SLAC-R-558
UC-414

Measurement of inclusive b semileptonic branching fractions at the Z resonance [1]

Aurel Trandafir

Stanford Linear Accelerator Center
Stanford University
Stanford, CA 94309

SLAC-Report-558
May 2000

Prepared for the Department of Energy
under contract number DE-AC03-76SF00515

[1] PhD Thesis, University of Massachusetts Amherst

Acknowledgments

SLD is a magnificent piece of craft, and I wish to thank all men and women who participated in its designing and manufacturing.

Regarding this analysis, I wish to thank Su Dong for bringing the topic to my attention, for his availability in spite of all the workload, and for the countless contributions to the analysis.

I especially wish to thank Dave Muller for the help with numerous details in the applications of the CRID system in the physics studies, and for thoroughly revising the entire chapters on the analysis method and on the electron and muon identification, which resulted in a number of significant corrections to the previous calculations.

Very special thanks go to Chris Damerell, for reviewing the entire document, and for the great experience while working with him on the VXD3.

I wish to thank Professor John Donoghue for useful inputs in phenomenology and theory, and other members of the dissertation committee: Professor Monroe Rabin, Professor Charles Weems, Professor Richard Kofler, and Professor Stephane Willocq, for thoroughly reading the manuscript, and for a number of truly useful suggestions that improved both its content and the form.

I also want to thank Achim Wiedemann for the help with the details of the luminosity measurement, and Danning Dong for supplying me with ready and easy to use fragmentation functions, which were essential in the systematic studies of this analysis. Some cross-checks in the lepton identification studies were done using τ event data that I got from Timothy Barklow, and I want to thank him as well.

Very special thanks go to Ray Cowan for the help with the LaTeX issues, and for all his availability and time spent formating this document.

Many thanks go to Mike Huffer, from whom I adopted philosophy of dealing with programming tasks in hierarchical and object oriented manner.

Finally, I wish to thank some members of the SLD computing guru: Joe Perl, Garry Bower and Tony Johnson, who were always available when I had to get things done.

Abstract

This document presents a new measurement of inclusive b semileptonic branching fractions $\mathcal{B}(b\to l)$ and $\mathcal{B}(b\to c\to l)$. The $b\to l$ and $b\to c\to l$ are separated by a means that uses correlation between the final state lepton charge and that of its parent b quark as a constraint. Monte Carlo counts of electrons and muons are calibrated to the data using a newly developed technique based on pairs of mutually independent tests for each particle hypothesis separately. The data sample consists of about $550,000$ hadronic Z decays collected at the SLD between 1993 and 1998. Upon analysis of electron and muon counts in 61602 hadronic event hemispheres tagged as containing either a b or a $\bar{b}$, we report

$$\begin{aligned}
\mathcal{B}(b\to e) &= 0.0949 \pm 0.0049 \pm 0.0050 \\
\mathcal{B}(b\to \mu) &= 0.1066 \pm 0.0038 \pm 0.0049 \\
\text{combined } \mathcal{B}(b\to l) &= 0.1015 \pm 0.0030 \pm 0.0035 \\
\\
\mathcal{B}(b\to c\to e) &= 0.0811 \pm 0.0053 \pm 0.0030 \\
\mathcal{B}(b\to c\to \mu) &= 0.0717 \pm 0.0045 \pm 0.0024 \\
\text{combined } \mathcal{B}(b\to c\to l) &= 0.0756 \pm 0.0034 \pm 0.0019\,.
\end{aligned}$$

Contents

List of Figures

List of Tables

1
Semileptonic decays of b hadrons

Semileptonic decays of b hadrons represent some of the simplest means for probing the structure of these hadrons. What makes semileptonic decays so special compared to other processes is that leptons do not interact strongly with the rest of the hadron, and can be factored in a form of an accurately know lepton tensor. Moreover, the energy release in b decays is high. As a consequence, the dynamics of these processes is dominated by the space-time separations close to the light cone, which up to the details of the hadronic form-factors can be treated by the formalism similar to that in the studies of deep inelastic scattering of electrons on nuclear targets. Although the rates are dominated by the electroweak amplitude, the QCD corrections are substantial, and it is actually the probe of the QCD sector that attracts the most attention.

The observables of interest are inclusive and exclusive branching fractions, decay shapes, and various hadronic form-factors. Recent theoretical calculations reliably relate these observables to the more fundamental quantities of the Standard Model (SM), like the KM matrix elements $|V_{cb}|$ and $|V_{ub}|$, strong coupling constant, and the quark masses. The $|V_{cb}|$, for example, can be calculated from the inclusive semileptonic branching fraction and the average B lifetime, or from the endpoint values of suitably parameterized hadronic form-factors in $\bar{B}\to D^{*+}l^-\bar{\nu}_l$ and $\bar{B}\to D^{+}l^-\bar{\nu}_l$ decays. A different parameterization of the hadronic form-factors, in terms of the fraction of hadron momentum carried by the b quark, allows for studies of the quark-gluon structure of b hadrons, including their scaling properties, as well as of non-perturbative effects associated with the light degrees of freedom.

As is the case with many other studies, the semileptonic decays cannot be completely separated from other processes in the events, especially not in busy environments like the hadronic Z decays. Depending on the analysis, a whole host of other variables may be involved: various hadronic branching fractions, total open charm multiplicity n_c, integrated mixing χ, lifetimes τ_B, b production quantities like asymmetry A_b, rate R_b, or energy x_B, and a number of others. They appear as factors in probability relations connecting the observables of interest to the kinematical and topological variables in the events, and are either taken from other experiments and theory, or are turned into independent variables and solved for in extended measurements.

If one assumes that the SM description of the semileptonic decays is qualitatively correct, then our measurement becomes an input into the calculations of the $|V_{cb}|$ and hence into analyses of CP violation and further tests of the Standard Model. On the other hand, any statistically significant discrepancy between the theory and the experiment has the potential of nullifying some of the assumptions of the Standard Model. In that regard processes for which discrepancies between the theory and the experiment persist, in spite of more accurate measurements and more detailed calculations, tend to attract some attention.

1.1 Why measure inclusive b semileptonic branching fractions?

An interesting problem in heavy flavor physics, for which a satisfactory answer has been sought for more than half a decade now, is related to the value of the inclusive b semileptonic branching fraction. Inclusive b semileptonic branching fraction is defined as

$$\mathcal{B}(b\to l) = \frac{\Gamma(b\to l)}{\Gamma}\,, \tag{1.1}$$

where b stands for b-hadron admixture at either $\Upsilon(4S)$ ($B^{\pm}/B^0$ admixture) or at the Z ($B^{\pm}/B^0/B^0_s$/b-baryon admixture), and

$$\Gamma = \Gamma(\mathrm{sl}) + \Gamma(\mathrm{had}) + \Gamma(\mathrm{radiative}) \tag{1.2}$$

for the total width of the admixture.

The theoretical lower bound of $\mathcal{B}(b\to l)$ from QCD corrections to the electroweak amplitude has been originally estimated to 12.5% [1], significantly higher than the world average values of around 10.2% measured at $\Upsilon(4S)$ [2], and around 11% measured at the Z [3]. More recent theoretical calculations based on HQET [4] report $10.9\pm1.0\%$ and $12.0\pm1.0\%$, depending on the renormalization scale μ and the m_c/m_b ratio used in the calculations [5, 6].

Given that the semileptonic branching fractions of B_s and Λ_b are also around 10% [7], and that the Λ_b lifetime is somewhat shorter than the lifetimes of other b-hadrons [8, 9], one would expect the $\mathcal{B}(b\to l)$ at the Z to be slightly lower than that at $\Upsilon(4S)$, while from the experiments it appears to be systematically higher. Recently, an underestimated value of $\mathcal{B}(b\to\bar{D}X)$ has been pointed out as a potential source of the discrepancy. This will be discussed in more detail later in the chapter.

Inclusive quantities like $\mathcal{B}(b\to l)$ usually constrain a fairly large number of more fundamental quantities, and can rarely be used alone to extract the value of any of them. For example, in order to get $|V_{cb}|$ from $\mathcal{B}(b\to l)$ one also has to know the average B lifetime τ_B and some theoretical inputs as will be shown later in the chapter. In spite of that, inclusive measurements have some advantages over exclusive ones. Their statistics is high, and the reconstruction of the relevant event attributes usually simple and more reliable compared to exclusive measurements, leading to much more accurate results. In our measurement of $\mathcal{B}(b\to l)$ for example, one does not need to know the vertices of origin of the lepton candidates, which substantially simplifies the analysis.

This document presents a new measurement of $\mathcal{B}(b\to l)$ with the simultaneous measurement of the cascade branching fraction $\mathcal{B}(b\to c\to l)$. In our technique $b\to l^-$ is separated from $b\to c\to l^+$ by making use of the correlation between the final state lepton charge and that of its parent b quark (we will continue to use just $b\to l$ and $b\to c\to l$). The drawback of the technique is that the $\mathcal{B}(b\to l)$ cannot be directly measured because of the backgrounds like $b\to\bar{c}\to l$ and $b\to\tau\to l$ which yield leptons of the same sign as the $b\to l$. The quantities that *are* directly measured instead, are the total *like sign branching fraction*

$$\mathcal{B}_L = \mathcal{B}(b\to l) + \mathcal{B}(b\to\bar{c}\to l) + \mathcal{B}(b\to\tau\to l)\,, \tag{1.3}$$

and the *unlike sign branching fraction*

$$\mathcal{B}_U \equiv \mathcal{B}(b\to c\to l)\,. \tag{1.4}$$

The $\mathcal{B}(b\to l)$ is then obtained from (1.3) and (1.4) by employing two auxiliary relations: $\mathcal{B}(b\to\bar{c}\to l) = r_D\,\mathcal{B}(b\to c\to l)$ and $\mathcal{B}(b\to\tau\to l) = \alpha\,\mathcal{B}(b\to l)$, where r_D is obtained from experiments on charm multiplicities in b decays, while α is taken from the theory.

1.2 Theoretical predictions

At the tree level, the matrix element for the b quark semileptonic decay is given by

$$\mathcal{M} = -\frac{g^2}{2}\,V_{cb}\left[\bar{u}(p_3)\gamma_\mu\frac{1-\gamma^5}{2}u(p_1)\right]\left[\frac{i(-g^{\mu\nu}+q^\mu q^\nu/m_W^2)}{q^2-m_W^2}\right]\left[\bar{u}(p_4)\gamma_\nu\frac{1-\gamma^5}{2}v(p_2)\right]\,,$$

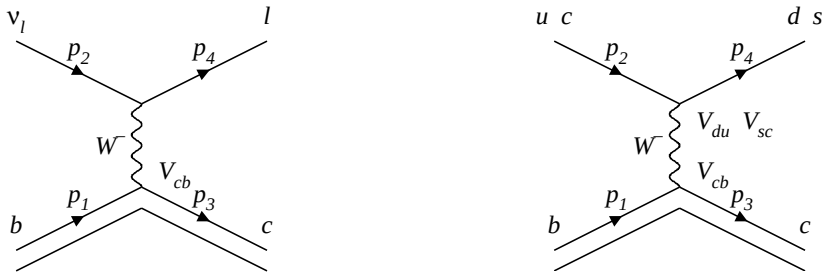


Figure 1.1: Matrix elements for the dominant semileptonic and hadronic b decay modes at the tree level.

where $q = p_2 - p_4$ is the momentum transfer, [14, 15] and Fig. 1.1. The differential decay rate is given by

$$d\Gamma = (2\pi)^4\,\delta^4(p_3 + p_4 - p_1 - p_2)\,\frac{|\,\mathcal{M}\,|^2}{2E_1}\,\prod_{i=2}^{4}\frac{d^3p_i}{(2\pi)^3 2E_i}\,. \tag{1.5}$$

In the limit $q^2 \ll m_W^2$ and only m_b and m_c being nonzero, one gets, upon integrating (1.5), the well know expression for the total decay rate [15]

$$\Gamma = \frac{G_F^2\, m_b^5}{192\,\pi^3}|\,V_{cb}\,|^2\,(1 - 8x^2 + 8x^6 - x^8 - 24x^4\ln x)\,, \qquad x = \frac{m_c}{m_b}\,,$$

where $G_F = g^2\sqrt{2}/(8\,m_W^2)$ defines the Fermi coupling constant.

The $\mathcal{B}(b{\to}l)$ can be roughly estimated by assuming that the matrix elements for $l\bar{\nu}_l$, $d\bar{u}$, and $s\bar{c}$ modes are the same (when $|V_{ud}|$ and $|V_{cs}|$ are taken to be ≈ 1), by taking into account 3 quark colors for each hadronic mode, and by taking into account that the $s\bar{c}$ and $\tau\bar{\nu}_\tau$ rates are phase space suppressed to about 20% of the light mode rates. This gives [15]

$$\mathcal{B}(b{\to}l) \approx \frac{1}{2 + 0.2 + 3\times(1 + 0.2)} \approx 17\%\,, \tag{1.6}$$

in rough agreement with the experimental value of around $10 - 11\%$.

The lowest order radiative gluon corrections to these processes are easy to calculate in the limit of only m_b being nonzero. Guberina, Peccei, and Rückl [16] go

to the bottom of it, and calculate gluon corrections to α_s order that include one-loop corrections to the propagators and the vertices, and gluon radiation corrections (bremsstrahlung). In that limit, the infrared and the mass singularities nicely cancel to give

$$\Gamma(b\to l) = \frac{G_F^2\, m_b^5}{192\,\pi^3}\,\left\{1 - \frac{2\alpha_s}{3\pi}\left(\pi^2 - \frac{25}{4}\right)\right\}\,,\quad \Gamma(\text{had}) = \frac{G_F^2\, m_b^5}{192\,\pi^3}\,\left\{1 - \frac{2\alpha_s}{3\pi}\left(\pi^2 - \frac{31}{4}\right)\right\}\,,$$

where $\Gamma(\text{had})$ is per color. A closer look at the two expressions reveals that $\Gamma(b\to l)$ is reduced slightly more than $\Gamma(\text{had})$ compared to their respective values at the tree level. For $\alpha_s \approx 0.12$ the $\mathcal{B}(b\to l)$ is reduced compared to the value in (1.6) by about 10%.

More detailed calculations of $\mathcal{B}(b\to l)$ have been made by taking into account the c and the τ masses. Most of these calculations focus on estimating the size of $\Gamma(\text{had})$. Bagan, Ball, Braun, and Gosdzinsky [5] calculate the leading order radiative corrections to the nonleptonic inclusive b decay modes $b\to c\bar{u}d$, $b\to c\bar{c}s$ and $b\to c\tau\bar{\nu}$ by taking the c and the τ masses into account, and by neglecting the s-quark mass. They conclude that the ratio of semileptonic to non-leptonic decays into a single final charm state remains practically unchanged, while the $\mathcal{B}(b\to l)$ reduces the previous benchmark value of 12.5% [1] by about 1%. They report

$$\mathcal{B}(b\to l) = (11.6 \pm 1.8)\%\,, \tag{1.7}$$

where the error comes from uncertainties in various inputs: the renormalization scale μ, HQET parameter λ_1, and strong coupling constant $\alpha_s(m_Z)$. These calculations do not take into account the radiative processes.

The same group of authors calculates the next-to-leading order radiative corrections in $b\to c\bar{c}s$ and report

$$\mathcal{B}(b\to l) = (12.0 \pm 0.9^{+0.9}_{-1.3})\% \qquad \text{and} \qquad \mathcal{B}(b\to l) = (11.2 \pm 1.0^{+1.0}_{-2.2})\%\,, \tag{1.8}$$

for two different schemes in which pole quark masses and running $\overline{\text{MS}}$ masses are respectively used [5].

In an independent numerical analysis with the same theoretical inputs as in [5], Neubert and Sachrajda obtain [6]

$$\mathcal{B}(b\to l) = \begin{cases} 12.0 \pm 1.0\%; & \mu = m_b\,, \\ 10.9 \pm 1.0\%; & \mu = m_b/2\,, \end{cases}$$

$$n_c = \begin{cases} 1.20 \mp 0.06; & \mu = m_b\,, \\ 1.21 \mp 0.06; & \mu = m_b/2\,, \end{cases}$$

where n_c is the total open charm multiplicity (see Fig. 1.3), and where $m_b = (4.8 \pm 0.2)$ GeV, $0.25 < m_c/m_b < 0.33$, and $\alpha_s(m_Z) = 0.117 \pm 0.004$ have been used.

1.3 Phenomenology and related studies

The theoretical uncertainties in the calculations of $\mathcal{B}(b\to l)$ are large. This is partly due to the fact that these calculations are mostly focused on processes like those in Fig. 1.2 a) and b) (which are the same for inclusive decays when the differential rates are integrated over all final state momenta). Contributions from other processes, Fig. 1.2 c) - e), are small but not negligible.

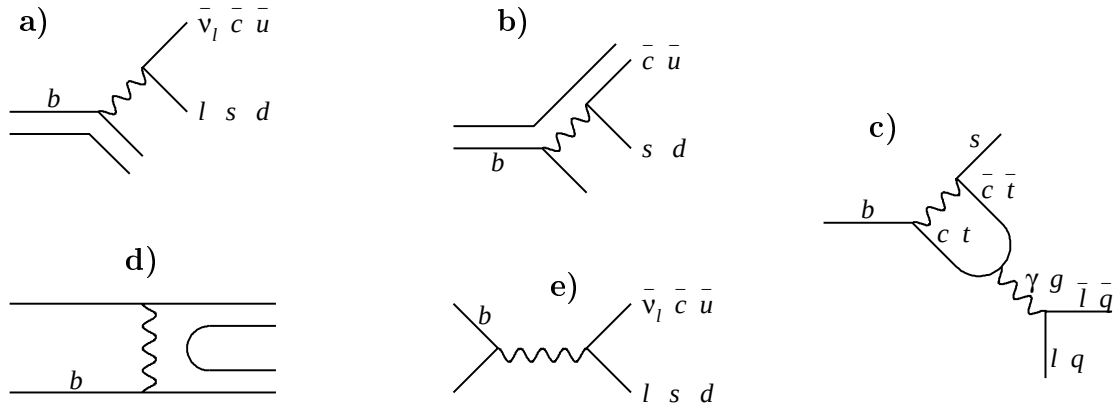


Figure 1.2: Various b decay mechanisms. The theoretical calculations have been so far focused to estimating the size of the radiative corrections to processes in a) and b).

Processes in b) also involve color mixing, and are for the exclusive two body decays suppressed roughly by a factor of 3 in the amplitude (9 in the rate). The W exchange processes in d) and e) are proportional to the square of the amplitude of finding a b quark and its antiquark companion at the same point, which roughly goes

as $|\psi(0)|^2 \ll m_b^3$ in cases where the hadronic radius does not depend strongly on the quark mass [14]. This has to be compared to m_b^5 dependence in the rates of graphs a) and b). Moreover, processes in e) are restricted to only neutral B mesons. Radiative processes c) are KM suppressed. Interference effects between processes a) and b) in charged B mesons also matter and will be mentioned later in this section.

The current discrepancy between the theoretical prediction and the measured value of $\mathcal{B}(b\to l)$ is about 2σ, Fig. 1.3. It is not clear, however, whether the problem has been actually overplayed, given the large uncertainties in the theoretical calculations, and the potential biases in both the theory and the experiment. To get a better overview of the problem, we briefly present here some of the studies that relate the $\mathcal{B}(b\to l)$ to other measurable quantities, like the total open charm multiplicity n_c, ratios of the exclusive lifetimes, and the flavor specific branching fractions into open charm. The overall picture of the b semileptonic decays is therefore a rather patchy one, consisting of pieces of theory and experiment put together by the common sense.

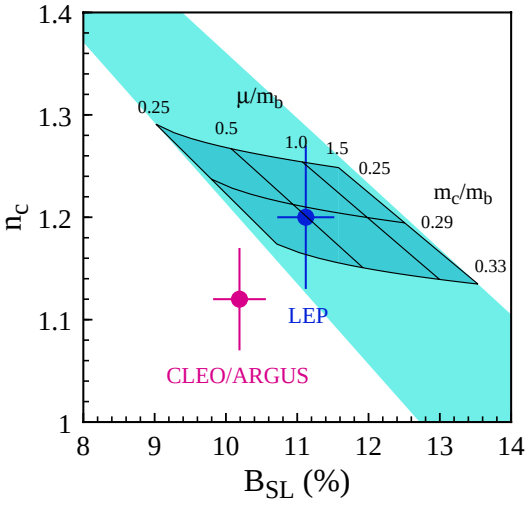


Figure 1.3: This figure, taken from Neubert [17], compares early experimental results for n_c and $\mathcal{B}(b\to l)$ with the newer theoretical calculations of these quantities. Newer L3 and SLD numbers for $\mathcal{B}(b\to l)$ better agree with the CLEO/ARGUS average (see Fig. 6.22 at the end of this document).

