SM. In the minimally supersymmetric standard model (MSSM) the branching fraction for these decays can be enhanced by orders of magnitude [2]. In particular, for MSSM models with modified minimal flavor violation (MFV) and large $\tan \beta$ [3], the branching fraction can be increased by up to 4 orders of magnitude. Experimental bounds can restrict allowed regions of parameter space, specifically the mass of the charged Higgs boson. In models with two Higgs doublets and natural flavor conservation at large $\tan \beta$, an increase in the branching fraction of several orders of magnitude is expected [4]. $B^0 \rightarrow \ell^+ \ell^-$ decays are also allowed in specific models containing leptoquarks [5] and supersymmetric (SUSY) models without $R$ parity [6]. The branching fractions for the flavor violating channels $B^0 \rightarrow \ell^+_i \ell^-_j (i \neq j)$ are expected to be exceedingly small but can be enhanced by leptoquarks or $R$ parity violating operators in SUSY models.

To date, $B^0 \rightarrow \ell^+ \ell^-$ decays have not been observed. As shown in Table I, experimental limits are approaching a level of sensitivity that will restrict the allowed parameter space of models that produce $B^0 \rightarrow \ell^+ \ell^-$ branching fraction enhancements of a few orders of magnitude with respect to the SM rates.

The data used in these analyses were collected with the \textit{BABAR} detector at the PEP-II $e^+e^-$ storage ring and correspond to an integrated luminosity of 111 fb$^{-1}$ accumulated at the $\Upsilon(4S)$ resonance (“on resonance”) and 11.9 fb$^{-1}$ accumulated at a center-of-mass (c.m.) energy about 40 MeV below the $\Upsilon(4S)$ resonance (“off resonance”). The latter sample is used for nonresonant $q\bar{q}$ ($q = u, d, s, \text{ and } c$) background studies. The collider is operated with asymmetric beam energies, producing a boost ($\beta \gamma = 0.55$) of the $\Upsilon(4S)$ along the collision axis.

The \textit{BABAR} detector is optimized for the asymmetric beam configuration at PEP-II and is described in detail in [7]. The 1.5-T superconducting solenoidal magnet, whose cylindrical volume is 1.4 m in radius and 3 m long, contains a charged-particle tracking system, a Cherenkov detector dedicated to charged-particle identification, and central and forward electromagnetic CsI calorimeters (EMC). The segmented flux return, including end caps, is instrumented with resistive plate chambers for muon and $K_L^0$ identification.

The presence of two charged high-momentum leptons provides a very clean signature for the three decay modes under consideration. We require two oppositely-charged high-momentum leptons (i.e., $|p_\ell^*| \sim m_\ell/2$ where $p_\ell^*$ is the c.m. momentum of lepton $\ell$) from a common vertex consistent with the decay of a $B^0$ meson. Since the signal events contain two $B^0$ mesons and no additional particles, the total energy of each $B^0$ in the c.m. must be equal to half of the total beam energy. We define

$$m_{ES} = \sqrt{(E_{\text{beam}}^*)^2 - \left(\sum p_i^*\right)^2} \quad (1)$$

$$\Delta E = \sum p_i^* - E_{\text{beam}}^* \quad (2)$$

where $E_{\text{beam}}^*$ is the $(e^+ \text{ or } e^-)$ beam energy in the c.m. frame, $p_i^*$ is the momentum of lepton $i$ in the c.m. frame, and $m_i$ is the mass of lepton $i$. In Eq. (1), $E_{\text{beam}}^*$ is used as opposed to $E_B^0$ because $E_{\text{beam}}^*$ is known with much greater precision. For correctly reconstructed $B^0$ mesons, $m_{ES}$ peaks at the mass of the $B^0$ meson with a resolution of about 2.8 MeV/$c^2$ and $\Delta E$ peaks near zero.

To reduce background from lepton misidentification, we require the leptons to satisfy stringent electron and muon identification criteria [8]. The electron identification efficiency is greater than 93\% with a misidentification rate of less than 0.3\%. The muon identification efficiency ranges from (55–70)\% (depending on run period) with a misidentification rate of 3\%. Electron energy lost through bremsstrahlung is partially recovered by adding the energy of photons that lie within a 3 degree cone about the electron direction.

Suppression of background from nonresonant $q\bar{q}$ production is provided by a series of topological requirements.

\begin{figure}[h]
\centering
\includegraphics[width=0.8\textwidth]{Fig1.png}
\caption{Representative Feynman diagrams for $B^0 \rightarrow \ell^+ \ell^-$ in the standard model.}
\end{figure}
Monte Carlo (MC) for the e background. Sideband data are compared with signal quire particles that form the reconstructed $B^0$ candidate and the thrust axis of the remaining tracks and neutral clusters in the event. In addition, we employ cuts on the invariant mass of the “rest of the event” (ROE) (all tracks not associated with the $B^0$ candidate where all nonleptonic tracks are assumed to be pions) of $m_{\text{ROE}} > 0.5$ GeV and on the second normalized Fox-Wolfram moment of $R_2 < 0.8$ [9]. We also cut on the total multiplicity of both charged tracks and neutral particles by means of the variable $N_{\text{mult}}$ defined as $N_{\text{mult}} = N_{\text{trk}} + N_\gamma / 2$, where $N_{\text{trk}}$ is the total number of tracks in the event and $N_\gamma$ is the number of photons found with an energy $E_\gamma > 80$ MeV. We require $N_{\text{mult}} \geq 5.5$ for the $ee$ and $e\mu$ channels and $N_{\text{mult}} \geq 5.0$ for the $\mu\mu$ channel. This variable is especially useful in the rejection of radiative Bhabha events. We also require that the total energy in the EMC ($E_{\text{EMC}}$) be less than 11 GeV. This cut is effective in reducing background from QED $e^+e^-$ events, including radiative Bhabhas with many conversions.

Four of the selection criteria given above ($|\cos\theta_\gamma| < 0.8$, where $\theta_\gamma$ is the angle in the c.m. frame between the thrust axis of the particles that form the reconstructed $B^0$ candidate and the thrust axis of the remaining tracks and neutral clusters in the event. In addition, we employ cuts on the invariant mass of the “rest of the event” (ROE) (all tracks not associated with the $B^0$ candidate where all nonleptonic tracks are assumed to be pions) of $m_{\text{ROE}} > 0.5$ GeV and on the second normalized Fox-Wolfram moment of $R_2 < 0.8$ [9]). We also cut on the total multiplicity of both charged tracks and neutral particles by means of the variable $N_{\text{mult}}$ defined as $N_{\text{mult}} = N_{\text{trk}} + N_\gamma / 2$, where $N_{\text{trk}}$ is the total number of tracks in the event and $N_\gamma$ is the number of photons found with an energy $E_\gamma > 80$ MeV. We require $N_{\text{mult}} \geq 5.5$ for the $ee$ and $e\mu$ channels and $N_{\text{mult}} \geq 5.0$ for the $\mu\mu$ channel. This variable is especially useful in the rejection of radiative Bhabha events. We also require that the total energy in the EMC ($E_{\text{EMC}}$) be less than 11 GeV. This cut is effective in reducing background from QED $e^+e^-$ events, including radiative Bhabhas with many conversions.

The $B^0 \to \ell^+\ell^-$ candidates are selected by simultaneous requirements on the energy difference $\Delta E$ and the energy-substituted mass $m_{\text{ES}}$. For the $B^0 \to \mu^+\mu^-$ decay mode, the size of this “signal box” is chosen to be $[+2, -2]\sigma$ of the $m_{\text{ES}}$ distribution and $[+2, -2]\sigma$ for the $\Delta E$. In the cases of the $B^0 \to e^+e^-$ and $B^0 \to e^-\mu^+$ decay modes, the signal box sizes in $m_{\text{ES}}$ are also $[+2, -2]\sigma$ but in $\Delta E$ are relaxed to $[+2, -3]\sigma$ and $[+2, -2.5]\sigma$, respectively, to accommodate the tail in the distribution resulting from uncorrected bremsstrahlung and final state radiation. The resolution in $m_{\text{ES}}$ is obtained from a fit to a Gaussian distribution, whereas the resolution in $\Delta E$ is obtained from a fit to an empirical function [10] that gives a good description of this tail.