It has been speculated for a while [18, 19, 20] that a somewhat smaller mass of the c quark, or the failure of some of the assumptions used in the calculations of the hadron currents (in connection with the Operator Product Expansion (OPE) technique used in these calculations [20, 21]) can be responsible for the increase of the total open charm multiplicity n_c at the expense of $\mathcal{B}(b\to l)$. In order to accommodate $\mathcal{B}(b\to l) \approx 0.104$, the n_c has to be around 1.25, while from the experiments it is more like $n_c = 1.10 \pm 0.05$, Fig. 1.3. From the theory, $\mathcal{B}(b\to l)$ and n_c are mutually constrained based on [5]

$$r_{ud} \equiv \frac{\Gamma(b\to c\bar{u}d')}{\Gamma(b\to ce\bar{\nu})} = 4.0 \pm 0.4\,, \quad d' \equiv V_{ud}d + V_{us}s\,, \tag{1.9}$$

which is considered a reliable quantity [20].

No clear-cut explanation for the discrepancy between the theory and the experiment for the n_c versus $\mathcal{B}(b\to l)$ "puzzle" has been offered so far. However, a very nice by-product of the efforts to explain the "puzzle" is the prediction of the flavor-specific charm branching ratios in b-hadron decays. This turned out to be critical in providing a clear-cut explanation for another discrepancy, one between the values of $\mathcal{B}(b\to l)$ measured at $\Upsilon(4S)$ and the Z.

At the Z, the composition of the b hadron admixture $f_{B^+} : f_{B^0} : f_{B_s^0} : f_{\Lambda_b}$ is roughly $40 : 40 : 12 : 8$, while at $\Upsilon(4S)$, $f_{B^+} : f_{B^0}$ is about $50 : 50$. Ratios of the b hadron lifetimes are also related to $\mathcal{B}(b\to l)$. LEP recently measured [9]

$$\frac{\tau_{B^+}}{\tau_{B^0}} = 1.06 \pm 0.04\,, \quad \frac{\tau_{B_s}}{\tau_{B^0}} = 0.98 \pm 0.05\,, \quad \frac{\tau_{\Lambda_b}}{\tau_{B^0}} = 0.79 \pm 0.06\,. \tag{1.10}$$

Hence

$\mathcal{B}(b\to l)(\text{at the } Z) =$

$$f_{B^+}\,\frac{\Gamma(B^+\to l\overline{\nu}_l X)}{\Gamma(B^+\to \text{all})} \;+\; f_{B^0}\,\frac{\Gamma(B^0\to l\overline{\nu}_l X)}{\Gamma(B^0\to \text{all})} \;+\; f_{B_s}\,\frac{\Gamma(B_s\to l\overline{\nu}_l X)}{\Gamma(B_s\to \text{all})} \;+\; f_{\Lambda_b}\,\frac{\Gamma(\Lambda_b\to l\overline{\nu}_l X)}{\Gamma(\Lambda_b\to \text{all})}\,,$$

where $f_{B^+} = \Gamma(B^+\to\text{all})/\Gamma$, etc. One can similarly write $\mathcal{B}(b\to l)$ at $\Upsilon(4S)$. Assuming the semileptonic decay widths of all b hadrons approximately equal, say to $\Gamma(B^0\to l\overline{\nu}_l X)$, one has

$$\begin{aligned}
\mathcal{B}(b\to l)(\text{at the } Z) &= (f_{B^+}\,\tau_{B^+} + f_{B^0}\,\tau_{B^0} + f_{B_s}\,\tau_{B_s} + f_{\Lambda_b}\,\tau_{\Lambda_b})\;\Gamma(B^0\to l\overline{\nu}_l X) \\
&\approx 1\,\tau_{B^0}\,\Gamma(B^0\to l\overline{\nu}_l X)\,, \\
\mathcal{B}(b\to l)(\text{at } \Upsilon(4S)) &= [f_{B^+}\,\tau_{B^+} + f_{B^0}\,\tau_{B^0}]\;\Gamma(B^0\to l\overline{\nu}_l X) \\
&\approx 1.03\,\tau_{B^0}\,\Gamma(B^0\to l\overline{\nu}_l X)\,.
\end{aligned}$$

The value of the $\mathcal{B}(b\to l)$ at the Z should therefore be around 3% lower than that at $\Upsilon(4S)$, while from the early LEP measurements it appears to be systematically higher.

Upon study of the Dalitz plot distributions of $b\to c\bar{c}s$, Buchalla, Dunietz, and Yamamoto [19] predicted a while ago the size of the "wrong-charm" yield

$$\mathcal{B}(b\to\bar{D}X)\approx 0.2 \qquad \text{for } D \equiv D^+/D^0 \text{ admixture}\,.$$

The prediction has been independently confirmed at about half that level by CLEO [22]:

$$\frac{\Gamma(b\to\bar{D}X)}{\Gamma(b\to DX)} = 0.100\pm 0.026\pm 0.016\,. \tag{1.11}$$

A series of measurements of the ratios of flavor specific charm yields (ratios of upper to lower vertex charm yields) by ALEPH [23], and of the exclusive and inclusive open charm multiplicities by OPAL [24], independently confirm these results. [These studies are nicely summarized in already cited [20] which uses a somewhat older and unpublished ALEPH result $\mathcal{B}(b\to D\bar{D}X)\equiv\mathcal{B}(b\to D^0\bar{D}^0X,\, D^0D^-X,\, D^+\bar{D}^0X) = 0.128\pm 0.027\pm 0.026$.]

Recent LEP measurements take more carefully into account the ratio between the $\mathcal{B}(b\to\bar{c}\to l)$ and $\mathcal{B}(b\to c\to l)$, and report somewhat lower value of $\mathcal{B}(b\to l)$ (see Fig. 6.22 at the end of this document).

Several other explanations for the n_c versus $\mathcal{B}(b\to l)$ discrepancy have been proposed: enhancement of $b\to c\bar{u}d$ due to non-perturbative effects, Eq. (1.9), enhancement of radiative processes like $b\to sg$ and $b\to dg$ due to new physics [7], etc. The rates of the radiative processes, for example, have been measured at SLD and elsewhere, but no substantial deviations from the theoretical predictions have been observed.

An interesting alternative explanation of the discrepancy between the observed and predicted value of the $\mathcal{B}(b\to l)$, that does not affect the n_c, is offered in terms of the interfering amplitudes in charged B mesons, [25] and Fig. 1.4. Indirectly, the constructive interference in charged B mesons can be seen through the larger lifetimes compared to the neutral B mesons, Eq. (1.10). This has been studied to some degree, with the prevailing conclusion that these effects "comprise only a small fraction of the total hadronic rate" (see [7] p. 527 and references therein).

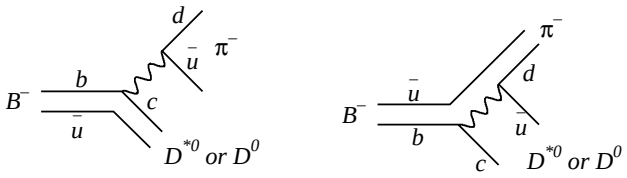


Figure 1.4: Interfering amplitudes in charged B mesons.

1.4 The role of mixing

We first examine the role of the integrated mixing χ in our analysis, then compare it with the role of χ in the seemingly same method that CLEO uses to measure $\mathcal{B}(b\to l)$ [2]. The comparison of the roles of χ in the two methods does not affect our analysis, but is interesting in its own right.

For the b-hadron admixture at the Z, the average mixing into all final states is given by

$$\chi = f_{B_d}\,\chi_{B_d} + f_{B_s}\,\chi_{B_s}\,, \tag{1.12}$$

where f_{B_d} and f_{B_s} are the fractions of B_d and B_s mesons in the admixture. If, however, only leptons from prompt decays are counted in the final state, the χ is given by

$$\chi = f_{B_d}\,\frac{\mathcal{B}(B_d\to l)}{\mathcal{B}(b\to l)}\,\chi_{B_d} + f_{B_s}\,\frac{\mathcal{B}(B_s\to l)}{\mathcal{B}(b\to l)}\,\chi_{B_s}\,. \tag{1.13}$$

The role of this correction, first pointed out by Dunietz [18], is important in CLEO and SLD methods of measuring $\mathcal{B}(b\to l)$. By definition

$$\mathcal{B}(b\to l) = f_{B_d}\mathcal{B}(B_d\to l) + f_{B_u}\mathcal{B}(B_u\to l) + f_{B_s}\mathcal{B}(B_s\to l) + f_{\Lambda_b}\mathcal{B}(\Lambda_b\to l)\,.$$

For a prompt decay of a b quark which *does not* mix one has

$$
\begin{aligned}
& f(b\to b\to l) \\
&= (1-\chi_{B_d}) f_{B_d} \mathcal{B}(B_d\to l) + f_{B_u}\mathcal{B}(B_u\to l) + (1-\chi_{B_s}) f_{B_s}\mathcal{B}(B_s\to l) + f_{\Lambda_b}\mathcal{B}(\Lambda_b\to l) \\
&= \mathcal{B}(b\to l) - [\chi_{B_d} f_{B_d}\mathcal{B}(B_d\to l) + \chi_{B_s} f_{B_s}\mathcal{B}(B_s\to l)] \\
&= (1-\chi)\,\mathcal{B}(b\to l)\,,
\end{aligned}
$$

according to the definition (1.12) of inclusive integrated χ at the Z. Similar relations can be obtained for processes that involve mixing, like $b\to\bar{b}\to l$, and others.

For prompt semileptonic decays all exclusive branching fractions are approximately equal, around 10%. Therefore $\mathcal{B}(B_d\to l)/\mathcal{B}(b\to l) \approx \mathcal{B}(B_s\to l)/\mathcal{B}(b\to l) \approx 1$, and χ calculated using (1.13) approximately equal to χ calculated using (1.12).

For cascade decays, this is somewhat different. Here we use the MC numbers provided to us by Su Dong [29]:

```
            b->c->l+    b->cbar->l-
Bu            7.469       1.718
Bd           10.539       1.719
Bs            8.166       1.794
Baryon        4.966       1.394
<B had>       8.611       1.704
```

from which

$$
\frac{\mathcal{B}(B_d\to D\to l)}{\mathcal{B}(b\to c\to l)} \approx 1.22 \quad \text{and} \quad \frac{\mathcal{B}(B_s\to D\to l)}{\mathcal{B}(b\to c\to l)} \approx 0.95\,,
$$

where $D \equiv D^+/D^0/D_s^+/c$-baryon admixture in b decays. This translates into effective mixing as

```
Effective mixing Chi-bar
Chi-bar B-had    =    0.13000
Chi-bar b->l     =    0.13043
Chi-bar b->c->l+ =    0.14341
```

also copied from [29]. These numbers will be used only as ratios and will be scaled down to the world average value $\chi \approx 0.12$, which corresponds to the value 0.13000 in the table above. The effective mixing for like and unlike sign decays will be denoted χ_L and χ_U respectively. In general discussions we will continue to use one single value χ to keep the notation simple.

Clearly, mixing of neutral B mesons dilutes information about the parent b flavor. If χ were 0, the probability of the parent flavor would be the same as the probability for the initial state flavor. If χ were 0.5 instead, the information about the parent flavor would be completely scrambled, and it would be impossible to calculate the branching fractions using our method. For the b-hadron admixture at the Z, the average mixing is around 0.12, small enough for our method to work.

It is interesting to compare the role of χ at the Z and $\Upsilon(4S)$. At the Z, a $B\bar{B}$ pair is produced incoherently as either

$$\mid B(\mathbf{k_1})\rangle \mid \bar{B}(\mathbf{k_2})\rangle \quad \text{or} \quad \mid \bar{B}(\mathbf{k_1})\rangle \mid B(\mathbf{k_2})\rangle$$

along $\mathbf{k_1}$ and $\mathbf{k_2}$, where the flavor phase is fixed by other degrees of freedom in the event. For example, the rest of the jet with the $b\bar{d}$ bound state will likely carry the d quark while the $\bar{b}d$ jet will likely carry the $\bar{d}$ [30]. For an incoherent $B\bar{B}$ pair the time evolution of one meson is independent of the other [31]. Therefore, if the probability of the initial state B meson is p, probability that the flavor is unchanged at the decay is $p(1-\chi)$.

At $\Upsilon(4S)$ a $B\bar{B}$ pair is created as a single partial wave $\mid l = 1\rangle \mid B\bar{B}\rangle$. Its orbital part is determined by $J = 1$ of the initial $\Upsilon(4S)$, and by $I(J^P) = \frac{1}{2}(0^-)$ of the final state B mesons. Its antisymmetric flavor phase

$$\mid B\bar{B}\rangle = \frac{1}{\sqrt{2}}\left(\mid B(\mathbf{k})\rangle \mid \bar{B}(-\mathbf{k})\rangle - \mid \bar{B}(\mathbf{k})\rangle \mid B(-\mathbf{k})\rangle\right) \tag{1.14}$$

is fixed by the requirement that the overall wave function must be symmetric.

A neutral $B\bar{B}$ therefore oscillates coherently until one of them decays. This means that before one of the two B mesons decays the system can be either in $B\bar{B}$ or $\bar{B}B$ state in the $\Upsilon(4S)$ rest frame. The probability of the final state pair can be expressed as a function of the decay times of the two mesons t_1 and t_2. These derivations are given in a beautifully written review paper on neutral kaons by Radoje Belusevic [31] (see pp. 79 - 81 for the derivations). Here we only present some of the final results.

The amount of mixing is experimentally given by

$$\mathcal{R} = \frac{p(BB) + p(\bar{B}\bar{B})}{p(B\bar{B}) + p(\bar{B}B)} ,$$

where the probabilities are proportional to the counts. For an incoherent pair

$$\mathcal{R} = \frac{2\chi(1-\chi)}{(1-\chi)^2 + \chi^2}$$

($p(BB) = p(\bar{B}\bar{B}) = \chi(1-\chi)$ (*probability that one meson oscillates*) $\times$ (*probability that the other does not oscillate*), etc.), while for a coherent pair [31]

$$\mathcal{R} = \frac{\chi}{1-\chi} .$$

For a coherent pair therefore, $p(BB) + p(\bar{B}\bar{B}) = \chi$, and $p(BB) = p(\bar{B}\bar{B}) = \chi/2$. We introduce indices 1 and 2 for the two B mesons in the pair. Then $p(B_1) = p(B_1B_2) + p(B_1\bar{B}_2) = 1/2$. Conditional probability that the flavor of meson 2 is B when the flavor of meson 1 is also B is $p(B_2 \mid B_1) = p(B_1B_2)/p(B_1) = \chi$. Along the same line of reasoning, $p(\bar{B}_2 \mid B_1) = 1-\chi$, etc. The CLEO method [2] naturally follows:

$$\begin{aligned} p(l^+ \mid l^+) = p(l^- \mid l^-) &= \chi\mathcal{B}(b{\to}l) + (1-\chi)\mathcal{B}(b{\to}c{\to}l) , \\ p(l^- \mid l^+) = p(l^+ \mid l^-) &= (1-\chi)\mathcal{B}(b{\to}l) + \chi\mathcal{B}(b{\to}c{\to}l) , \end{aligned}$$

where the role of $b{\to}l$ and $b{\to}c{\to}l$ is merely to ensure proper charge correlation between the final state leptons. High momentum lepton used to tag one of the two B meson is always correlated to the b quark flavor of that meson. The other final state lepton in the like sign combination has to be from $b{\to}l$ when the system decays as either a BB or a $\bar{B}\bar{B}$, or from $b{\to}c{\to}l$ when the system decays as either a $B\bar{B}$ or a $\bar{B}B$. Similar reasoning applies to unlike sign combination of the final state leptons.

We will see in Ch. 3, that in the limit of precisely known initial state flavor our system of equations has the same form as that of CLEO. The difference between the two is fundamental however: CLEO correlates final state leptons on the tagging and the lepton counting side, while our equations correlate final state leptons and the initial state b quark flavor in each individual hemisphere separately.

1.5 Determination of $|V_{cb}|$ from inclusive b semileptonic decays

The $|V_{cb}|$ is related to the inclusive b semileptonic branching fraction $\mathcal{B}(b\to l)$ and average b lifetime τ_B as

$$\Gamma(b\to l) \equiv \frac{\mathcal{B}(b\to l)}{\tau_B} = \gamma_c |V_{cb}|^2 + \gamma_u |V_{ub}|^2$$

where γ_c and γ_u have to be obtained from the theory [26, 10, 27]. Given $|V_{ub}| \ll |V_{cb}|$ and $\gamma_u \sim \gamma_c$ one approximately has

$$|V_{cb}|^2 \approx \frac{\mathcal{B}(b\to l)}{\gamma_c \tau_B}\,.$$

For a reference value $\mathcal{B}(b\to l) = 0.109$, Ball [27] estimates

$$\left(\frac{\tau_B}{1.5\ \text{ps}}\right)^{1/2} |V_{cb}| = 0.041 \pm 0.002 \pm 0.002$$

where the error on $\mathcal{B}(b\to l)$ is taken from [2]. For a different $\mathcal{B}(b\to l)$ one has

$$|V_{cb}| = (0.041 \pm 0.002 \pm 0.002) \times \left(\frac{1.5\ \text{ps}}{\tau_B}\right)^{1/2} \left(\frac{\mathcal{B}(b\to l)}{0.109}\right)^{1/2} = 0.039 \pm 0.004\,, \tag{1.15}$$

where the final number has been obtained by inserting our preliminary value $\mathcal{B}(b\to l) = 0.1015 \pm 0.0046$, Fig. 6.22, and by using $\tau_B = 1.582 \pm 0.030$ ps from [28].

Clearly, the determination of $|V_{cb}|$ depends on theoretical inputs. For γ_c from [11] for example, one has

$$\begin{aligned} |V_{cb}| = \left(\frac{\mathcal{B}(b\to l)}{\gamma_c \tau_B}\right)^{1/2} &= \left[\frac{0.1015 \pm 0.0046}{(49 \pm 9\ \text{ps}^{-1}) \times (1.582 \pm 0.030\ \text{ps})}\right]^{1/2} \\ &= 0.0361 \pm 0.0035\,. \end{aligned} \tag{1.16}$$

1.6 Measurements of the hadronic form-factors

We conclude this introductory chapter with a brief overview of the measurements of the hadronic form-factors, which have the potential to further illuminate the structure of b hadrons, and can be used to independently measure the $|V_{cb}|$ in an almost model independent way.

The most general matrix element at the tree level for decay of a $|B\rangle$ into $|l\rangle|\bar{\nu}_l\rangle|n\rangle$, $|n\rangle$ being an arbitrary final n hadron state, is given by

$$\mathcal{M}_n = -i\frac{G_F}{\sqrt{2}}\, V_{qb}\, \bar{u}(p_l)\gamma^\mu(1-\gamma^5)v(p_\nu)\, \langle n|j_\mu(x)|B\rangle\,,$$

where

$$j_\mu(x) = \bar{q}(x)\gamma_\mu(1-\gamma^5)b(x)\,.$$

is the weak hadron current. Eq. (1.5) for the decay rate then translates into

$$\begin{aligned} d\Gamma &= \frac{1}{2E_B}\sum_n \mid \mathcal{M}_n \mid^2 \frac{d^3p_l}{(2\pi)^3 2E_l}\,\frac{d^3p_\nu}{(2\pi)^3 2E_\nu}\,\left[\prod_{i=1}^{n}\frac{d^3p_i}{(2\pi)^3 2E_i}\right] \\ &\quad \times(2\pi)^4\,\delta^4(p_B - q - \sum_i p_i)\,, \end{aligned} \tag{1.17}$$

where $q = -(p_l + p_\nu)$ is the momentum transfer, and the sum over n is over all final states containing $n = 1, 2, \ldots$ hadrons.

Upon integrating Eq. (1.17) over final state hadron momenta p_i one gets

$$d\Gamma = \frac{G_F^2|V_{cb}|^2}{(2\pi)^5 E_B}\, L^{\mu\nu}\, W_{\mu\nu}\,\frac{d^3p_l}{2E_l}\,\frac{d^3p_\nu}{2E_\nu}\,,$$

where

$$\begin{aligned} L^{\mu\nu} &= 2(p_l^\mu p_\nu^\nu + p_\nu^\mu p_l^\nu - g^{\mu\nu}p_l\cdot p_\nu + i\varepsilon^{\mu\nu}{}_{\alpha\beta}p_l^\alpha p_\nu^\beta) \\ W^{\mu\nu} &= \sum_n \int \prod_{i=1}^{n}\frac{d^3p_i}{(2\pi)^2 2E_i}(2\pi)^3\delta^4(p_B - q - \sum_{i=1}^{n} p_i)\langle B \mid j_\nu^\dagger(0) \mid n\rangle\langle n \mid j_\mu(0) \mid B\rangle \end{aligned}$$

are the lepton and the hadron tensors respectively.

From the symmetries of the matrix element, the $W^{\mu\nu}$ can be expressed as

$$W^{\mu\nu} = W_1 p_\mu p_\nu + W_2 p_\mu q_\nu + W_3 q_\mu p_\nu + W_4 \varepsilon_{\mu\nu\alpha\beta} p_\alpha q_\beta + W_5 g_{\mu\nu} \,,$$

where $W_1, \ldots, W_5$ are scalar functions of only $\nu \equiv q \cdot p/m$ and q^2. The details of these derivations are given in [10], in almost complete analogy to the derivations of similar expression in deep inelastic scattering cross sections of electrons on nucleons (see for example Halzen and Martin: "Quarks and Leptons"). In the limit of zero lepton masses the above expression reduces to

$$\frac{d^2\Gamma}{d\nu\, dq^2} = \frac{G_F^2 |V_{cb}|^2}{48\pi^3 E} \sqrt{\nu^2 - q^2} \left[W_1\, 3q^2 + W_2\, (\nu^2 - q^2) \right] \,,$$

with only $W_1(\nu, q^2)$ and $W_2(\nu, q^2)$ being non-zero.