Comparisons between data and MC indicate that two-photon processes result in substantial electron backgrounds which are not modeled in the generic MC. We thus parameterize the background level in the signal box from the data sidebands with the ARGUS function [11] in $m_{\text{ES}}$ and an exponential function in $\Delta E$. We use these parametrizations to extrapolate the background level found in the sidebands into the signal box. As indicated in Table II, three different sideband boxes are used. The grand sideband boxes are used to estimate the functional form of the $\Delta E$ distributions. The $\chi^2$/d.o.f. for these fits are 24.5/38, 28.6/38, and 27.0/38 for the $ee$, $\mu\mu$, and $e\mu$ channels, respectively. The upper and lower $m_{\text{ES}}$ sideband boxes are used to estimate the functional form of the $m_{\text{ES}}$ distribution. The $\chi^2$/d.o.f. for these fits are 27.1/34, 39.3/34, and 29.8/34 for the $ee$, $\mu\mu$, and $e\mu$ channels, respectively. The remaining background is dominated by pairs of real leptons ($e^+e^-, \mu^+\mu^-, e^+\mu^-$) from $c\bar{c}$ decay, resulting in a proportionately larger background in the $e\mu$ channel (appearing almost entirely in the lower $m_{\text{ES}}$ sideband box). Peaking backgrounds from misidentified two-body

<table>
<thead>
<tr>
<th>Decays</th>
<th>$e^+e^-$</th>
<th>$\mu^+\mu^-$</th>
<th>$e^+\mu^-$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SM</td>
<td>$2.4 \times 10^{-15}$</td>
<td>$8.3 \times 10^{-7}$</td>
<td>$1.9 \times 10^{-7}$</td>
</tr>
<tr>
<td>CLEO</td>
<td>$9.1 \times 10^{-7}$</td>
<td>$6.1 \times 10^{-7}$</td>
<td>$1.5 \times 10^{-7}$</td>
</tr>
<tr>
<td>Belle</td>
<td>$78 \times 10^{-1}$</td>
<td>$1.5 \times 10^{-7}$</td>
<td>$1.7 \times 10^{-7}$</td>
</tr>
<tr>
<td>CDF</td>
<td>$0.17 \times 10^{-1}$</td>
<td>$1.7 \times 10^{-7}$</td>
<td>$1.7 \times 10^{-7}$</td>
</tr>
</tbody>
</table>

![Image](https://via.placeholder.com/150)

FIG. 2 (color online). Distributions of signal MC (hatched) and sideband data (points) for the $e^+e^-$ channel after the initial selection cuts for (a) $|\cos\theta_\gamma|$, (b) $m_{\text{ROE}}$, (c) $R_2$, and (d) $N_{\text{mult}}$. Arrows indicate final cut values. All distributions are normalized to unit area.

which are not modeled in the generic MC. We thus parameterize the background level in the signal box from the data sidebands with the ARGUS function [11] in $m_{\text{ES}}$ and an exponential function in $\Delta E$. We use these parametrizations to extrapolate the background level found in the sidebands into the signal box. As indicated in Table II, three different sideband boxes are used. The grand sideband boxes are used to estimate the functional form of the $\Delta E$ distributions. The $\chi^2$/d.o.f. for these fits are 24.5/38, 28.6/38, and 27.0/38 for the $ee$, $\mu\mu$, and $e\mu$ channels, respectively. The upper and lower $m_{\text{ES}}$ sideband boxes are used to estimate the functional form of the $m_{\text{ES}}$ distribution. The $\chi^2$/d.o.f. for these fits are 27.1/34, 39.3/34, and 29.8/34 for the $ee$, $\mu\mu$, and $e\mu$ channels, respectively. The remaining background is dominated by pairs of real leptons ($e^+e^-, \mu^+\mu^-, e^+\mu^-$) from $c\bar{c}$ decay, resulting in a proportionately larger background in the $e\mu$ channel (appearing almost entirely in the lower $m_{\text{ES}}$ sideband box). Peaking backgrounds from misidentified two-body