The W functions can be parameterized in many different ways. For the parameterization in terms of $\omega = v_B \cdot v_{D^*} = (m_B^2 + m_{D^*}^2 - q^2)/2m_B m_{D^*}$, for example, the differential decay rates are [17]

$$\begin{aligned}
\frac{d\Gamma(\bar{B} \to D^{*+} l^- \bar{\nu}_l)}{d\omega} &= \frac{G_F^2 |V_{cb}|^2}{48\pi^3}\, m_B^5\, r_*^3 (1 - r_*)^2 \sqrt{\omega^2 - 1} (\omega + 1)^2 \\
&\quad \times \left[1 + \frac{4\omega}{\omega + 1} \frac{1 - 2\omega r_* + r_*^2}{(1 - r_*)^2} \right] \mathcal{F}^2(\omega) \,, \\
\frac{d\Gamma(\bar{B} \to D^{+} l^- \bar{\nu}_l)}{d\omega} &= \frac{G_F^2 |V_{cb}|^2}{48\pi^3}\, m_B^5\, r^3 (1 + r)^2\, (\omega^2 - 1)^{3/2} \mathcal{G}^2(\omega) \,,
\end{aligned}$$

where v_B and v_{D^*} are the velocities of the initial and the final states, and $r_* = m_{D^*}/m_B$. In the limit of $m_B \to \infty$ the $\mathcal{F}(\omega)$ and $\mathcal{G}(\omega)$ coincide with what is known as the universal function $\xi(\omega)$ which is normalized to unity at zero recoil $\omega = 1$. By using theoretical inputs $\mathcal{F}(1) = 0.91 \pm 0.03$ and $\mathcal{G}(1)/\mathcal{F}(1) = 1.08 \pm 0.06$ [32], ALEPH measures [12]

$$|V_{cb}| = 0.0344 \pm 0.031 \,,$$

which compares well with the results in (1.15) and (1.16) obtained by techniques that use the value of the $\mathcal{B}(b \to l)$ as an input.

Another parameterization of the form-factors is in terms of fraction of hadron momentum carried by the b quark. It can be easily shown that in the free quark limit the hadron tensor $W_{\mu\nu}$ has a form similar to the leptonic tensor $L_{\mu\nu}$, with a delta function factor $\delta((p_b + q)^2 - m_q^2)$ that is needed for the conservation of four on-shell momenta. It can be easily shown [10] that

$$\delta((p_b + q)^2 - m_q^2) = \frac{1}{m_B(m_b - m_B\xi_-)}\,\delta\left(\xi_+ - \frac{m_b}{m_B}\right) ,$$

where

$$\xi_\pm = \frac{\nu \pm \sqrt{\nu^2 - q^2 + m_q^2}}{m_B} .$$

Therefore, in the free quark limit the system exhibits the property of scaling, known from the studies of the deep inelastic scattering of electrons on nuclear targets. The large momentum transfer in b hadron decays corresponds to the space-time separations close to the light cone. It can be easily shown in analogy with the deep inelastic scattering that in the light-cone limit the W functions can be expressed as a function of a single probability distribution $f(\xi)$, normalized to 1 on $0 \le \xi \le 1$. In the simple limit of massless leptons this looks like [10]

$$\begin{aligned} W_1 &= 2\,[f(\xi_+) + f(\xi_-)] , \\ W_2 &= \frac{2}{\xi_+ + \xi_-}\,[\xi_+ f(\xi_+) - \xi_- f(\xi_-)] . \end{aligned}$$

In the free quark limit

$$f(\xi) = \delta\left(\xi - \frac{m_b}{m_B}\right) ,$$

i.e., the ξ is the fraction of momentum carried by the b quark inside the hadron. Calculations using OPE and HQET [10] indicate that the $f(\xi)$ is sharply peaked

around $\xi = m_b/m_B$. The $f(\xi)$ has a potential of revealing the structure of the b hadrons. For example the width of $f(\xi)$ would reveal the role of the bound valence quarks, while the behavior of $f(\xi)$ at the lower values of ξ has a potential of revealing the role of softer $g \to q\bar{q}$ processes inside the hadron (again in complete analogy with the deep inelastic scattering experiments).

In our opinion, direct or indirect measurement of $f(\xi)$ would be a prime one to complement the measurement of $\mathcal{B}(b \to l)$ and to get a better insight into the structure of b hadrons.

2

Detector, event selection, and the datasets

The SLD experiment is designed to study the physics of $e^+e^- \to Z \to f\bar{f}$ or $q\bar{q}$ at the Z resonance. An overview of the detector subsystems is given in this chapter, including a brief presentation of event classification, MC simulations, and some of the most important features of the datasets. At the end, a little digression is made by presenting a new approach to Vertex Detector stand-alone tracking.

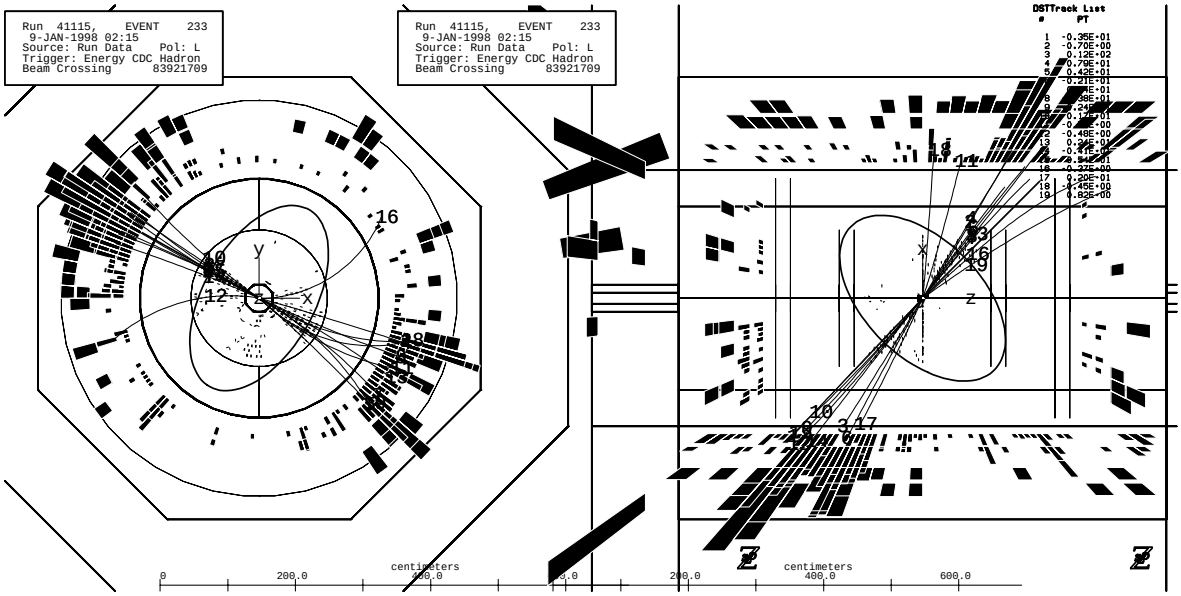
2.1 Overview of the SLD experiment

Fig. 2.1 shows the detector cross sections across and along the beams, featuring an event interpreted as $b\bar{b}$. A detailed SLD quadrant layout is given at the bottom of the figure.

Bunches of incident electrons and positrons, each about a millimeter in length and about a micron in diameter, collide head on in the center of the detector at their center of mass energy approximately equal to that of the Z resonance. Each bunch contains about 10^{10} particles. Under normal conditions the beam crossing frequency is 120 Hz. The SLD xyz reference frame is centered approximately at the interaction point (IP), and is oriented so that positrons go in the $+z$ and electrons in the $-z$ direction (theoretical expressions for the scattering cross sections usually take the direction of the incident *electrons* to be in the $+z$ direction, which is taken into account in our calculations by reversing the sign of the polar angle $\cos\theta$ in these expressions).

The main properties of incident e^+ and e^- are their absolute energies, luminosity, and the polarization of e^-.

Absolute energies (relative to the SLD reference frame) of both incident beams are measured for each beam crossing separately, using a pair of Wire Imaging Synchrotron Radiation Detectors (WISRD). The knowledge of the beam energies is important to accurately estimate the actual fraction of the Z events in the $Z-\gamma$ admixture at the Z energy peak, and to accurately estimate various asymmetry parameters, like A_e, A_b, and others.



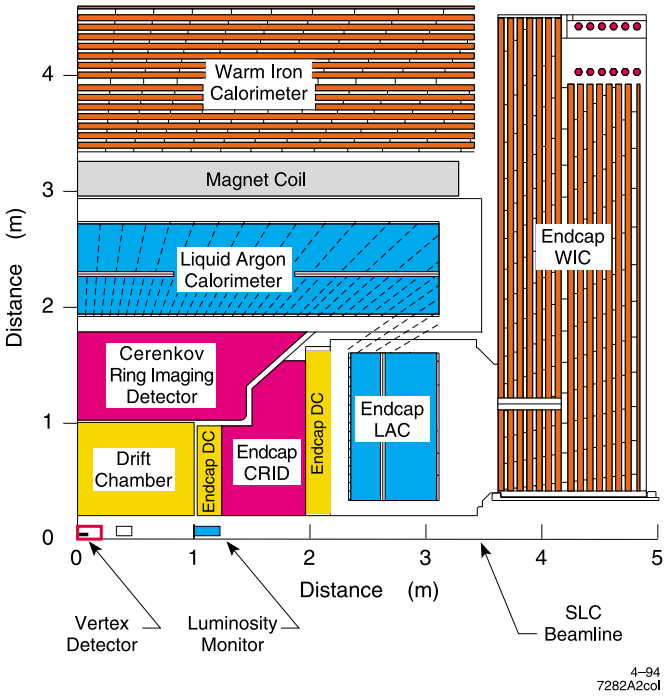


Figure 2.1: Top row: cross sections of the detector across and along the beams featuring an event interpreted as $b\bar{b}$. The superimposed ellipse centered at the IP represents the event hemisphere plane. Right: quadrant view of the SLD along the beams [33]. Calorimeter modules are radially segmented to match radial-like Z decay topologies such as one in the top two images.

The luminosity is measured using the SLD Luminosity Monitor (LUM). It is related to the small angle elastic e^+e^- scattering cross section, which itself can be related to the Z production rate by knowing the density distributions in the electron and positron bundles. The luminosity is useful to cross check the Z reconstruction efficiencies.

The polarization of the incident electrons affects the coupling of the Z to leptons and quarks. It enhances the forward-backward asymmetry in the scattering cross section. In our analysis, as in a number of other SLD studies, it is used to enhance the separation between the initial state b and $\bar{b}$ flavors. Polarized electrons are a unique SLD feature and will be given somewhat more attention.

The Z decay products are analyzed by the SLD subsystems. The task boils down to first determining the event types: hadronic, τ pair, e^+e^-, or $\mu^+\mu^-$, then to determining the relevant kinematical and topological properties of the events: reconstructing tracks and determining their momenta and velocities, identifying decay vertices, etc.

Based on expected topological and kinematical properties of the Z decays, certain conditions of the detector, known as the *readout triggers*, are specified, that have to be fulfilled for a beam crossing to be read out and processed for further analysis. This is euphemistically called an *event*. The triggers are so tuned as to record real events with maximum efficiency, and at the same time to reject backgrounds and noisy beam crossings. The latter are most often caused by glitches in the LINAC which result in either defocusing or the misalignment of the beams (just a few particles drifting away from the beam and hitting a portion of the beam-pipe is enough to fill the entire detector with interaction products).

The data acquisition system, the triggers, and the way the data are processed and classified will be described following a brief overview of the incident beam measurements, and of the SLD subsystems.

2.2 Incident beams: luminosity, energy, and polarization measurements

A very brief description of the energy and the luminosity measurement is followed by a more detailed description of the measurement of the polarization, which explicitly enters our relations for the initial state flavor probability (see Ch. 5).

2.2.1 Energy measurement

The current uncertainty of the center of mass energy E_{cm} at the SLD (around 0.03%) is negligible for our analysis. In principle, the knowledge of the E_{cm} is important to correctly asses the values of the asymmetry parameters like A_e and A_b which depend on the fraction of the Z events in the $Z-\gamma$ admixture.

The energy of a beam that passes through the IP is measured by Wire Imaging Synchrotron Radiation Detectors (WISRD). The beam is first bent horizontally, then vertically, then horizontally again. The purpose of the two horizontal sharp "kinks" is to generate two swathes of synchrotron radiation, while the purpose of the softer vertical bending is to vertically separate the swathes, and to direct them toward a WISRD component known as the wire array [34]. The vertical separation of the two swathes is proportional to the beam energy.

During the 1997-98 data taking, the WISRD was calibrated using $3,700$ and $3,000$ 120 Hz events at energies above and below the Z resonance respectively. From a precise knowledge of the energy spectrum near the Z resonance, it was determined that the WISRD has a systematic energy bias equal to -46 ± 25 MeV. The resulting luminosity weighted mean center-of-mass energy for the 1997-98 run was then determined to be $E_{cm} = 91.237 \pm 0.029$ GeV [35].

2.2.2 Luminosity Monitor

Luminosity is defined as the proportionality factor between the the event rate R and the interaction cross section σ_{int}, $R = \mathcal{L}\sigma_{int}$, and is given by

$$\mathcal{L} = f\frac{n_1 n_2}{4\pi\sigma_x\sigma_y}H_d$$

where n_1 and n_2 are the number of particles in the bunches, f the collision frequency, σ_x and σ_y half-widths of the Gaussian beam profiles in horizontal and vertical direction normal to the beam axis, and H_d the disruption enhancement factor due to the pinch the charge beams feel when going through each other. When the two beams are of opposite charge they attract each other, $H_d > 1$. When they are of like charge they repel each other, $H_d < 1$ ($H_d = 1$ for neutral beams) [36].

The luminosity is determined by measuring the rate of small angle elastic e^+e^- (Bhabha) scattering cross section, which is given by $d\sigma/d\cos\theta \approx \alpha^2/(s\,\theta^4)$, where α is fine structure constant, s the center of mass energy, and θ the scattering angle. The rate is measured by two luminosity monitors, Fig. 2.2, which are located symmetrically about 1m on each side of the IP along the beams (see Fig. 2.1 for the location of the LUM). Each consists of two components, one called Luminosity Monitor/Small Angle Tagger (LMSAT) that covers the angular region 23-68 mrad with respect to the IP, and another called Medium Angle Silicon Calorimeter (MASiC) that covers angular region 68-190 mrad (see [37] for details). The rate R is calculated for a set of appropriate angular, energy, and a-collinearity cuts and and the luminosity calculated as $\mathcal{L} = R/\sigma_{int}$.

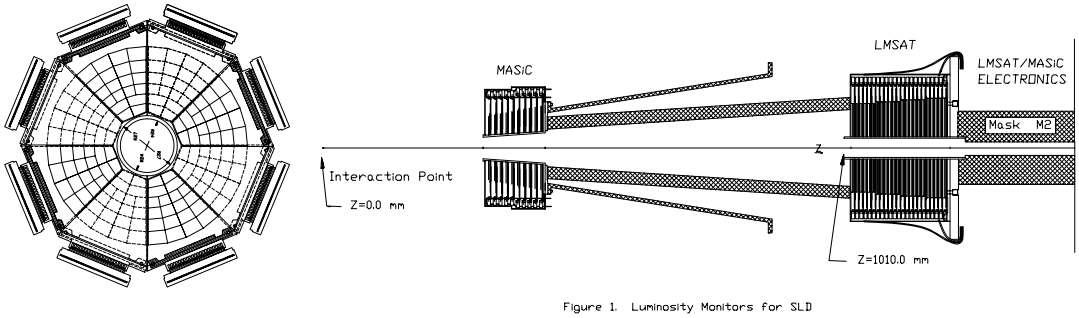


Figure 2.2: Cross sections of the SLD Luminosity Monitor across and along the beams.

The luminosity of the SLAC Linear Collider (SLC) increased by a factor of two during data taking in 1997-98, reaching beam sizes as small as 1.5×0.65 microns and the full beam intensity of 4×10^{10} particles per bunch [38]. This increase in the luminosity was independently verified by the increased rate of recorded Z decays.

2.2.3 Polarization measurement

The polarization of the incident electrons is measured with the SLD Compton Polarimeter which is located 33 m from the IP down the electron beam. It consists of a 532 nm (2.33 eV/photon) frequency modulated Nd:YAG pulsed laser, polarization and transport optics, and a multichannel Cherenkov counter.

Incident electrons that pass the IP undeflected collide with the circularly polarized laser beam produced by the polarimeter. The cross section of the scattered electrons is then measured as a function of their energy in the interval 17 to 27 GeV/c [39], and the electron polarization calculated from the asymmetry in the cross sections between parallel and anti-parallel spin configurations for the electron-photon pair.

The total spin projection $|J_z|$ of the pair can take values 3/2 for parallel configurations, (L_e, R_γ) and (R_e, L_γ), and 1/2 for anti-parallel configurations, (L_e, L_γ) and (R_e, R_γ). If $f(L_\gamma) = (1 - P_\gamma)/2$ and $f(R_\gamma) = (1 + P_\gamma)/2$ are fractions of left and right-handed photons in the laser beam, then for a left-handed electron

$$\begin{aligned} \sigma(L_e) &= \sigma(L_e \mid R_\gamma) f(R_\gamma) + \sigma(L_e \mid L_\gamma) f(L_\gamma) \\ &= \sigma_{3/2} \frac{1 + P_\gamma}{2} + \sigma_{1/2} \frac{1 - P_\gamma}{2} \end{aligned}$$

$$= \frac{1}{2}(\sigma_{3/2}+\sigma_{1/2})\left(1+P_\gamma\frac{\sigma_{3/2}-\sigma_{1/2}}{\sigma_{3/2}+\sigma_{1/2}}\right)$$
$$\equiv \sigma_{unp}(1+P_\gamma A_{LR})\,,$$

where P_γ is the laser beam polarization and A_{LR} theoretical asymmetry. For the right-handed electron one similarly obtains $\sigma(R_e)=\sigma_{unp}(1-P_\gamma A_{LR})$ so that for the two polarized beams the cross section reads

$$\begin{aligned}\sigma(P_e,P_\gamma) &= \sigma(L_e)f(L_e)+\sigma(R_e)f(R_e)\\ &= \sigma_{unp}(1+P_\gamma A_{LR})\frac{1-P_e}{2}+\sigma_{unp}(1-P_\gamma A_{LR})\frac{1+P_e}{2}\\ &= \sigma_{unp}(1-P_eP_\gamma A_{LR})\,, \qquad (2.1)\end{aligned}$$

where σ_{unp} is the unpolarized cross section (which has the same value if at least one of the two beams is unpolarized regardless of the polarization of the other beam).

Calculation of the Compton scattering matrix element itself is straightforward but lengthly: one starts with the most general matrix element for the Compton scattering when both momenta are non-zero (see for example [40]), does not average over the two initial photon (circular) polarization vectors, and uses "polarized" fermion projectors

$$u\bar{u}=(\not{p}+m)\,\Lambda(s) \quad \text{and} \quad v\bar{v}=(\not{p}-m)\,\Lambda(s)$$

in the trace formulas, where $\Lambda(s)=(1+\gamma^5\not{s})/2$ is the electron spin projector along $s=(s^0,\mathbf{s})$ ($s^2=-1$). Since one needs to calculate only $\sigma(L_e \mid R_\gamma)$ and $\sigma(L_e \mid L_\gamma)$, both of which have the same $s^0=\gamma\beta$ and $\mathbf{s}=-\gamma\mathbf{p_1}/|\mathbf{p_1}|$, for electrons with incoming momentum $\mathbf{p_1}$, the difference in the two cross sections is from the opposite photon polarizations. At the lowest order [41]

$$A^0_{LR}=\frac{(AB^3-A^3B)+m^2(AB^2+A^2B-A^3-B^3)}{(AB^3+A^3B)+2m^2(A^2B-AB^2)+m^4(A-B)^2}\,, \qquad (2.2)$$

where

$$
\begin{aligned}
A &= p_1 k_1 = -E_1 E_{1\gamma} - |\mathbf{p_1}| E_{1\gamma}\,, \\
B &= p_2 k_1 = -E_2 E_{1\gamma} - |\mathbf{p_2}| E_{1\gamma} \cos\theta\,, \qquad (2.3) \\
\cos\theta &= \frac{E_1 E_2 + E_{1\gamma} E_2 - E_{1\gamma} E_1 - E_{1\gamma}|\mathbf{p_1}| - m^2}{|\mathbf{p_2}|(|\mathbf{p_1}| - E_{1\gamma})}\,.
\end{aligned}
$$

The indices 1 and 2 refer to the incoming and outcoming particles respectively. Radiative corrections appear to be important and are discussed in detail in [42, 43].