<table>
<thead>
<tr>
<th>Sideband box</th>
<th>$\Delta E$ (GeV)</th>
<th>$m_{\text{ES}}$ (GeV/TeV$^2$)</th>
<th>$ee$</th>
<th>$\mu\mu$</th>
<th>$e\mu$</th>
<th># events</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grand</td>
<td>$(-0.40, 0.40)$</td>
<td>$(5.20, 5.26)$</td>
<td>46</td>
<td>56</td>
<td>57</td>
<td></td>
</tr>
<tr>
<td>Upper $m_{\text{ES}}$</td>
<td>$(0.20, 0.50)$</td>
<td>$(5.20, 5.29)$</td>
<td>12</td>
<td>13</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>Lower $m_{\text{ES}}$</td>
<td>$(-0.50, -0.20)$</td>
<td>$(5.20, 5.29)$</td>
<td>33</td>
<td>40</td>
<td>52</td>
<td></td>
</tr>
</tbody>
</table>
TABLE III. Summary of the analyses where $N_{\text{obs}}$ and $N_{\text{exp}}$ are the observed and expected numbers of events in the signal box, $e$ is the efficiency, and $B_{\text{UL}}(B^0 \rightarrow \ell^+ \ell^-)$ is the upper limit on the branching fraction at the 90% C.L. Uncertainties on $N_{\text{exp}}$ and $e$ are statistical and systematic added in quadrature.

<table>
<thead>
<tr>
<th>Decay mode</th>
<th>$N_{\text{obs}}$</th>
<th>$N_{\text{exp}}$</th>
<th>$e[%]$</th>
<th>$B_{\text{UL}}(B^0 \rightarrow \ell^+ \ell^-)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$e^+e^-$</td>
<td>0</td>
<td>0.71 ± 0.31</td>
<td>21.8 ± 1.2</td>
<td>6.1 × 10^{-8}</td>
</tr>
<tr>
<td>$\mu^+\mu^-$</td>
<td>0</td>
<td>0.72 ± 0.26</td>
<td>15.9 ± 1.1</td>
<td>8.3 × 10^{-8}</td>
</tr>
<tr>
<td>$e^+\mu^-$</td>
<td>2</td>
<td>1.29 ± 0.44</td>
<td>18.1 ± 1.2</td>
<td>18 × 10^{-8}</td>
</tr>
</tbody>
</table>

$B$ decay modes were estimated using an MC sample equivalent to more than 20 times the data luminosity and found to be negligible. The total background expectations and signal efficiencies are given in Table III.

The systematic uncertainties on the signal efficiency $e$, the number of $B^0$ mesons produced in the data, and the background estimate are incorporated into the determination of the upper limit on $B(B^0 \rightarrow \ell^+ \ell^-)$. Since the signal efficiency is determined from MC simulation only, differences between data and the simulation would result in an error in our normalization. To estimate this uncertainty we perform comparisons of data and MC using high statistics control samples that have similar characteristics to our $B^0 \rightarrow \ell^+ \ell^-$ signal. The optimal control samples are $B^0 \rightarrow J/\psi K_S^0$, with $J/\psi \rightarrow e^+e^-$ for $B^0 \rightarrow e^+e^-$ and $J/\psi \rightarrow \mu^+\mu^-$ for $B^0 \rightarrow \mu^+\mu^-$, respectively. Since there exists no appropriate control sample for the $e^-\mu^+$ mode, we use the larger of the systematic errors derived from either the $ee$ or $\mu\mu$ modes. In performing these comparisons we found a substantial uncertainty on the signal efficiency to be due to differences between data and the MC simulation in the mean and resolutions of various quantities, depending on the channel. For the electron channels, the dominant quantities are $\Delta E$ and $m_{\text{ROE}}$. Whereas for the muon channels, they are $|\cos\theta_T|$, $N_{\text{mult}}$, and $m_{\text{ROE}}$. When combined with the uncertainties on tracking efficiency of 2.6% and that for particle identification (1.0% per electron, 3.0% per muon), the total systematic uncertainty on the efficiency is estimated to be 5.7%, 7.1%, and 6.8% for the $ee$, $\mu\mu$, and $e\mu$ modes, respectively.