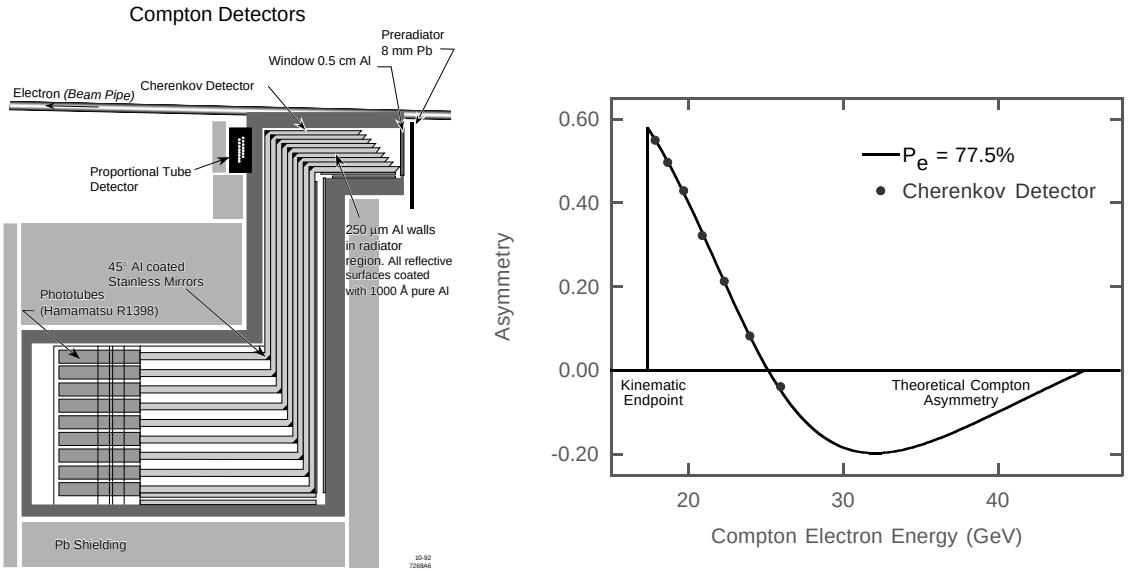


Figure 2.3: Counts from the seven (out of 9) Compton Detectors (left) are used to estimate the Asymmetry in the right hand side plot.

When $\cos\theta$ is inserted into B, Eq. (2.3), and A and B from (2.3) into (2.2), one obtains A_{LR} as a function of the energy of the Compton electron E_2. The value of $P_e P_\gamma$ is then extracted by fitting the $P_e P_\gamma A_{LR}$ from Eq. (2.1) to seven points obtained from the Cherenkov counter, Fig. 2.3.

The points in Fig. 2.3 correspond to seven channels in the Cherenkov counter. The channels are radially spaced from 10.5 cm to 16.5 cm, roughly 1 cm apart. Given

the kinematic constraints from the Compton scattering [44] this corresponds to E_2 range from about 25 down to about 17 GeV.

The Asymmetry versus Compton Electron Energy in Fig. 2.3 is given by

$$\text{Asymmetry} = \frac{n(3/2) - n(1/2)}{n(3/2) + n(1/2) - 2\,[n(\text{laser off}) + n(\text{pickup})]} = P_e P_\gamma A_{LR}\,, \quad (2.4)$$

where $n(3/2)$ and $n(1/2)$ are counts corresponding to $|J_z|$ equal to 3/2 and 1/2 respectively, n(laser off) is the average count when the laser is off, and n(pickup) the average count due to a small electronics noise pickup [45].

The photon polarization, P_γ, needed to extract the value of P_e, is measured using a laser beam analyzer located just after the Compton IP [39]. Combined statistical and systematic errors in polarization measurement are around $1.5 - 3.0\%$.

2.3 The SLD subsystems

The SLD subsystems are used to observe the Z decay products. Particles that leave visible tracks in the detector are e, μ, π, K, and p. Neutral particles like γ, K^0_L, or n are generally observable by the calorimetry. Neutrinos are only "observable" by means of their missing energy. Long-lived particles, those with lifetimes 10^{-10} s, like K^0_s and Λ_0 are observed by means of their displaced decay vertices, several cm from the IP in the average, and their invariant mass. Short-lived particles, those with lifetimes 10^{-13} - 10^{-12} s, like B and D mesons, Λ_b and τ, are observed by a means requiring accurate reconstruction of the decay vertices displaced from the IP by several hundred μm in the average. At the SLD and elsewhere this is achieved by using high precession Vertex Detectors.

2.3.1 Vertex Detector (VXD)

The purpose of the Vertex Detector is precision tracking and reconstruction of the decay vertices near the IP. Two Vertex Detector designs were employed at the SLD: VXD2, for 1992-95 data taking, and VXD3, for 1996-98 data taking. Their cross sections along the beams are shown in Fig. 2.4. Plots at the right hand side of Fig. 2.4 show impact parameter resolutions of VXD2 and VXD3 at $\cos\theta = 0$. The resolutions are degraded somewhat at lower polar angles due to the effectively higher thicknesses of the materials seen by the particles along their trajectories at these angles. For arbitrary θ, the overall VXD3 impact parameter resolutions deduced from the data are [46]

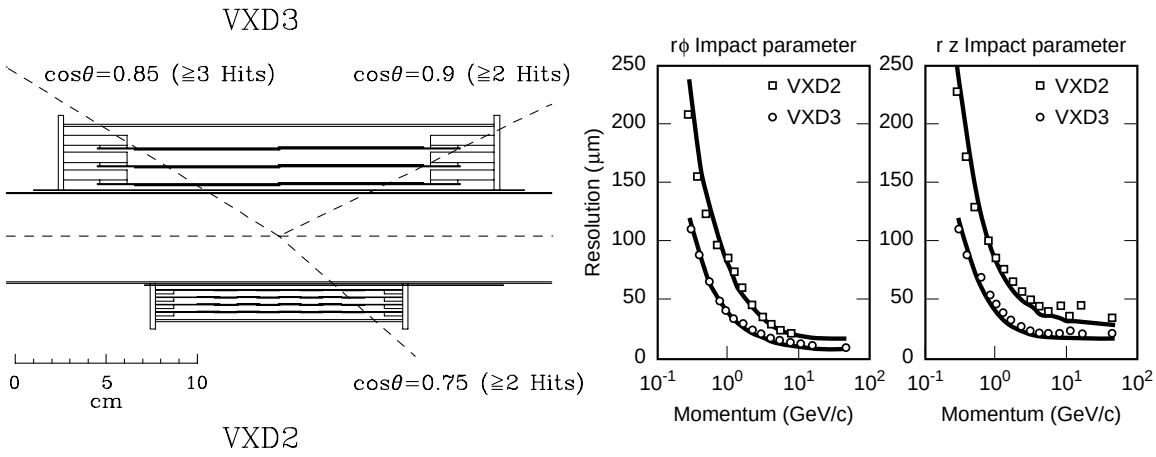


Figure 2.4: Left: cross sections of VXD2 and VXD3 along the beams. Right: impact parameter resolutions of VXD2 and VXD3 at $\cos\theta = 0$. Lines correspond to the MC simulations, while points correspond to the measured values. Improvement in the resolution form VXD2 to VXD3 is mostly due to the increase of spacing between the CCD layers.

$$\sigma_{r\phi} = \sqrt{14.0^2 + \left(\frac{33.0}{p\sin^{3/2}\theta}\right)^2}\ \mu m \qquad \text{and} \qquad \sigma_{rz} = \sqrt{26.5^2 + \left(\frac{33.0}{p\sin^{3/2}\theta}\right)^2}\ \mu m\,.$$

The details of the design and the operation of the VXD3, including the data acquisition system, are also given in [46]. Principles and applications of the silicon vertex detectors, as well as of many other systems for precision tracking, are nicely presented in [47].

An example of a VXD reconstructed event is shown in Fig. 2.5, where the picture on the right shows the topology of the event at distances roughly 1 cm around the IP.

Another useful application of the device is for the calibration of the tracking efficiency. Here we anticipate some of the results from Ch. 4, where a technique for the calibration of the electron and muon MC counts is described in detail. The technique uses statistical independence of two hypothesis tests (for the same hypothesis) as a mean of improving the accuracy of the calibration. In complete analogy, one can do the tracking efficiency calibration, by comparing outcomes of the two mutually independent tests for a single track hypothesis, where one of the tests uses only VXD information, and another only DC information. This was the original motivation for the development of the stand-alone tracking in the VXD3, which will be shortly described at the end of this chapter.

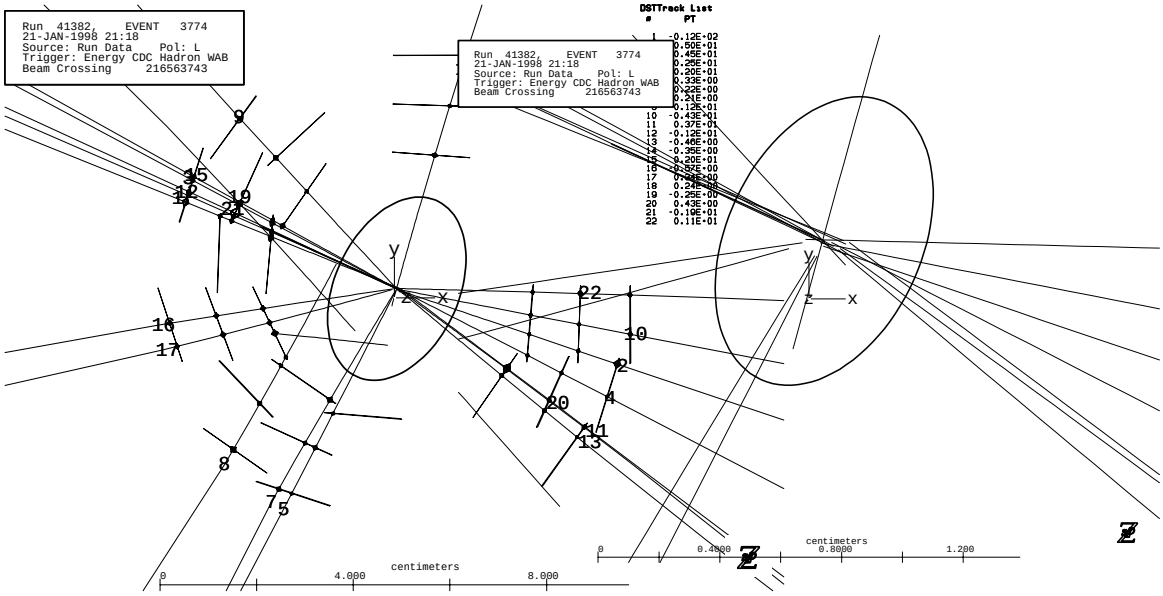


Figure 2.5: Precision tracking using the Vertex Detector: tracks extrapolated through the CCD clusters (diamonds) on the left, and zoomed IP region of the same event on the right.

2.3.2 Drift Chamber (DC)

The Drift Chamber is used to identify charged particles and to measure their momenta. The momenta are determined from the curvature of the charged tracks, which curl inside the uniform magnetic field of 0.6 T, provided by the magnet solenoid, Fig. 2.1. The Central Drift Chamber (CDC) is in a form of a hollow barrel 0.2 to 1 m in radius and 2 m in length. Two end-cap drift chambers (EDC) cover the faces of the CDC. Another EDC pair is located just behind the end-cap CRID (see Fig. 2.1 for the details).

Charged particles leave traces in the DC through the ionization of a gas admixture that fills the entire volume of the Drift Chamber. The admixture consists of 75% CO_2, 21% Argon, 4% Isobutane, and 0.2% water. Electrons produced in the process are transported (drifted) through the medium by the electrostatic field provided by the carefully configured arrays of *sense*, *guard*, and *field shape* wires, grouped into 10 "super-layers". A portion of a super-layer is shown on the left hand side of Fig. 2.6. Electrons that land on the sense wires produce charge waveforms, which propagate through the wires, and are read out by the fast coincidence electronics mounted directly on the faces of the CDC. From that information, segments of the charged tracks in each super-layer are reconstructed. They are called *vector hits* and are seen as little line segments on the right hand side picture in Fig. 2.6. Arrays of vector hits

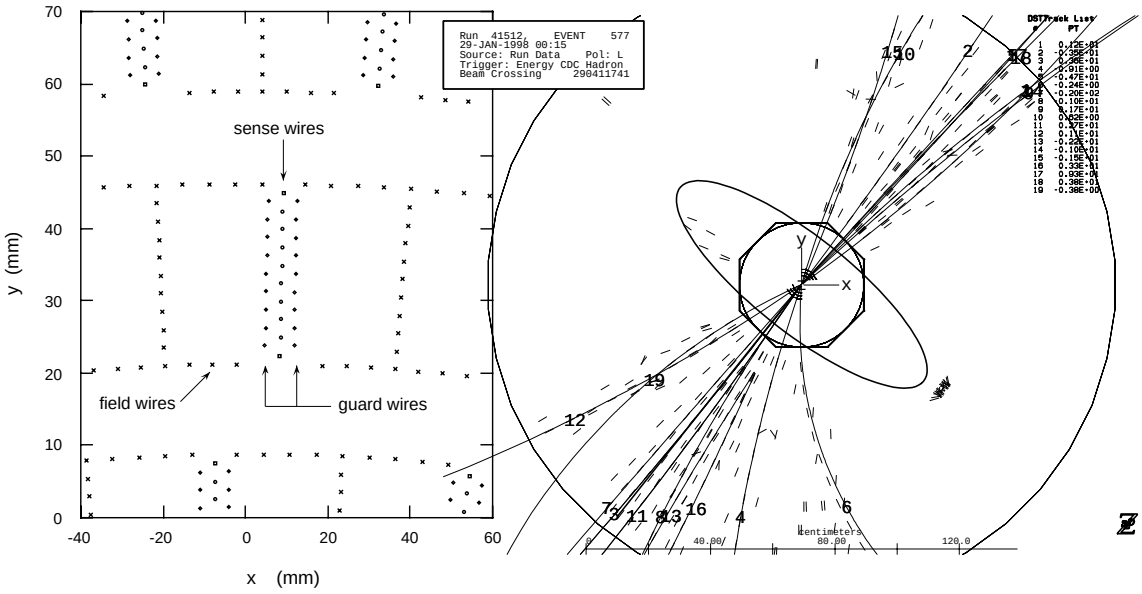


Figure 2.6: Left: portion of the CDC across the beams featuring one out of ten CDC super-layers. Right: vector hits from all 10 super-layers are shown, together with reconstructed tracks that link vector hits consistent with a track hypothesis.

consistent with a track hypothesis are linked into objects which we call *tracks*.

The coincidence measurement (timing) at the faces of the CDC is needed to determine the axial (z) projections of the vector hits. The z measurement is additionally enhanced by using a clever orientation of the sense wires relative to the z axis, that introduces a geometric 3D effect in the vector hit reconstructions, which would otherwise be only 2D (in xy plane). This is achieved by alternating the orientation of the wires with respect to the z axis from one super-layer to another. The bottom super-layer, one closest to the beams, is axial (wires parallel to the z axis), while direction of the wires in the two super-layers above the bottom one are slightly rotated, by $+42$ and -42 mrad respectively. This is repeated with the next two triplets of the super-layers, with the outermost, the 10th super-layer, being axial. The role of this configuration can simply be understood by thinking of determining the xy position of a point in a plane, by first measuring the x coordinate of the point in the xy coordinate system, then by measuring the x' coordinate in a slightly rotated $x'y'$ coordinate system.

Track reconstruction using DC, the central problem in the event reconstruction, is a highly nontrivial 3D geometry and pattern recognition task, and is described in some detail in [48].

2.3.3 Cherenkov Ring Imaging Detector (CRID)

CRID is used to measure the velocities of the charged particles. By knowing the momentum and the velocity, one can calculate the mass, and therefore deduce the particle identity.

A good account of the CRID principles, design, and performance, can be found in [49, 50]. The detector explores the effect of radiation named after Russian physicist P. A. Cherenkov, who was the first to observe, back in 1937, that charged particles traveling in a medium faster than the speed of light in that medium continuously emit light along their trajectories. When directed to a parabolic mirror, the reflected conical envelope formed by the light waves emitted along the trajectory has a form of a ring in the mirror focal plane. The radius of the ring is directly proportional to the particle velocity.

The range of velocity β is determined from $\beta n(\omega) > 1$, where $n(\omega)$ is the index of refraction of a medium for a given frequency. To cover wider range of velocities, CRID uses two radiators: liquid radiator, with higher $n(\omega)$ for lower velocities, and gas radiator, with lower $n(\omega)$ for the higher velocity range. The Cherenkov angle versus momentum dependence is shown in the bottom-right plot of Fig. 2.7. The same figure shows cross sections of the barrel and the end-cap portions of the CRID. An example of a ring is given in the lower-left plot.

2.3.4 Liquid Argon Calorimeter (LAC)

LAC is a sampling calorimeter designed to measure energy deposited by charged and neutral particles. It has two electromagnetic (EM) and two hadronic (HAD) modules, which consist of radially segmented towers of lead pads serving as the energy samplers [51].

The central application of the LAC, in coincidence with the Worm Iron Calorimeter, is as hadronic event trigger. It is also the main subsystem for the identification of electrons.

Position of LAC inside the SLD is shown in Fig. 2.1, and its isometric drawing in Fig. 2.8.

The two innermost layers of the LAC form the electromagnetic module and are primarily designed to separate electrons from other particles. At energies above the critical energy, which is given by the well known Bethe-Heitler formula $E_c \approx 1600 m_e c^2/Z$, the dominant mechanism of energy loss for electrons is by emission of radiation due to acceleration of electrons in the electric fields of the nuclei (*bremsstrahlung*). From the bremsstrahlung cross section it can be seen that the emission probability varies as the inverse square of the particle mass [52], making the radiation loss for muons, for example, around 40,000 times smaller than the radiation loss for electrons. The signature of an electron is thus a wide lateral distribution and short

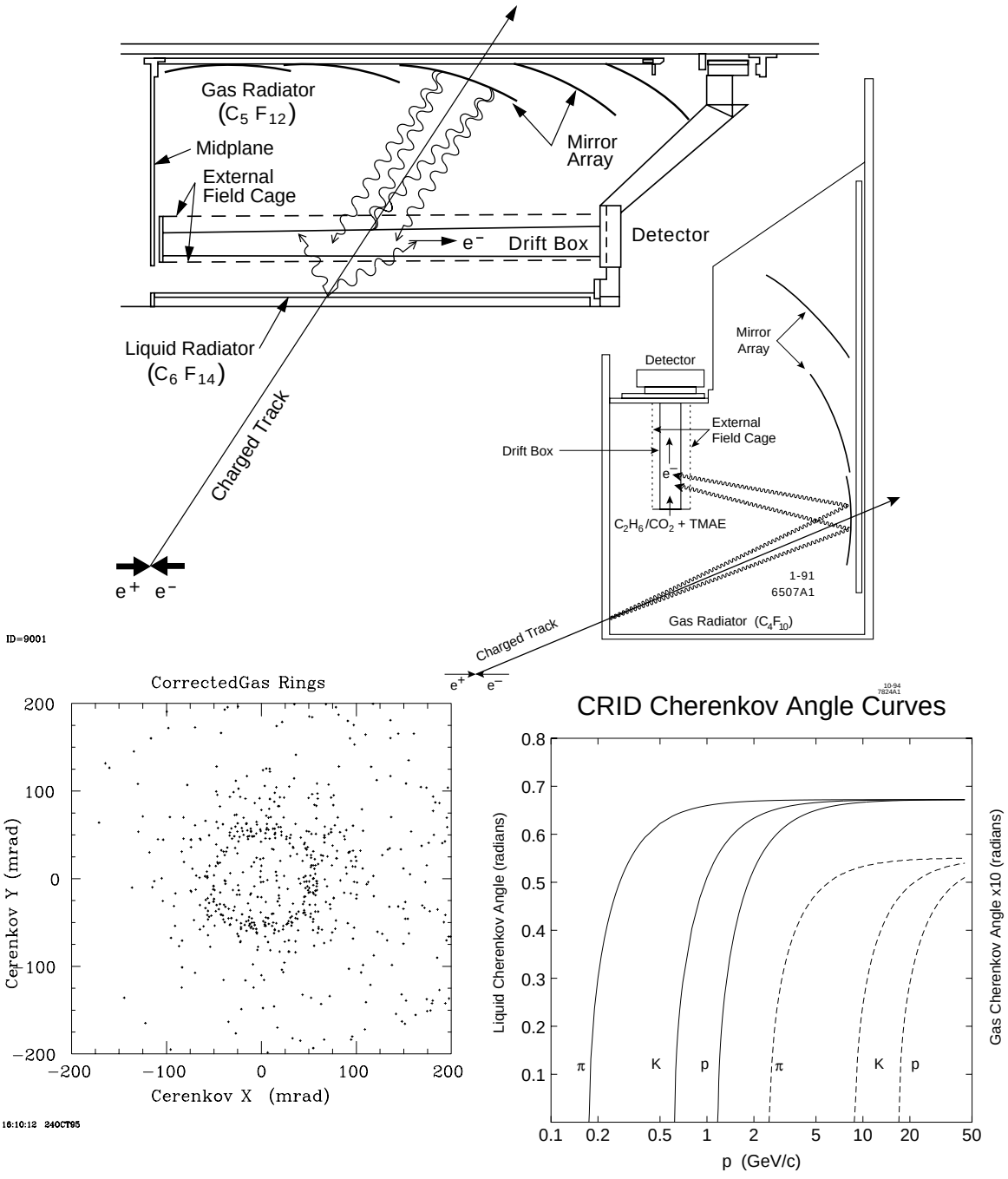


Figure 2.7: Top two drawings show barrel and end-cap CRID cross sections. An example of a reconstructed ring in the end-cap CRID is given in the lower-left plot [49], and the CRID ring radii as a function of momentum in the lower-right plot [50].

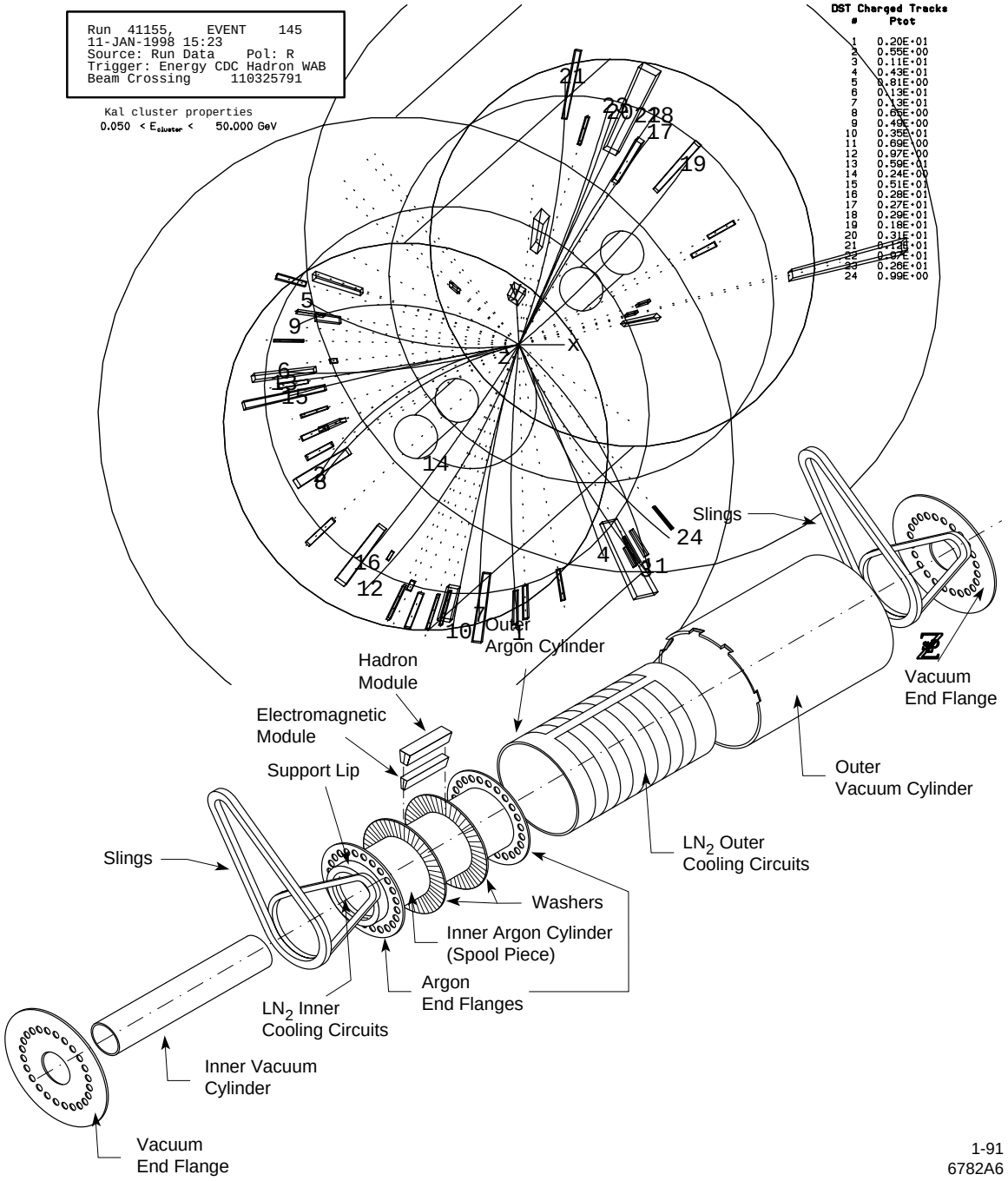


Figure 2.8: Top: a snapshot of a high track multiplicity event. The height of the towers is proportional to the energy stored in all four LAC modules. Bottom: isometric drawing of the LAC structure. EM and HAD modules are located between Argon End Flanges.

penetration depth. The lateral distribution of the scattered electrons is sampled using two finely segmented EM modules, while their short range is observed as a rapid decrease in the deposited energies along the four LAC modules. The total LAC thickness is 21 electron radiation lengths, sufficient to contain $98-99\%$ of electrons with energy of 50 GeV (electron radiation length is defined as distance at which the energy of an electron drops to $1/e$ of its initial value).

The HAD section of the LAC is positioned outward. HAD towers are twice as large as EM towers in both transverse dimensions. Each of the two subunits, HAD1 and HAD2, is 1 absorption length thick (hadronic absorption length is defined as the mean free path of a particle before undergoing a non-elastic interaction in a given medium). Combined thickness of EM + HAD modules is around 2.8 absorption lengths, enough to contain $80-90\%$ of all the hadrons.

An interesting alternative application of the hadronic LAC modules is to additionally discriminate muons from other particle, which is otherwise most effectively done using Warm Iron Calorimeter. We will see in Ch. 4 that this turns out to be important for the calibration of the MC muon counts.

2.3.5 Warm Iron Calorimeter (WIC)

WIC, in coincidence with LAC, serves as the hadronic event trigger. It is also the main subsystem for the identification of muons.

The WIC consists of 14 layers of 5 cm thick iron plates that serve as energy sampler (similar to lead pads in LAC) [55]. Sandwiched between the plates are layers of sampling chambers, Fig. 2.9, that implement streamer tubes developed by Iarocci and others [56]. Following the 7th and 14th iron plates are double layers of chambers including strips transverse to the tube directions, to completely determine space points for the tracking purposes [55].

Given the WIC thickness of 4 absorption lengths, the energy profiles across the layers serve as an additional means of discriminating muons from hadrons.

2.4 Trigger and event classification

We give here a very brief overview of the data acquisition system, the readout triggers, and of the event reconstruction and classification.

Aspects of the SLD design take particular advantage of the relatively low e^+e^- cross section and the low beam crossing rate of just 120 Hz [57]. Preamplified charge waveforms from active detector components (sense wires in the DC, Fig. 2.6, arrays of lead towers in the LAC [51], strip electrodes in the WIC [55], and others) are stored in various multiplexed systems of analog storage devices during an event. The only exception is the Vertex Detector, for which the analog storage devices, the

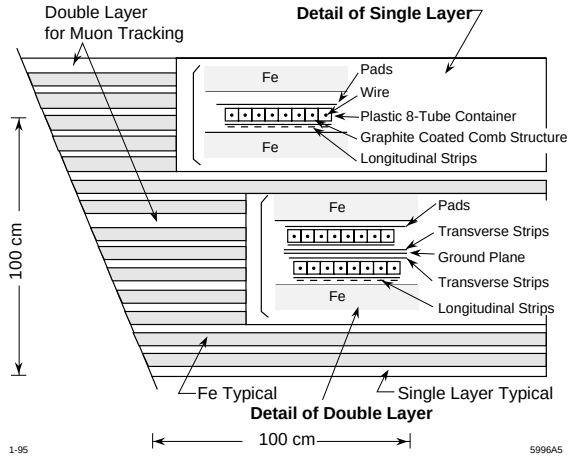
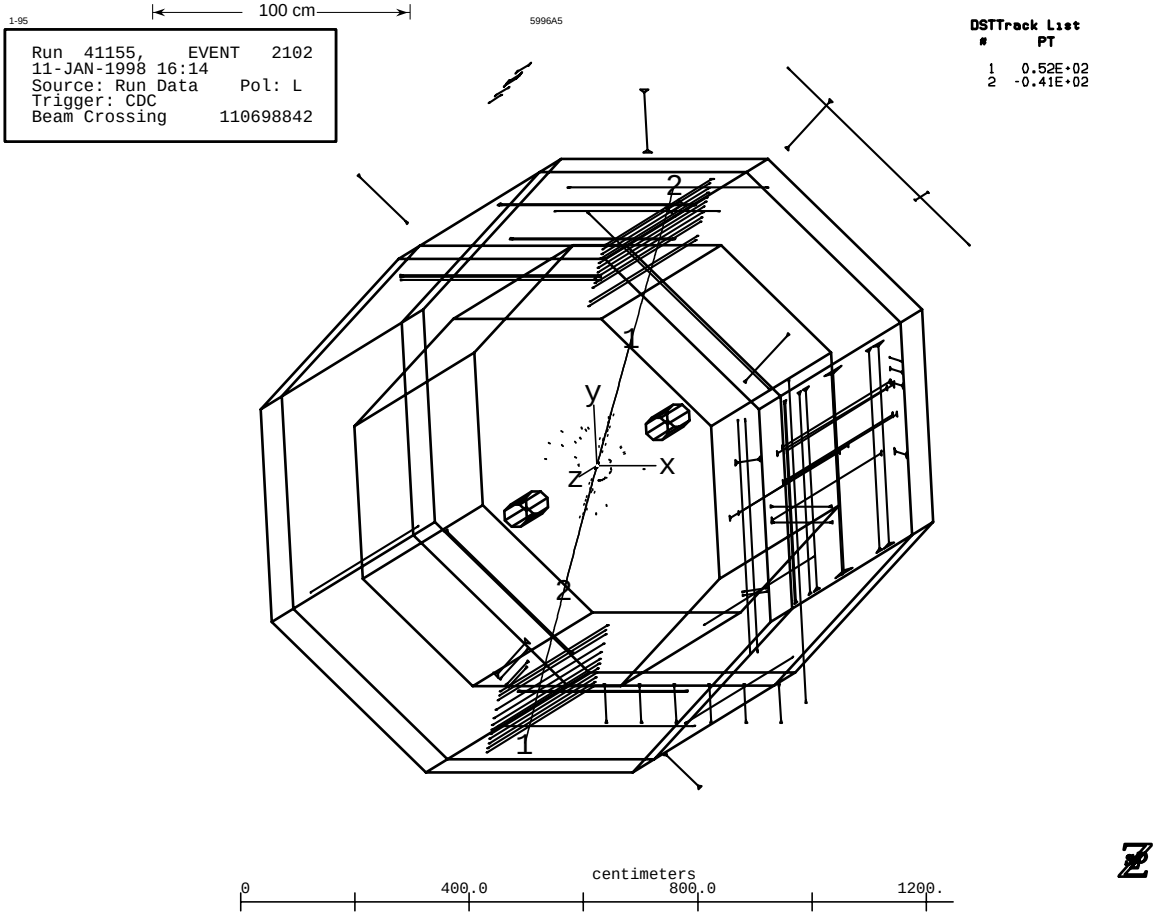


Figure 2.9: Left: schematics of one half of a WIC octant. Below: a back-to-back $\mu^+\mu^-$ event featuring muon finding hits in the WIC Iarocci tubes at approximately one and seven o'clock in xy plane.



CCD pixels, are an intrinsic part of the detector itself [46, 47]. These signals are digitized using accurate ADC units "glued" at the detector face. The front-end data acquisition electronics varies from one subsystem to another. Subsystems that use wires for signaling, like the DC and CRID, are coupled to a custom made device known as the SLAC Microstore [58]. LAC [51] and WIC [55] use similar systems.

Digitized outputs from the front-end electronics are then forwarded via optical cables to a network of FASTBUS modules known as ALEPH Event Builders (AEB) [59]. The task of event building is done by a collection of the AEB modules dedicated to each detector subsystem separately. An extra AEB module, known as the *trigger* AEB, is connected in parallel with all other AEB modules [57], and is responsible for making the decision of whether to accept or to reject a beam crossing. Beam crossings that pass the filter, about 3% of them, are recorded in files called RAWDATA. The trigger has efficiency of around 99% and purity of around 6%. Under normal running conditions the trigger rate is around $0.1 - 0.2$ Hz. During noisy running conditions it can exceed 2 Hz [49].

The triggers are fundamental for the selection of true Z events, so we describe them in some detail.

- **Energy:** This trigger is formed by summing the pulse-heights of LAC towers with energy exceeding a preset threshold. The high energy trigger thresholds are set to 60 ADC counts for towers in the two EM modules, and to 120 ADC counts for towers in the two HAD modules (in EM (HAD) towers one ADC count represents about 2.0 (5.4) MeV of energy loss for a minimum ionizing particle at muon energy scale). The sum of energies in all towers is required to be at least 8 GeV. Plots of the: pulse-height distributions, total energies in all four modules, total trigger energy distributions for Z events, and the Z efficiency of the energy trigger are given in [60].

- **CDC:** Requires at least two well separated tracks to be reconstructed. A good track is required to have > 30 hits, $p > 100$ MeV, $\mid z \mid < 20$ cm, $r_{xy} < 10$ cm.

- **Hadron:** Logical AND combination of Energy and CDC triggers, requiring at least one charged track in 9 CDC super-layers and an energy deposition in LAC sightly smaller than in the Energy trigger. This trigger is introduced to handle hadronic events with low track multiplicity.

- **WAB:** Requires two back-to-back tracks with good track specifications as in the CDC trigger.

- **Muon:** Pair of back-to-back tracks with WIC hits in opposite octants.

- **Tau:** At least one track with $p > 1$ GeV.

- **Random:** Records events every 20 seconds regardless of the detector status. It is used to determine background levels.

In addition, several vetoes in the triggering system are used that prevent excessive detector dead-time from noisy backgrounds.

Information stored in the RAWDATA files is subject to further processing, using various pattern-recognition procedures, to classify events into hadronic, τ, e^+e^-, and $\mu^+\mu^-$, Fig. 2.10, and to determine their kinematical and topological properties. The outputs of these reconstructions are stored in files called RECON. For our analysis the most important is hadronic event selection, although selection of other event types is important for the calibration purposes (see [44] for the τ, e^+e^-, and $\mu^+\mu^-$ event selection).

2.5 Hadronic event selection

Hadronic event filter is the first step in selecting clear hadronic events. Events that satisfy the **Energy** trigger are then "tightened" by imposing further requirement on the energy above the high tower threshold EHI (60 ADC in EM section and 120 ADC in HAD section), on the energy below the low tower threshold ELO (8 ADC in the EM section and 12 ADC in the HAD section), and on the number of towers in the EM section of the LAC above the high threshold NEHEMI [60]:

- NEHEMI $\geq$ 10,
- EHI $>$ 15 GeV,
- ELO $<$ 140 GeV,
- EHI $>$ 1.5 $\cdot$ (ELO $-$ 70).

This trigger is referred to as **Energy2**. It is estimated that 93% of hadronic Z decays pass the hadronic event filter [60, 61].

Further refinement of hadronic events is achieved by applying constraints which vary from analysis to analysis and depend on the efficiency versus purity requirements. We have used the SLD standard selection *as is*, implemented in routine `BBESEL` (written by Homer Neil), which uses the following cuts:

- Event contains at least 5 reconstructed tracks, each with a track momentum transverse to the beam axis $p_T > 200$ MeV/c and with the distance of closest approach to the IP < 5 cm.
- Visible energy calculated from the track momenta $E_{vis} > 12$ GeV, under the assumption that all tracks are pions.

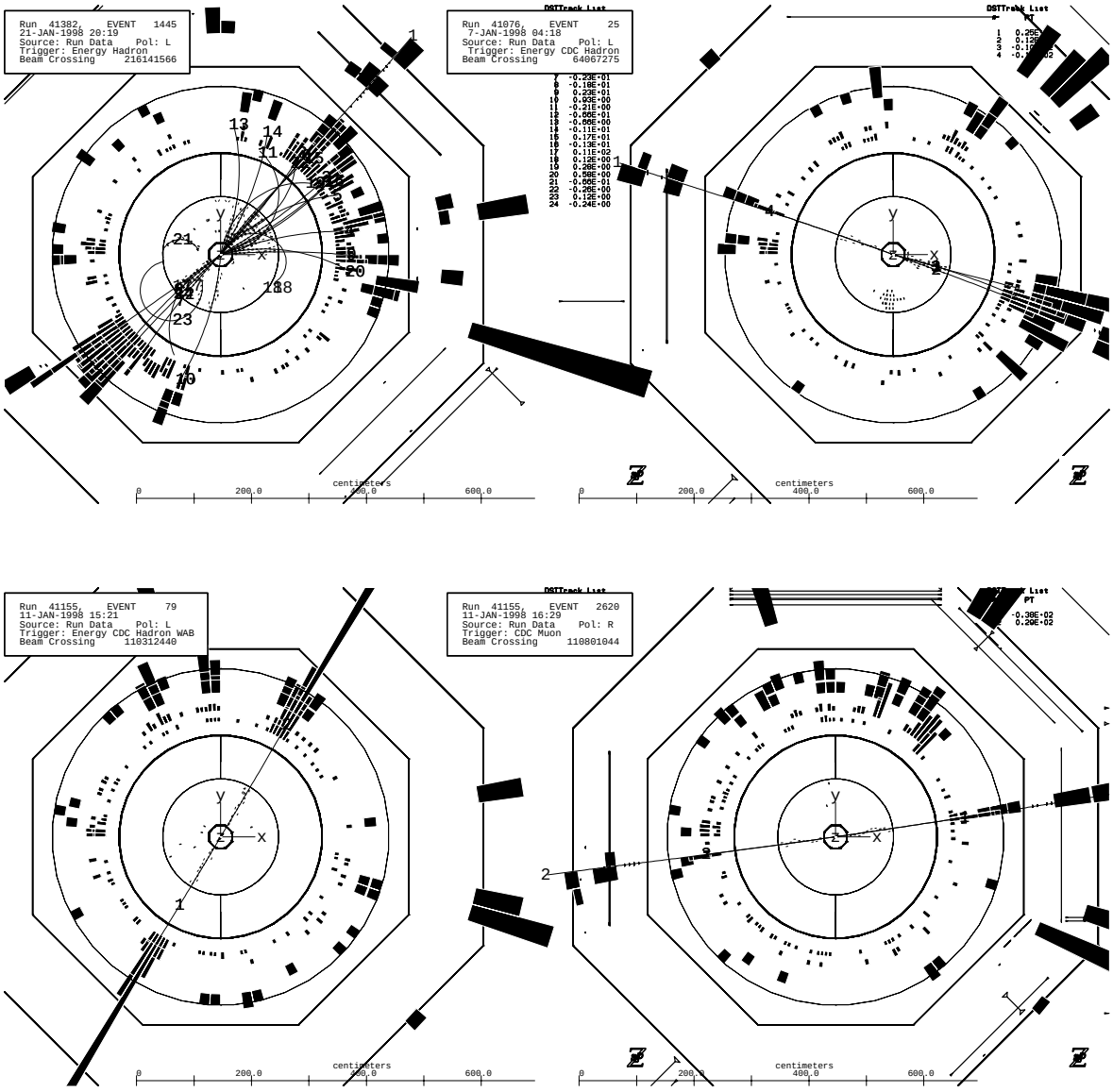


Figure 2.10: Event varieties at the Z: hadronic (top left), usually with large track multiplicity, τ event (top right), e^+e^- pair (bottom left), and $\mu^+\mu^-$ pair (bottom right). The τ pair in the top right plot decays into a muon in one arm (at around ten o'clock) and hadronically in another arm (at around four o'clock).

- Thrust $|\cos\theta| < 0.85$.
- At least two of the charged tracks must have two or more VXD hits.

The details of the $b\bar{b}$ event selection and the flavor tagging will be discussed within the context of our analysis later in Ch. 5.

2.6 Simulation and the Monte Carlo

The details of the SLD MC generation and the detector simulation is nicely presented in [48].

Hadronic events at the SLD were generated using Jetset Version 6.3 [62] for 1993-95 data, and Version 7.4 [63] for 1996-98 data. The differences between the two versions do not affect our analysis at all. Version 7.4 has a "structure" below the final parton level which allows for fragmentation studies, studies of $g \to b\bar{b}$ splitting, and the like. From the educational viewpoint Version 7.4 is interesting because it gives a nice primer for a simple applications of a data structure known as *graph* in representing the QCD processes.

The final state particles are produced by simulating multiple scattering, conversion, and other effects of particles interactions in the detector materials. The simulation is done using GEANT Version 3.21. For each data taking period there are several MC datasets which use either different reconstruction procedures, different input parameters, or different conjectures about the real detector geometry [48, 64, 29].