The background estimate is obtained from a fit to sideband data, so the primary uncertainty is due to fluctuations in the fit procedure as events fall in or out of the sideband box. We have studied the stability of the fit and the background estimate when adding or removing events from the $m_{\text{ES}}$ and $\Delta E$ histograms. We find that the fit is unbiased and stable to a level significantly less than the statistical uncertainty on the background estimate.

As shown in Fig. 3 and Table III, when the contents of the signal box were revealed, 0, 0, and 2 events were found in the $ee$, $\mu\mu$, and $e\mu$ channels, respectively. As can be seen in Table III, the numbers of events found in the signal boxes are compatible with the expected background for each mode.

An upper limit (U.L.) on the branching fraction is computed using

$$B(B^0 \rightarrow \ell^+ \ell^-) = \frac{N_{\text{UL}}(N_{\text{obs}})}{(N_{B^0} + N_{\overline{B}^0}) \cdot e},$$

where $N_{\text{UL}}(N_{\text{obs}})$ is the Poisson 90% U.L. on the number of events for $N_{\text{obs}}$ events having been observed, $N_{B^0}(N_{\overline{B}^0})$ is the number of $B^0(\overline{B}^0)$ mesons produced in the data, and $e$ is the signal efficiency. We have $N_{B^0} + N_{\overline{B}^0} = N_{B^0}$ under the assumption of equal production of $B^0\overline{B}^0$ and $B^+\overline{B}^-$ in $Y(4S)$ decays. For our data set, $N_{B^0} = (122.5 \pm 1.2) \times 10^6$.

We follow the technique of [12] in order to account for the presence of background and to include our systematic uncertainties in the determination of the upper limit. As summarized in Table III, the resulting upper limits at the 90% confidence level (C.L.) for $B(B^0 \rightarrow e^+e^-)$, $B(B^0 \rightarrow \mu^+\mu^-)$, and $B(B^0 \rightarrow e^-\mu^+)$ are $6.1 \times 10^{-8}$, $8.3 \times 10^{-8}$, and $18 \times 10^{-8}$, respectively. The corresponding non-background-subtracted upper limits are $8.6 \times 10^{-8}$, $1.2 \times 10^{-7}$, and $2.4 \times 10^{-7}$ respectively.

These bounds are stringent enough to place interesting constraints on popular models. For example, for the MSSM (MFV) models, the relation between $B(B^0 \rightarrow \mu^+\mu^-)$ and the mass of the charged Higgs boson $m_H$ is given as a function of $\tan\beta$ in [7]. We find that for $\tan\beta = 60$, $m_H > 138$ GeV/$c^2$ (90% C.L.).

We are grateful for the excellent luminosity and machine conditions provided by our PEP-II colleagues, and for the substantial dedicated effort from the computing organizations that support BABAR. 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 and DFG (Germany), INFN (Italy), FOM (The Netherlands), and particle centers, universities, and other institutions.
Netherlands), NFR (Norway), MIST (Russia), and PPARC (United Kingdom). Individuals have received support from CONACyT (Mexico), A. P. Sloan Foundation, Research Corporation, and Alexander von Humboldt Foundation.

*Now at Department of Physics, University of Warwick, Coventry, United Kingdom.
†Also with Università della Basilicata, Potenza, Italy.
‡Also with IFIC, Instituto de Física Corpuscular, CSIC-Universidad de Valencia, Valencia, Spain.
§Deceased.

[10] BABAR defines this empirical function to be $f(E) = A \cdot \exp[-\frac{1}{2} \left( \frac{\log(1 - \tau(E - \nu))}{\sigma} \right)^2 + \tau^2]$, where $\tau$ is the “tail parameter” (describing how much is contained in the tail), $\sigma$ is the width, and $\nu$ is the peak position.