2.7 A digression: the Vertex Detector stand-alone tracking

The subject of this section is only very loosely related to our analysis, but since a good year and a half was dedicated to it we shall let it stand.

One of the goals in the SLD event reconstructions was to obtain portions of tracks in the VXD region using the information from the VXD *alone*. These portions of tracks, which link arrays of the VXD clusters consistent with the track hypothesis, are called VXD *vector hits*.

The original motivation behind the VXD stand-alone tracking was to improve the calibration of the tracking efficiency, which can be done by comparing outcomes of the two mutually independent tests for a track hypothesis, one of which uses only VXD and another which uses only CDC. This is similar to the approach that we use to calibrate electron and muon MC counts to the data in Ch. 4.

The upper portion of Fig. 2.11 illustrates the nature of the problem of reconstructing the VXD vector hits using the VXD information alone. For the average

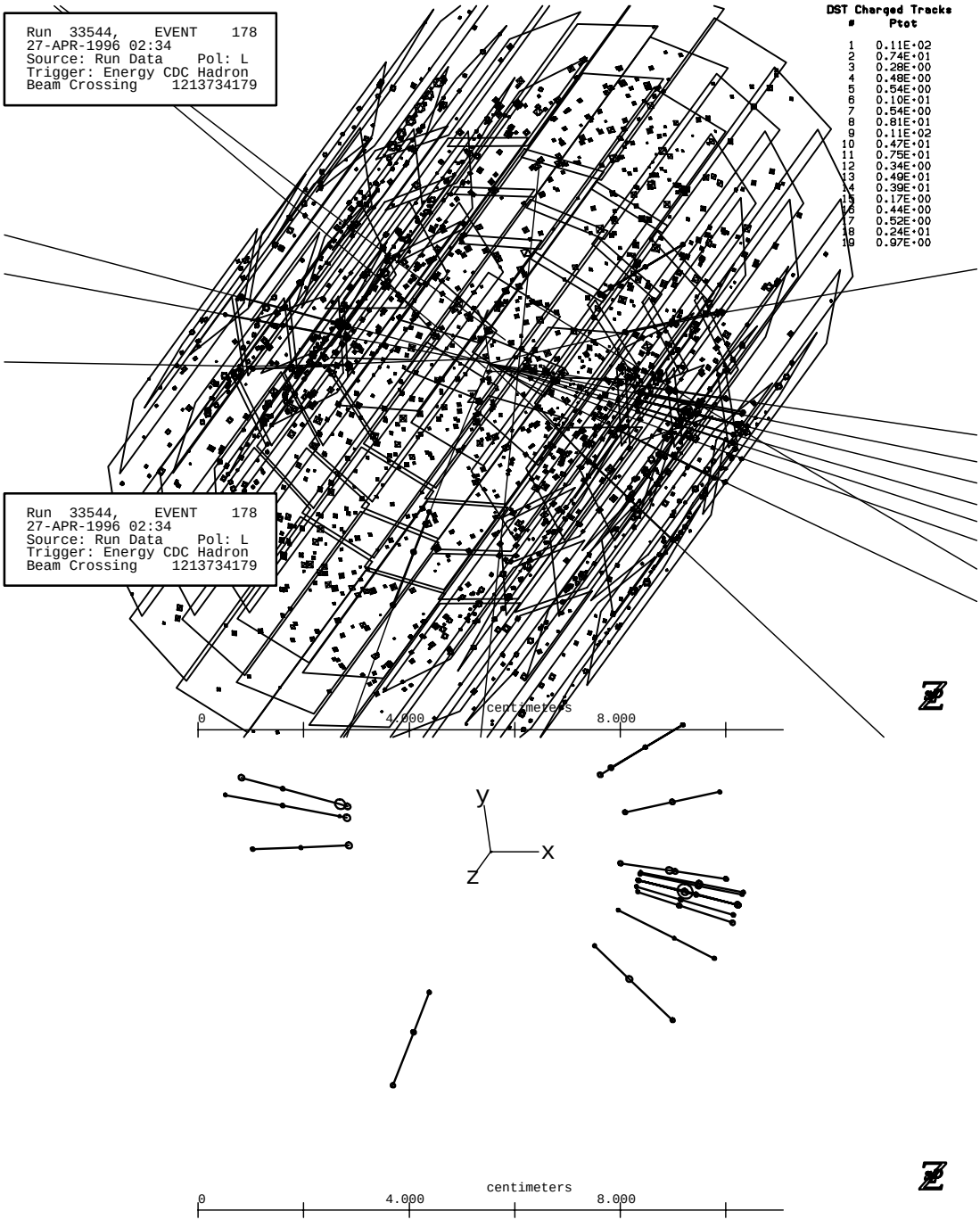


Figure 2.11: The VXD record of an hadronic event. Diamonds on the CCD-s represent energy clusters that passed the standard thresholds. The stand-alone vectors are separately plotted in the bottom portion of the picture. They are independently reconstructed using the the VXD information alone, and the requirement that the extrapolated vector hits cross sphere 1 cm in radius centered at the IP.

number of hits per CCD n, the number of vector hits that can be formed without any constraints is roughly n^3, assuming a vector hit links 3 clusters in the average. A number of constraints can be imposed to reduce the total number of possible vector hits. Asymptotically however, the search will still run roughly as n^3.

To solve the problem of the asymptotic behavior of the search routines, and to address a number of other issues in connection with the efficiency of the reconstructions, we developed a radically new solution.

The first step was to organize hits in a CCD as two-dimensional binary search trees. This reduces the search time of the vector hits to about $\mathcal{O}(n(\log_2 n)^2)$. Next, the representation of the arrangement of the CCD-s in the VXD has been decided to be in a form of an undirected graph, which in a very elegant way takes into account all the complexity of the arrangement of the CCD-s, like the overlapping of the CCD-s in the same layer and the similar. The net effect of combining the graph structure with the 2D binary search trees in a single CCD is "towering" of the portions of the CCD-s in either different layers or in the same layer when they overlap. This reduced the asymptotic behavior of the pattern search to roughly $\mathcal{O}(n(\log_2 n)^2)$.

With a very preliminary version of the code we achieved efficiency of about 95% and purity of about 80% in reconstructing vector hits with at least 3 clusters. The bottom portion of Fig. 2.11 shows collection of reconstructed vector hits in a hadronic event, which can be compared with the extrapolated tracks from the drift chamber (CDC) in the upper portion of the figure.

We also experimented with a number of data structures as substrates for variable event topologies in that the number of track attachments per vertex and the number of vertices were also dynamical quantities. In particular we studied graphs and binary heaps, where the nodes represented the vector hits, and links between the nodes attachments of tracks to their point of origin (physical vertices). The combinatorial set of all the spanning trees representing the potential event topologies is very large. An encouraging preliminary result is that these sets can be examined very quickly without the need to introduce any offhand cuts whatsoever. The work has not been developed from that point on.

3

The analysis method

The two branching fractions, $\mathcal{B}_L$ and $\mathcal{B}_U$, are in this analysis estimated from probability relations that make use of the correlation between the final state lepton charge and that of its parent b-quark. Since the highly energetic primary $b\bar{b}$ pairs at the Z tend to fragment into fairly narrow, back-to-back jets, the $b\bar{b}$ events are usually divided into two hemispheres, each hemisphere presumably containing only one of the two primary quarks and most of its fragmentation products. Therefore the initial state b-quark and the final state lepton, the suspected product of the b-quark decay, are tagged in the same hemisphere. The initial state b-quark is indirectly tagged by tagging the flavor of the b-quark in the *opposite hemisphere*, which is done by first using the invariant mass tag to identify the $b\bar{b}$ event, then by combining the jet charge variables with the variables in the polarized $e^+e^- \to b\bar{b}$ scattering (which are hemisphere unrelated) to identify the flavor. The probabilities of the initial state flavor and the final state leptons are then related to $\mathcal{B}_L$ and $\mathcal{B}_U$ through the set of probability relations, the derivation of which is the main subject of this chapter.

3.1 Topology of $b\bar{b}$ events and the outline of the method

An event identified as $b\bar{b}$ is shown in Fig. 3.1, featuring the characteristic topology of narrow back-to-back streams of particles. A quantity known as *thrust*,

$$T = 2\,\frac{\sum_i (\mathbf{p}_i \cdot \mathbf{t})^2}{\sum_i \mathbf{p}_i^2} - 1\,,$$

is sometimes used to quantify these topologies, where $\mathbf{t}$, the thrust axis, is a unit vector that maximizes the expression, $\mathbf{p}_i$ is momentum of a single track, and the sum is over all tracks in the event. The thrust is designed to be ≈ 1 for back-to-back topologies and ≈ 0 for isotropic topologies. The sum is usually over all charged tracks, the momenta of which are accurately measured by the CDC. The sum can be extended to include momenta of the neutral particles, which are indirectly estimated from the variables of the energy cluster in the LAC modules. Their momenta are far less accurately known than the momenta of the charged particles.

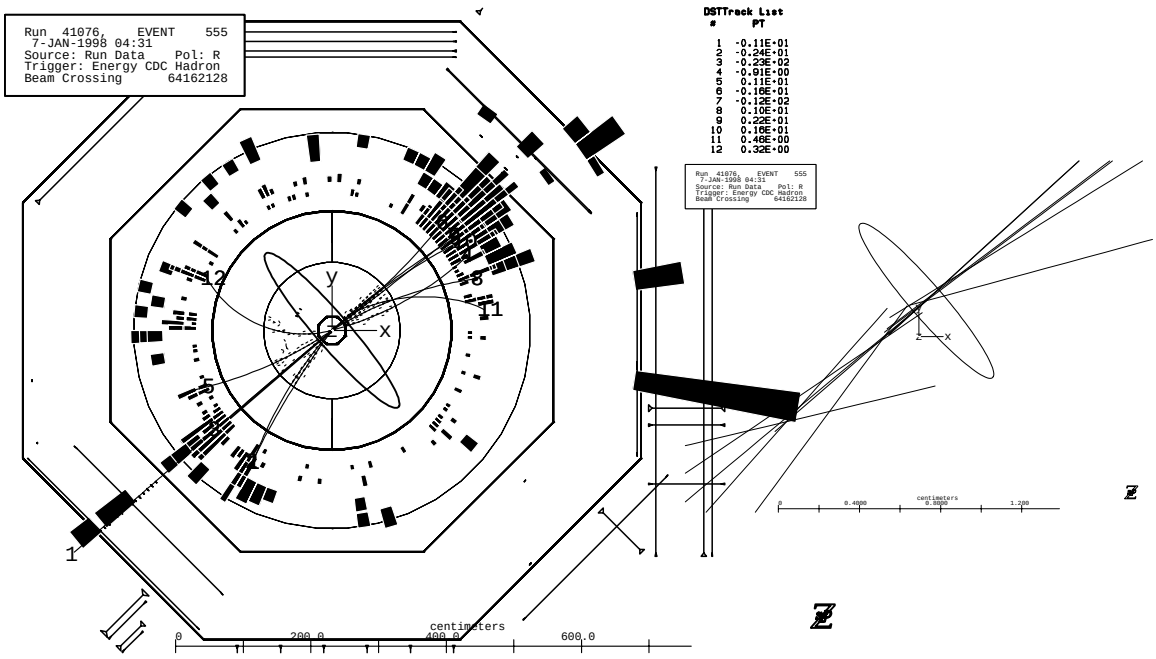


Figure 3.1: A "snapshot" of an event seen as $b\bar{b}$, featuring a 23.6 GeV/c track tagged as μ^- at around 8 o'clock. The opposite hemisphere (at around 2 o'clock') is tagged using the invariant mass tag. Ellipse centered at the IP represents the hemisphere plane.

Topologies of the $b\bar{b}$ events, like one in Fig. 3.1, allow us to assume that a b quark and almost all its decay products are contained in one hemisphere. The assumption holds even for more isotropic topologies or 3-jet events like one in Fig. 3.2. The correctness of our assumption about the $b\bar{b}$ event topologies will be indirectly confirmed by comparing the estimated values of $\mathcal{B}(b\to l)$ and $\mathcal{B}(b\to c\to l)$ with their true 4π MC values in Ch. 6. We will see there that the estimators produce a negative bias of around 2% for $\mathcal{B}(b\to l)$ and around 4% for $\mathcal{B}(b\to c\to l)$.

The flavor of the b quark is indirectly tagged by tagging the flavor of the b quark in the opposite hemisphere. This is done by first using the invariant mass tag in the opposite hemisphere, to estimate the $b\bar{b}$ event probability. The *opposite* hemisphere invariant mass tag is needed to avoid large correlations between the final state lepton sign and the flavor of the tagged b. The flavor of the b hadron is determined using the jet charge in the *opposite* hemisphere, and from the variables in the polarized $e^+e^-\to b\overline{b}$ scattering, as will be described in Ch. 5. As with the invariant mass tag, jet charge is measured in the opposite hemisphere to avoid large correlations between the signs of the final state leptons and the jet charge from the same hemisphere.

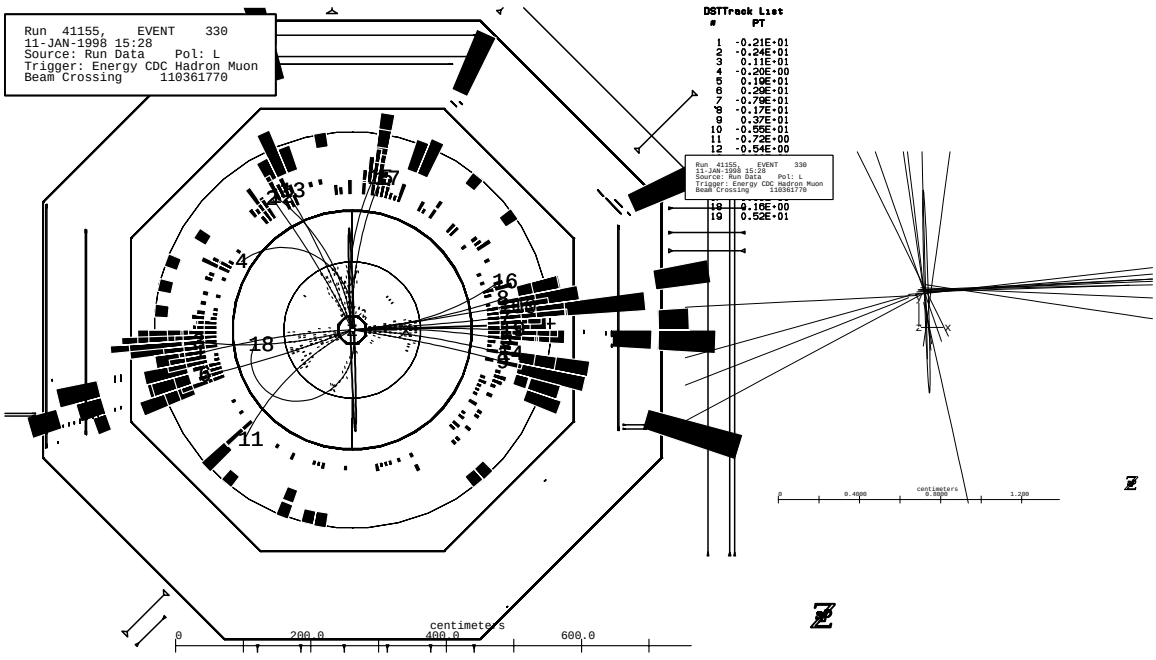


Figure 3.2: A 3-jet event identified as $b\bar{b}g$. Hemisphere at around 3 o'clock is tagged by the invariant mass tag. Our probability relations do not take into account events like this, which is what appears to be causing some problems when yields of pairs of leptons in the hemisphere are calculated (see the text, Sec. 3.2.2, Sec. 3.2.3, and parts of Ch. 6 in which the main results are cross-checked using pairs of leptons in the same hemisphere).

By using the opposite hemisphere tag one also simplifies handling of events that involve high energy gluon radiation, which sometimes have both primary quarks contained in a single hemisphere. In that case, the gluon hemisphere, which rarely has a displaced vertex, is almost never tagged. Therefore complicated topologies with two b quarks per hemisphere are avoided. If, on the other hand, the two quark hemisphere is tagged, the opposite hemisphere, the gluon one, will only contain tagged leptons whose sign is completely uncorrelated to the sign of the b tag, and only contribute to the random background.

For the rest of this chapter, we assume that probability of the initial state b-quark flavor, the lepton efficiencies, and the lepton mis-ID rates are known. They will be evaluated in Ch. 4 and Ch. 5.

3.2 Derivations of the probability relations

In the limit of precisely known flavor of the parent b quark, calculation of the branching fractions is trivial: one only has to divide the number of like (unlike) sign leptons with the number of the parent b quarks to get the $\mathcal{B}_L$ ($\mathcal{B}_U$). The like and unlike signs are relative to the sign of the parent b quark.

In practice the probability of the parent b quark flavor is not 1 (or 0). In our analysis it is determined from the probability of the *initial* state b quark flavor, which is directly measured, and from the value of the integrated mixing χ, which is taken from other experiments. The origins of the final states are also not precisely known; they could be leptons from b decays, background leptons, or hadrons misidentified as leptons.

Other constraints, such as known p and $p_\perp$ distributions of final state leptons, can also be used to separate $\mathcal{B}_L$ and $\mathcal{B}_U$. It turns out, however, that our method, which uses the charge correlation *alone*, separates $\mathcal{B}_L$ and $\mathcal{B}_U$ as well as other techniques. Statistical errors turn out to be approximately the same, while our systematic uncertainties are estimated to be somewhat lower than in other approaches.

Throughout the document, we will use somewhat more mathematical notation, not standard in HEP. This is needed to obtain correct probability relations, and to show how are lepton counts related to the probabilities. We will also see that the notation appears natural in the context of hypothesis testing in Ch. 4.

The probability that an event tagged by the invariant mass tag is $b\bar{b}$ will be denoted $f(b\bar{b})$. Strictly, the quantity should read $f(b\bar{b} \mid \text{mvtx} > 2\ \text{GeV})$ for an invariant mass cut of 2 GeV, stressing the conditional probability.

Variables used in the b *flavor* tagging will be collectively denoted

$$\text{B} \equiv (\text{mvtx}, Q, P_e, \cos\theta), \tag{3.1}$$

where mvtx is the invariant mass tag, Q the jet charge, P_e polarization of the incident electrons, and $\cos\theta$ the thrust axis polar angle. The probability that the primary initial state quark is b will be denoted $f(b \mid \text{B})$. Clearly

$$f(b \mid \text{B}) + f(\bar{b} \mid \text{B}) = f(b\bar{b}) \leq 1\,.$$

Positive outcomes of the lepton tests will be denoted E for electrons, M for muons, and L for either electrons or muons. Electron hypothesis will be denoted e, electron null hypothesis non-e. *Efficiency* $f(E \mid e)$ is defined as probability that a

"true" electron is accepted by the electron test, *mis-ID probability* $f(E \mid \text{non-}e)$ as probability that a non-electron is accepted by the electron test, and *purity* $f(e \mid E)$ as probability that a track tagged by the test is "true" electron. Clearly, the values of these quantities can only be established from the MC (or any other model that provides a mean of counting electrons). Everything goes the same for muons. The electron and muon tests themselves will be more of a subject in Ch. 4.

The *lepton tag multiplicity* in the b-tagged hemisphere is given by

$$\frac{n(L,\mathrm{B})}{n(\mathrm{B})} = 0 \cdot P_0 + 1 \cdot P_1 + 2 \cdot P_2 + \cdots , \tag{3.2}$$

where $n(L,\mathrm{B})$ is the number of tracks tagged as leptons in the b-tagged hemisphere, $n(\mathrm{B})$ the number of b-tagged hemispheres, and P_k probability that k (and only k) tracks are identified as leptons in the hemisphere. In (3.2) L indicates leptons that are tagged in a *hemisphere*. Individual probabilities P_k turn out to be very complicated functions of the physical branching fractions, efficiencies, and mis-ID probabilities. This will be discussed in more details in Sec. 3.4.

A simple way around is to do the calculations in individual bins, say in track momentum, that are chosen narrow enough that one can safely assume that $P_2, P_3, \cdots \approx 0$ in each individual bin. In i-th bin therefore,

$$f_i(L \mid \mathrm{B}) = \frac{n_i(L,\mathrm{B})}{n(\mathrm{B})} \approx 0 \cdot P_{0i} + 1 \cdot P_{1i} \tag{3.3}$$

is *lepton tag probability*, which can be expressed as a sum of probabilities for the lepton coming from various sources: like and unlike sign b decays, backgrounds, or the mis-ID leptons. In (3.3) L indicates that leptons are tagged in a *momentum bin*. Strictly, L in (3.2) and (3.3) is a random variable on $\{0,1\}$ indicating whether there are any tracks tagged as leptons in the hemisphere, or in the momentum bin.

A thing to note is that $f_i(L \mid \mathrm{B})$ is marginal probability with respect to a whole host of other variables. We introduce for a moment a slightly different notation to illustrate the point. Let L_i and L_j indicate that tracks labeled as leptons are observed in momentum bins i and j. Then

$$f(L_i \mid \mathrm{B}) = \sum_j [\, f(L_i, L_j \mid \mathrm{B}) + f(L_i, \bar{L}_j \mid \mathrm{B}) \,]$$

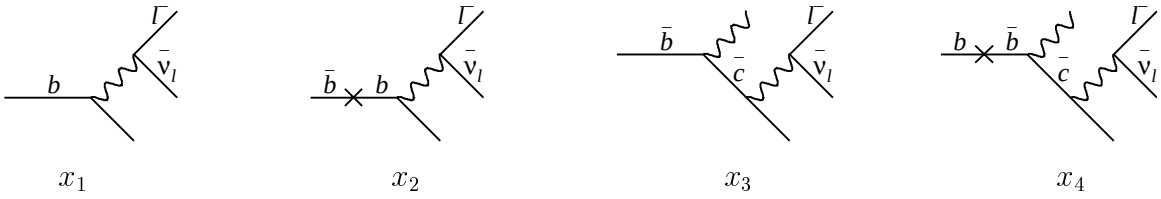
where the sum is over all momentum bins of the second track. Similarly $f(L_i, L_j \mid \mathrm{B}) = \sum_k [\, f(L_i, L_j, L_k \mid \mathrm{B}) + f(L_i, L_j, \bar{L}_k \mid \mathrm{B}) \,]$, and so on. The consequence for the

counting is that in order to correctly estimate $f_i(L \mid \mathrm{B})$ one has to count *every* lepton falling within bin i regardless of whether it is accompanied by another lepton or not.

A track that is tagged as a lepton is either a true lepton or is not, so in i-th momentum bin

$$\begin{aligned} f_i(L \mid \mathrm{B}) &= f_i(L, l \mid \mathrm{B}) + f_i(L, \text{non-}l \mid \mathrm{B}) \\ &= f_i(L \mid l, \mathrm{B}) f_i(l \mid \mathrm{B}) + f_i(L, \text{non-}l \mid \mathrm{B}) \\ &\equiv \eta_i f_i(l \mid \mathrm{B}) + \delta_i \qquad (3.4) \\ &\equiv \text{efficiency}_i \times (\text{lepton yield probability})_i + (\text{mis-ID rate})_i \,, \end{aligned}$$

where the efficiency and the mis-ID rate have been factored out. The lepton yield probability is the sum of probabilities for a lepton yield from either prompt and cascade decays of b hadrons, or from various background sources like γ, semileptonic decays of $c\bar{c}$ events and other.



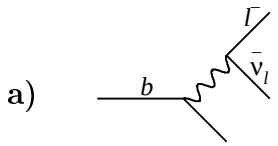
x_1 x_2 x_3 x_4

label	process	probability
x_1	$b\,, b{\to}b\,, b{\to}l^-$	$\eta_i\, \mathcal{B}_{L,i}\, (1-\chi_L)\, f(b \mid \mathrm{B})$
x_2	$\bar{b}\,, \bar{b}{\to}b\,, b{\to}l^-$	$\eta_i\, \mathcal{B}_{L,i}\, \chi_L\, f(\bar{b} \mid \mathrm{B})$
x_3	$\bar{b}\,, \bar{b}{\to}\bar{b}\,, \bar{b}{\to}\bar{c}{\to}l^-$	$\eta_i\, \mathcal{B}_{U,i}\, (1-\chi_U)\, f(\bar{b} \mid \mathrm{B})$
x_4	$b\,, b{\to}\bar{b}\,, \bar{b}{\to}\bar{c}{\to}l^-$	$\eta_i\, \mathcal{B}_{U,i}\, \chi_U\, f(b \mid \mathrm{B})$
x_5	bkgs + mis-ID l^-	$\eta_i\, \beta_i + \delta_i$

Figure 3.3: Processes yielding a negatively charged lepton in the final state. The sign of the final state lepton is not maximally correlated to the sign of the initial state b quark due to the mixing ($\times$.)

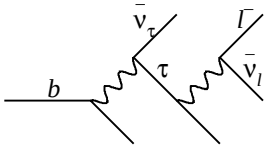
3.2.1 The master equation

Probabilities $f_i(L^- \mid \mathrm{B})$ and $f_i(l^- \mid \mathrm{B})$ are therefore marginal probabilities. Processes contributing to $f_i(l^- \mid \mathrm{B})$ are listed in Table 3.3; $x_1, \cdots, x_4$ label b semileptonic decays, while x_5 labels all processes contributing to the combined background + mis-ID rate $\eta_i\,\beta_i + \delta_i$. The efficiency η_i is included in these probabilities. The background and the misidentified leptons are considered to be uncorrelated to the b flavor. As a consequence, the $\eta_i\,\beta_i + \delta_i$ is treated as a single quantity, which we shall show can be independently estimated using events with pairs of leptons in a hemisphere.



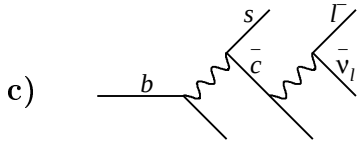
dominant semileptonic modes

$B^- \to D^0 l^- \overline{\nu}_l$	$(1.86 \pm 0.33)\%$
$B^- \to D^*(2007)^0 l^- \overline{\nu}_l$	$(5.3 \pm 0.8)\%$
$\overline{B}^0 \to D^+ l^- \overline{\nu}_l$	$(2.00 \pm 0.25)\%$
$\overline{B}^0 \to \overline{D}^*(2010)^+ l^- \overline{\nu}_l$	$(4.60 \pm 0.27)\%$



$b = B^\pm/B^0/B_s^0/b$-baryon admixture

$b \to \tau^- \overline{\nu}_\tau$ anything	$(2.6 \pm 0.4)\%$
$\hookrightarrow \mu^- \overline{\nu}_\mu \nu_\tau$	$(17.37 \pm 0.09)\%$
$\hookrightarrow \mu^- \overline{\nu}_\mu \nu_\tau \gamma$	$(3.0 \pm 0.6)\%$
$\hookrightarrow e^- \overline{\nu}_e \nu_\tau$	$(17.81 \pm 0.07)\%$



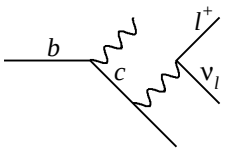
$b = B^\pm/B^0/B_s^0/b$-baryon admixture

$b \to D_s^-$ anything	$(12.11 \pm 1.68)\%$ [65]
$\hookrightarrow e^-$anything	$(8\,^{+\,6}_{-\,5})\%$
$\hookrightarrow \varphi l^- \overline{\nu}_l$	$(2.0 \pm 0.5)\%$
$\hookrightarrow \eta l^- \overline{\nu}_l + \eta' l^- \overline{\nu}_l$	$(3.4 \pm 1.0)\%$

Figure 3.4: The parton level and the dominant hadronic level processes yielding a lepton of the *same sign* as that of its parent b quark. These processes cannot be experimentally distinguished within our method; their joint branching fraction is referred to as $\mathcal{B}_L$. The complete list of the measured inclusive and exclusive modes is given in [7], pp. 533-579.

An important note: process x_1 (and similarly process x_2) in Fig. 3.3 represents all processes yielding a final state lepton of the same sign as that of the parent b

quark (at the decay vertex), Fig. 3.4.



$\underline{b = B^{\pm}/B^0/B_s^0/b\text{-baryon admixture}}$

$$
b \to \begin{cases}
D^+ \text{ anything} \quad (23.7 \pm 2.3)\% \\
\quad \hookrightarrow \begin{cases}
\overline{K}^0 l^+ \nu_l & (6.8 \pm 0.8)\% \\
K^- \pi^+ e^+ \nu_e & (4.2 \,{}^{+0.9}_{-0.7})\% \\
K^- \pi^+ \mu^+ \nu_\mu & (3.2 \pm 0.4)\%
\end{cases} \\
D^0 \text{ anything} \quad (60.1 \pm 3.2)\% \\
\quad \hookrightarrow \begin{cases}
K^- l^+ \nu_l & (3.50 \pm 0.17)\% \\
K^- \pi^0 e^+ \nu_e & (1.6 \,{}^{+1.3}_{-0.5})\% \\
\overline{K}^0 \pi^- e^+ \nu_e & (2.8 \,{}^{+1.7}_{-0.9})\% \\
\pi^- e^+ \nu_e & (3.7 \pm 0.6)\%
\end{cases} \\
\Lambda_c^+ \text{ anything} \quad (9.7 \pm 2.9)\% \\
\quad \hookrightarrow \begin{cases}
e^+ \text{anything} & (4.5 \pm 1.7)\% \\
p e^+ \text{anything} & (1.8 \pm 0.9)\%
\end{cases}
\end{cases}
$$

Figure 3.5: The parton level and some of the hadronic level processes yielding a cascade lepton of the *sign opposite* to that of its parent b quark. The branching fraction of these processes is referred to as $\mathcal{B}_U$. The complete list of the measured exclusive and inclusive modes is given in [7], pp. 486-551 for D mesons, and pp. 729-733 for Λ_c.

Some of the cascade semileptonic decays are listed in Fig. 3.5. The number of hadronic modes is large due to a large number of combinations between the initial state B and the intermediate state D varieties. This composition of exclusive decay modes "creeps into" our calculations, and is seen through the different values of the integrated mixing for the like sign, χ_L, and unlike sign processes, χ_U, as described in Sec. 1.4.

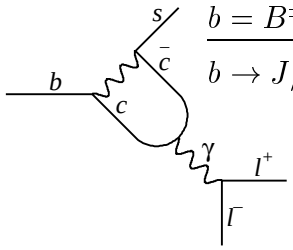
Finally, the sources of background leptons are listed in Fig. 3.6. They are all considered to be uncorrelated to the flavor of the tagged b hadron, even when they actually do originate in a b hadron decay, such as in J/ψ decays.

Conversions
$\pi^0 \to \gamma\,\gamma \quad (98.798 \pm 0.032)\%$
$\quad\hookrightarrow e^+\,e^-$ (dominant)
$\eta \to \gamma\,\gamma \quad (39.21 \pm 0.34)\%$
$\quad\hookrightarrow e^+\,e^-$

Decays
$\pi^0 \to e^+\,e^-\,\gamma \qquad (1.198 \pm 0.032)\%$

$\mu^- \to e^-\,\overline{\nu}_e$
$K^- \to e^-\,\overline{\nu}_e$
$K^0_L \to e^-\,\overline{\nu}_e$ X

$\pi^- \to \mu^-\,\overline{\nu}_\mu$
$K^- \to \mu^-\,\overline{\nu}_\mu$
$K^0_L \to \mu^-\,\overline{\nu}_\mu$ X



$b = B^\pm/B^0/B^0_s/b$-baryon admixture
$b \to J/\psi(1S)$ anything $\quad (1.16 \pm 0.10)\%$
$\hookrightarrow \begin{cases} e^+\,e^- & (6.02 \pm 0.19)\% \\ \mu^+\,\mu^- & (6.01 \pm 0.19)\% \end{cases}$

Figure 3.6: The background sources. For electrons these are mostly γ conversions to electron positron pairs and from Dalitz decays of π^0. There is also some small contribution from $c\bar{c}$ events. The main background source for muons is from the $c\bar{c}$ events as well as from the π and K decays. Some b decays, like that of a J/ψ, [7], pp. 486-551, also produce leptons that are uncorrelated to the flavor of their parent b-quark, and effectively behave as the background. The misidentified leptons are all considered uncorrelated to the flavor of the tagged b-quark.

Eq. (3.4) therefore translates into

$$
\begin{aligned}
f_i(L^- \mid \mathrm{B}) &= \eta_i\, f_i(l^- \mid \mathrm{B}) + \delta_i \equiv \sum_{a=1}^{5} f_i(x_a \mid \mathrm{B}) \qquad (3.5)\\
&= \eta_i \left(f_L(b \mid \mathrm{B})\mathcal{B}_{L,i} + f_U(\bar{b} \mid \mathrm{B})\mathcal{B}_{U,i}\right) + \eta_i\beta_i + \delta_i \\
f_L(b \mid \mathrm{B}) &= (1-\chi_L) f(b \mid \mathrm{B}) + \chi_L f(\bar{b} \mid \mathrm{B}) \\
f_U(\bar{b} \mid \mathrm{B}) &= \chi_U f(b \mid \mathrm{B}) + (1-\chi_U) f(\bar{b} \mid \mathrm{B})
\end{aligned}
$$

where $f_L(b \mid \mathrm{B})$ and $f_U(\bar{b} \mid \mathrm{B})$ are probabilities that the parent quark (at the vertex) is b or $\bar{b}$. Also, the momentum dependent efficiency η_i is different for $\mathcal{B}_{L,i}$, $\mathcal{B}_{U,i}$, and

β_i. The difference in η_i between $\mathcal{B}_{L,i}$ and $\mathcal{B}_U$i is about 2% and is taken into account in our calculations. The efficiency for the background processes is not separately estimated, but is absorbed in $\eta_i\beta_i$ instead, which is estimated in the "bulk".

For the charge conjugated processes one equivalently has

$$\begin{aligned} f_i(L^+ \mid \mathrm{B}) &= \eta_i \left(f_L(\bar{b} \mid \mathrm{B})\mathcal{B}_{L,i} + f_U(b \mid \mathrm{B})\mathcal{B}_{U,i}\right) + \eta_i\beta_i + \delta_i \qquad (3.6) \\ f_L(\bar{b} \mid \mathrm{B}) &= (1-\chi_L) f(\bar{b} \mid \mathrm{B}) + \chi_L f(b \mid \mathrm{B}) \\ f_U(b \mid \mathrm{B}) &= \chi_U f(\bar{b} \mid \mathrm{B}) + (1-\chi_U) f(b \mid \mathrm{B}) \end{aligned}$$

The two can be jointly written as

$$\begin{pmatrix} f_i(L^- \mid \mathrm{B}) \\ f_i(L^+ \mid \mathrm{B}) \end{pmatrix} = \eta_i \begin{pmatrix} f_L(b \mid \mathrm{B}) & f_U(\bar{b} \mid \mathrm{B}) \\ f_L(\bar{b} \mid \mathrm{B}) & f_U(b \mid \mathrm{B}) \end{pmatrix} \begin{pmatrix} \mathcal{B}_{L,i} \\ \mathcal{B}_{U,i} \end{pmatrix} + \begin{pmatrix} \eta_i\beta_i + \delta_i \\ \eta_i\beta_i + \delta_i \end{pmatrix}, \qquad (3.7)$$

where the matrix of the parent b and $\bar{b}$ quark probabilities is called the *correlation* or the *transition* matrix. The $\eta_i\beta_i + \delta_i$ is assumed equal for both positively and negatively charged tracks. Clearly, in order to solve the system for $\mathcal{B}_{L,i}$ and $\mathcal{B}_{U,i}$, the terms $\eta_i\beta_i + \delta_i$ have to be known. It this analysis the $\eta_i\beta_i + \delta_i$ are determined using two independent techniques.

The first technique, Ch. 4, uses $> 99.9\%$ pure hadronic samples obtained from K_s and Λ decays to calibrate the MC values of δ_i to the data. This applies equally to both electrons and muons. The $\eta_i\beta_i$ for electrons is calibrated from a sample of $\approx 90\%$ pure electrons from γ conversions. For muons, the $\eta_i\beta_i$ is taken directly from the Monte Carlo.

A truly nice alternative to that technique is to independently cross-check the value of the bulk quantity $\sum_i(\eta_i\beta_i + \delta_i)$, by counting the number of pairs of leptons in the same hemisphere. The pairs could be of either like or unlike sign. [The reason for not cross-checking the $\eta_i\beta_i + \delta_i$ bin-by-bin is because of the relatively low statistics, which can only be increased by simultaneously working with both lepton species, which is far more complicated than handling them separately.] Clearly, when tracks in a pair are of like sign, at least one of them is either a background or a mis-ID. When tracks are of unlike sign, additional possibility arises that the two originate in simultaneous decays of a single b-quark. Probability relations for these configurations are straightforward to derive by summing entries from Table 3.1, pretty much the same as in the derivation of the master equation, Eq. (3.7).

label	process	probability
z_1	b, $b\to b$, $b\to cl^-\to l^-l^+$	$\eta_i\eta_j\mathcal{B}_{L,i}\,\mathcal{B}_{U,j}\,(1-\chi)\,f(b \mid \mathrm{B})$
z_2	$\bar{b}$, $\bar{b}\to b$, $b\to cl^-\to l^-l^+$	$\eta_i\eta_j\mathcal{B}_{L,i}\,\mathcal{B}_{U,j}\,\chi\,f(\bar{b} \mid \mathrm{B})$
$\bar{z}_1$	$\bar{b}\,, \bar{b}\to\bar{b}\,, \bar{b}\to\bar{c}l^+\to l^-l^+$	$\eta_i\eta_j\mathcal{B}_{U,i}\mathcal{B}_{L,j}(1-\chi)f(\bar{b} \mid \mathrm{B})$
$\bar{z}_2$	$b\,, b\to\bar{b}\,, \bar{b}\to\bar{c}l^+\to l^-l^+$	$\eta_i\eta_j\mathcal{B}_{U,i}\mathcal{B}_{L,j}\chi f(b \mid \mathrm{B})$
$x_1\bar{x}_5$	$b\,, b\to b\,, b\to l^-$ and bkg or mis-ID l^+	$\eta_i\mathcal{B}_{L,i}(1-\chi)f(b \mid \mathrm{B})(\eta_j\beta_j+\delta_j)$
$x_2\bar{x}_5$	$\bar{b}\,, \bar{b}\to b\,, b\to l^-$ and bkg or mis-ID l^+	$\eta_i\mathcal{B}_{L,i}\chi f(\bar{b} \mid \mathrm{B})(\eta_j\beta_j+\delta_j)$
$x_3\bar{x}_5$	$\bar{b}\,, \bar{b}\to\bar{b}\,, \bar{b}\to\bar{c}\to l^-$ and bkg or mis-ID l^+	$\eta_i\mathcal{B}_{U,i}(1-\chi)f(\bar{b} \mid \mathrm{B})(\eta_j\beta_j+\delta_j)$
$x_4\bar{x}_5$	$b\,, b\to\bar{b}\,, \bar{b}\to\bar{c}\to l^-$ and bkg or mis-ID l^+	$\eta_i\mathcal{B}_{U,i}\chi f(b \mid \mathrm{B})(\eta_j\beta_j+\delta_j)$
$x_5\bar{x}_1$	bkg or mis-ID l^- and $b\,, b\to b\,, b\to l^+$	$(\eta_i\beta_i+\delta_i)\eta_j\mathcal{B}_{L,j}(1-\chi)f(b \mid \mathrm{B})$
$x_5\bar{x}_2$	bkg or mis-ID l^- and $b\,, b\to\bar{b}\,, \bar{b}\to l^+$	$(\eta_i\beta_i+\delta_i)\eta_j\mathcal{B}_{L,j}\chi f(b \mid \mathrm{B})$
$x_5\bar{x}_3$	bkg or mis-ID l^- and $b\,, b\to b\,, b\to c\to l^+$	$(\eta_i\beta_i+\delta_i)\eta_j\mathcal{B}_{U,j}(1-\chi)f(b \mid \mathrm{B})$
$x_5\bar{x}_4$	bkg or mis-ID l^- and $\bar{b}\,, \bar{b}\to b\,, b\to c\to l^+$	$(\eta_i\beta_i+\delta_i)\eta_j\mathcal{B}_{U,j}\chi f(\bar{b} \mid \mathrm{B})$
$x_5\bar{x}_5$	bkg or mis-ID l^- and l^+	$(\eta_i\beta_i+\delta_i)(\eta_j\beta_j+\delta_j)$

Table 3.1: Processes yielding an l^- in i-th momentum bin and an l^+ in j-th momentum bin. Processes in which a b hadron decays simultaneously into two final state leptons through $b\to l$ and $b\to c\to l$ branches are labeled z_1 and z_2. The lepton charges in these processes are maximally correlated, and their probability cannot be expressed as a product of the probabilities of the two statistically independent processes. Processes involving at least one background or a mis-ID lepton are on the other side considered statistically independent with respect to all other processes. The probabilities of such pairs are just the products of the probabilities for the individual processes. An important note: probabilities for all these processes are marginal probabilities with respect to any other variables.

3.2.2 Pairs of unlike sign leptons

Let $\tilde{\mathcal{B}}_{L,i}=\eta_i\mathcal{B}_{L,i}$, $\tilde{\mathcal{B}}_L=\sum_i\tilde{\mathcal{B}}_{L,i}$, $\tilde{\mathcal{B}}_{U,i}=\eta_i\mathcal{B}_{U,i}$, $\tilde{\mathcal{B}}_U=\sum_i\tilde{\mathcal{B}}_{U,i}$, $\tilde{\beta}_i=\eta_i\beta_i$, $\tilde{\beta}=\sum_i\tilde{\beta}_i$, and $\delta=\sum_i\delta_i$. Let i be a fixed (momentum) bin containing a *negatively* charged lepton, and j a running index of a bin containing a *positively* charged companion. Then the probability for such a configuration is the sum of the probabilities from Table 3.1 over all processes that can contribute to the configuration. The probability that the negatively charged lepton in i-th bin is accompanied with a positively charged lepton *anywhere* in the momentum range, is then obtained by the integration over j:

$$
\begin{aligned}
& f_i^-(L^- \mid \mathrm{B})_\pm \\
&= \sum_j [\sum_{a=1,2} (f_{ij}(z_a) + f_{ij}(\bar{z}_a)) + \sum_{a=1,4} (f_i(x_a) f_j(\bar{x}_5) + f_i(x_5) f_j(\bar{x}_a))\,] \\
&\quad + \sum_j f_i(x_5) f_j(\bar{x}_5) \\
&= (f_p(b \mid \mathrm{B}) \tilde{\mathcal{B}}_{L,i} \tilde{\mathcal{B}}_U + f_p(\bar{b} \mid \mathrm{B}) \tilde{\mathcal{B}}_{U,i} \tilde{\mathcal{B}}_L) \\
&\quad + (f_p(b \mid \mathrm{B}) \tilde{\mathcal{B}}_{L,i} + f_p(\bar{b} \mid \mathrm{B}) \tilde{\mathcal{B}}_{U,i})(\tilde{\beta} + \delta) \\
&\quad + (\tilde{\beta}_i + \delta_i)\,(f_p(b \mid \mathrm{B}) \tilde{\mathcal{B}}_L + f_p(\bar{b} \mid \mathrm{B}) \tilde{\mathcal{B}}_U) + (\tilde{\beta}_i + \delta_i)(\tilde{\beta} + \delta)\,.
\end{aligned} \tag{3.8}
$$

where $f_p(b \mid \mathrm{B})$ and $f_p(\bar{b} \mid \mathrm{B})$ are now parent quark probabilities based on counting the pairs of leptons, and χ the weighted average between the χ_L and χ_U. In a very much the same manner, when i-th lepton bin is *positively* and the running j-th bin negatively charged (just conjugate of Table 3.1), one has

$$
\begin{aligned}
& f_i^+(L^+ \mid \mathrm{B})_\pm \\
&= \sum_{j,a=1,2} (f_{ij}(\bar{z}_a) + f_{ij}(z_a)) + \sum_{j,a=1,4} (f_i(\bar{x}_a) f_j(x_5) + f_i(\bar{x}_5) f_j(x_a)) + \sum_j f_i(\bar{x}_5) f_j(x_5) \\
&= (f_p(\bar{b} \mid \mathrm{B}) \tilde{\mathcal{B}}_{L,i} \tilde{\mathcal{B}}_U + f_p(b \mid \mathrm{B}) \tilde{\mathcal{B}}_{U,i} \tilde{\mathcal{B}}_L) \\
&\quad + (f_p(\bar{b} \mid \mathrm{B}) \tilde{\mathcal{B}}_{L,i} + f_p(b \mid \mathrm{B}) \tilde{\mathcal{B}}_{U,i})(\tilde{\beta} + \delta) \\
&\quad + (\tilde{\beta}_i + \delta_i)\,(f_p(\bar{b} \mid \mathrm{B}) \tilde{\mathcal{B}}_L + f_p(b \mid \mathrm{B}) \tilde{\mathcal{B}}_U) + (\tilde{\beta}_i + \delta_i)(\tilde{\beta} + \delta)
\end{aligned} \tag{3.9}
$$

by simply interchanging $f_p(b \mid \mathrm{B})$ and $f_p(\bar{b} \mid \mathrm{B})$ in Eq. (3.8). One can also sum (3.8) and (3.9) to get

$$
\begin{aligned}
f_i(L \mid \mathrm{B})_{\pm} &= \eta_i f(b\bar{b})\,(\tilde{\mathcal{B}}_{L,i}\tilde{\mathcal{B}}_U + \tilde{\mathcal{B}}_{U,i}\tilde{\mathcal{B}}_L) + f(b\bar{b})(\tilde{\mathcal{B}}_{L,i} + \tilde{\mathcal{B}}_{U,i})(\tilde{\beta} + \delta) \\
&\quad + f(b\bar{b})(\tilde{\mathcal{B}}_L + \tilde{\mathcal{B}}_U)(\tilde{\beta}_i + \delta_i) + 2(\tilde{\beta}_i + \delta_i)(\tilde{\beta} + \delta) \\
&= f(b\bar{b})(\tilde{\mathcal{B}}_{L,i}\tilde{\mathcal{B}}_U + \tilde{\mathcal{B}}_{U,i}\tilde{\mathcal{B}}_L) + [f(b\bar{b})(\tilde{\mathcal{B}}_{L,i} + \tilde{\mathcal{B}}_{U,i}) + 2(\tilde{\beta}_i + \delta_i)](\tilde{\beta} + \delta) \\
&\quad + [f(b\bar{b})(\tilde{\mathcal{B}}_L + \tilde{\mathcal{B}}_U) + 2(\tilde{\beta} + \delta)](\tilde{\beta}_i + \delta_i) - 2(\tilde{\beta}_i + \delta_i)(\tilde{\beta} + \delta) \\
&\equiv f(b\bar{b})(\tilde{\mathcal{B}}_{L,i}\tilde{\mathcal{B}}_U + \tilde{\mathcal{B}}_{U,i}\tilde{\mathcal{B}}_L) \\
&\quad + f_i(L \mid \mathrm{B})(\tilde{\beta} + \delta) + f(L \mid \mathrm{B})(\tilde{\beta}_i + \delta_i) - 2(\tilde{\beta} + \delta)(\tilde{\beta}_i + \delta_i) \qquad (3.10)
\end{aligned}
$$

where the last expression has been obtained by moving $2(\tilde{\beta}_i + \delta_i)$ and $2(\tilde{\beta} + \delta)$ around and by using $f_i(L \mid \mathrm{B}) = f(b\bar{b})(\mathcal{B}_{L,i} + \mathcal{B}_{U,i}) + 2(\tilde{\beta} + \delta)$. [Recall that $\tilde{\beta} + \delta = (\tilde{\beta} + \delta)^- = (\tilde{\beta} + \delta)^+$, and that when averaged over the initial state b and $\bar{b}$ flavors, the "branching fractions" for positively and negatively charged leptons *separately* are $(\tilde{\mathcal{B}}_L + \tilde{\mathcal{B}}_U)^+ = (\tilde{\mathcal{B}}_L + \tilde{\mathcal{B}}_U)^- = (\tilde{\mathcal{B}}_L + \tilde{\mathcal{B}}_U)/2$. This is all true under the assumption that leptons of both signs behave equally when they interact with the detector materials, which is in agreement with the observations.]

In principle, any of equations (3.8), (3.9), or (3.10) can be used to cross check the values of $\tilde{\beta}_i + \delta_i$. Due to the low statistics however, only the bulk value $\tilde{\beta} + \delta$ is cross-checked. The probability for a yield of a pair of oppositely charged tracks, both of which are tagged as leptons, is obtained by integrating either (3.8), (3.9), or Eq. (3.10), with a 1/2 factor because the pairs are counted twice. Integration of Eq. (3.8), for example, gives

$$
\begin{aligned}
f(L)_{\pm} = \text{number of oppositely charge pairs per } b\text{-tag} &= \sum_i \int f_i^-(L \mid \mathrm{B})_{\pm}\, f(\mathrm{B})\, d\mathrm{B} \\
= \sum_i \int &[(f_p(b \mid \mathrm{B})\tilde{\mathcal{B}}_{L,i}\tilde{\mathcal{B}}_U + f_p(\bar{b} \mid \mathrm{B})\tilde{\mathcal{B}}_{U,i}\tilde{\mathcal{B}}_L) \\
&+ (f_p(b \mid \mathrm{B})\tilde{\mathcal{B}}_{L,i} + f_p(\bar{b} \mid \mathrm{B})\tilde{\mathcal{B}}_{U,i})(\tilde{\beta} + \delta) \\
&+ (\tilde{\beta}_i + \delta_i)(f_p(b \mid \mathrm{B})\tilde{\mathcal{B}}_L + f_p(\bar{b} \mid \mathrm{B})\tilde{\mathcal{B}}_U) + (\tilde{\beta}_i + \delta_i)(\tilde{\beta} + \delta)]\, f(\mathrm{B})\, d\mathrm{B}
\end{aligned}
$$

$$= \frac{1}{2}(\tilde{\mathcal{B}}_L\tilde{\mathcal{B}}_U + \tilde{\mathcal{B}}_U\tilde{\mathcal{B}}_L) + \frac{1}{2}(\tilde{\mathcal{B}}_L + \tilde{\mathcal{B}}_U)(\tilde{\beta}+\delta) + \frac{1}{2}(\tilde{\beta}+\delta)(\tilde{\mathcal{B}}_L + \tilde{\mathcal{B}}_U) + (\tilde{\beta}+\delta)^2$$

$$= \tilde{\mathcal{B}}_L\tilde{\mathcal{B}}_U + f(\mathrm{L})\,(\tilde{\beta}+\delta) \;-\; (\tilde{\beta}+\delta)^2 \tag{3.11}$$

where $\int f_p(b \mid \mathrm{B})\, f(\mathrm{B})\, d\mathrm{B} = \int f_p(\bar{b} \mid \mathrm{B})\, f(\mathrm{B})\, d\mathrm{B} = 1/2$, because in the average one half of the parents are b and another half are $\bar{b}$. The $f(\mathrm{B})$ is the b-tag frequency ($\int f(\mathrm{B})\, d\mathrm{B} = 1$), and $f(\mathrm{L}) = \tilde{\mathcal{B}}_L + \tilde{\mathcal{B}}_U + 2(\tilde{\beta}+\delta)$.

Note that Eq. (3.10) is marginal probability with respect to processes with more than 2 leptons. If, for example, 3 tracks were identified as leptons in a hemisphere, one of which is, say negative, and other two positive, than the negative track will be counted twice in $\pm$ combinations, and these counts have to be divided with 2, while the $++$ pair is counted only once. Fortunately, the number of events with more than 2 leptons per hemisphere is negligible, which substantially simplifies the analysis.

3.2.3 Pairs of like sign leptons

A pair of like sign leptons can be created by all processes from Table 3.1 except by z_1 and z_2, and their conjugate processes $\bar{z}_1$ and $\bar{z}_2$. Therefore the probability relations are the same as in (3.8) and (3.9), only without the term containing products of $\tilde{\mathcal{B}}_L$ and $\tilde{\mathcal{B}}_U$:

$$f_i^-(L^- \mid \mathrm{B})_{--}$$

$$= \eta_i(f_p(b \mid \mathrm{B})\mathcal{B}_{L,i} + f_p(\bar{b} \mid \mathrm{B})\mathcal{B}_{U,i})(\tilde{\beta}+\delta) + (\eta_i\beta_i + \delta_i)(f_p(b \mid \mathrm{B})\tilde{\mathcal{B}}_L + f_p(\bar{b} \mid \mathrm{B})\tilde{\mathcal{B}}_U)$$

$$+(\eta_i\beta_i + \delta_i)(\tilde{\beta}+\delta) \tag{3.12}$$

$$f_i^+(L^+ \mid \mathrm{B})_{++}$$

$$= \eta_i(f_p(\bar{b} \mid \mathrm{B})\mathcal{B}_{L,i} + f_p(b \mid \mathrm{B})\mathcal{B}_{U,i})(\tilde{\beta}+\delta) + (\eta_i\beta_i + \delta_i)(f_p(\bar{b} \mid \mathrm{B})\tilde{\mathcal{B}}_L + f_p(b \mid \mathrm{B})\tilde{\mathcal{B}}_U)$$

$$+(\eta_i\beta_i + \delta_i)(\tilde{\beta}+\delta) \tag{3.13}$$

and their sum

$$f_i(L \mid \mathrm{B})_{\pm\pm} = f_i(L \mid \mathrm{B})(\tilde{\beta}+\delta) + f(L \mid \mathrm{B})(\eta_i\beta_i+\delta_i) - 2(\tilde{\beta}+\delta)(\eta_i\beta_i+\delta_i)\,. \quad (3.14)$$

The integration of (3.14) over i yields

$$\begin{aligned} f(\mathrm{L})_{\pm\pm} &= \text{ track multiplicity in equally charged pairs} \\ &= \frac{1}{2}\sum_i \int f_i(\mathrm{L} \mid \mathrm{B})_{\pm\pm}\, f(\mathrm{B})\, d\mathrm{B} \\ &= (\tilde{\beta}+\delta) \int f(\mathrm{L} \mid \mathrm{B})\, f(\mathrm{B})\, d\mathrm{B} \;-\; (\tilde{\beta}+\delta)^2 \int f(\mathrm{B})\, d\mathrm{B} \\ &= f(\mathrm{L})\,(\tilde{\beta}+\delta) \;-\; (\tilde{\beta}+\delta)^2\,, \end{aligned} \quad (3.15)$$

where the factor $1/2$ is introduced because each track in the like sign pairs is counted twice when the integration is performed over both i and j indices. Again, $\int f(\mathrm{B})\, d\mathrm{B} = 1$.

3.3 The estimators

The goal is to find the best (having minimum variance) unbiased point estimator of $\mathcal{B}_{L,i}$ and $\mathcal{B}_{U,i}$. The confidence region for the $\mathcal{B}_{L,i}$ and $\mathcal{B}_{U,i}$ combined is then taken to be the highest probability density regions of the corresponding likelihood function, with the point estimators as the "central value". In addition to that, when the binomial distributions are used in the likelihood function, the confidence regions for $\mathcal{B}_{L,i}$ and $\mathcal{B}_{U,i}$, in each momentum bin i separately, are *required* to lay within $[0,1]$ in order to make the estimator unbiased. This is shown in more detail in Sec. 3.4.

In this analysis we use the entire range of probabilities $[0,1]$ for $f(b \mid \mathrm{B})$. Continuous $f(b \mid \mathrm{B})$ is turned into a discrete variable: $f(b \mid \mathrm{B}) \to f_j(b)$, $j = 1,\ldots,20$, and so is momentum: p_i, $i = 1,\ldots,30$. Therefore, we end up working with $p_i \times f_j(b)$ matrix of size 30×20. In a single ij-th bin of that matrix one has

$$\begin{pmatrix} f_{ij}^{-} \\ f_{ij}^{+} \end{pmatrix} = \begin{pmatrix} f_{L,j}(b) & f_{U,j}(\bar{b}) \\ f_{L,j}(\bar{b}) & f_{U,j}(b) \end{pmatrix} \begin{pmatrix} \tilde{\mathcal{B}}_{L,i} \\ \tilde{\mathcal{B}}_{U,i} \end{pmatrix} + \begin{pmatrix} \tilde{\beta}_{ij}+\delta_{ij} \\ \tilde{\beta}_{ij}+\delta_{ij} \end{pmatrix}, \quad (3.16)$$

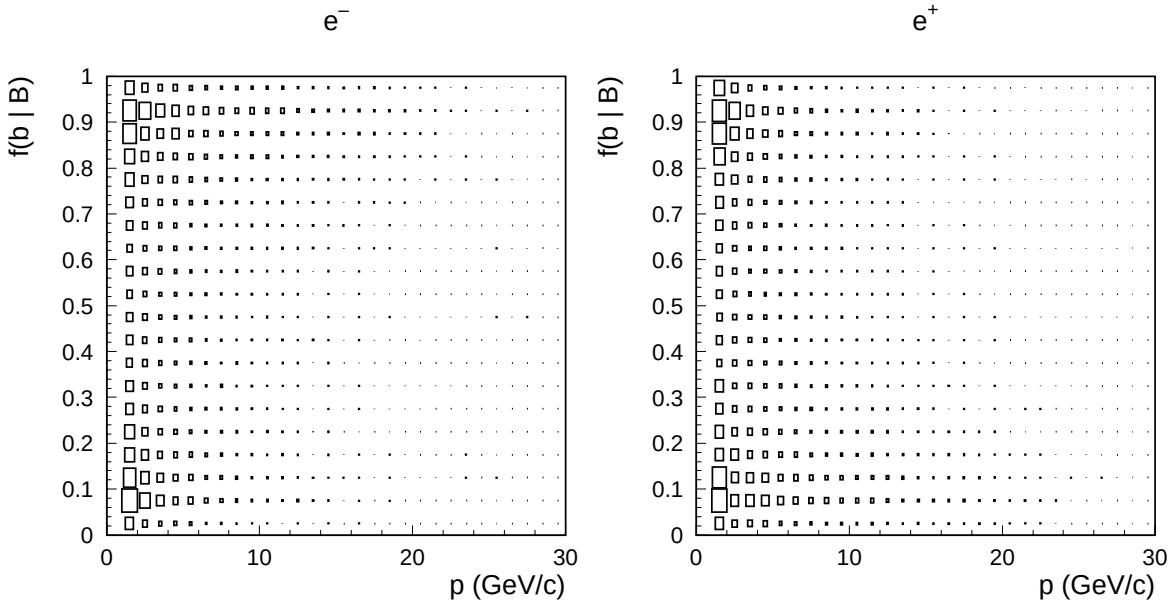


Figure 3.7: Distributions f_{ij}^- and f_{ij}^+ when electron polarization is $P_e = 1$.

where $f_i(L^- \mid \mathrm{B}) \to f_{ij}^-$ and $f_i(L^+ \mid \mathrm{B}) \to f_{ij}^+$. The $\tilde{\beta}_{ij}$ and δ_{ij} are approximately flat in j since the sign of the background and the misidentified leptons is (almost) completely random with relative to the b flavor. Distributions f_{ij}^- and f_{ij}^+ are shown in Fig. 3.7.

The likelihood function in a single ij-th bin is

$$\mathcal{L}_{ij} = \left(\begin{array}{c} n_j \\ k_{ij}^- \end{array} \right) (f_{ij}^-)^{k_{ij}^-} (1 - f_{ij}^-)^{n_j - k_{ij}^-} \times \left(\begin{array}{c} n_j \\ k_{ij}^+ \end{array} \right) (f_{ij}^+)^{k_{ij}^+} (1 - f_{ij}^+)^{n_j - k_{ij}^+} \qquad (3.17)$$

where $n(\mathrm{B}) \to n_j$, $n_i(L^-, \mathrm{B}) \to k_{ij}^-$ and $n_i(L^+, \mathrm{B}) \to k_{ij}^+$. The two factorial terms are irrelevant and are only present as a reminder that the distributions are binomial.

The central values for $\mathcal{B}_{L,i}$ and $\mathcal{B}_{U,i}$ can be obtained algebraically from

$$\begin{aligned} \frac{\partial \ln \mathcal{L}_{ij}}{\partial \tilde{\mathcal{B}}_{L,i}} &= \left(\frac{k_{ij}^-}{f_{ij}^-} - \frac{n_j - k_{ij}^-}{1 - f_{ij}^-} \right) f_{L,j}(b) + \left(\frac{k_{ij}^+}{f_{ij}^+} - \frac{n_j - k_{ij}^+}{1 - f_{ij}^+} \right) f_{L,j}(\bar{b}) = 0 \,, \\ \frac{\partial \ln \mathcal{L}_{ij}}{\partial \tilde{\mathcal{B}}_{U,i}} &= \left(\frac{k_{ij}^-}{f_{ij}^-} - \frac{n_j - k_{ij}^-}{1 - f_{ij}^-} \right) f_{U,j}(\bar{b}) + \left(\frac{k_{ij}^+}{f_{ij}^+} - \frac{n_j - k_{ij}^+}{1 - f_{ij}^+} \right) f_{U,j}(b) = 0 \,, \end{aligned}$$

or equivalently from

$$\begin{pmatrix} f_{L,j}(b) & f_{U,j}(\bar{b}) \\ f_{L,j}(\bar{b}) & f_{U,j}(b) \end{pmatrix} \begin{pmatrix} x_{ij}^- \\ x_{ij}^+ \end{pmatrix} \equiv C_j\, x_{ij} \;=\; 0\,,$$

where

$$x_{ij}^- = \left(\frac{k_{ij}^-}{f_{ij}^-} - \frac{n_j - k_{ij}^-}{1 - f_{ij}^-} \right) \qquad \text{and} \qquad x_{ij}^+ = \left(\frac{k_{ij}^+}{f_{ij}^+} - \frac{n_j - k_{ij}^+}{1 - f_{ij}^+} \right)$$

have been introduced for convenience.

The system is either singular or has solution $x_{ij}^- = 0$, $x_{ij}^+ = 0$. When it is singular,

$$\det \begin{pmatrix} f_{L,j}(b) & f_{U,j}(\bar{b}) \\ f_{L,j}(\bar{b}) & f_{U,j}(b) \end{pmatrix} \;=\; 0\,,$$

and it cannot be solved for $\mathcal{B}_L$ and $\mathcal{B}_U$. This happens when $f(b \mid \mathrm{B}) = 0.5$, the condition sometimes referred to as the "zero analyzing power" (the analyzing power is defined as $2f(b \mid \mathrm{B}) - 1$). Note that the system would also be singular if χ were 0.5 instead of around 0.12, which would be approximately the case if instead of the $B^\pm/B^0/B_s^0/b$-baryon admixture we had a pure B_s^0 sample. Here the singularity argument is in agreement with our intuition which tells us that the information about the lepton origin is completely scrambled if one cannot tell what the flavor of the parent b-quark is, either by completely not knowing what the flavor of the initial state b-quark is, or by having a completely random b mixing phase.

When the system is regular $\det C_j \neq 0$ and the solutions are

$$\tilde{\mathcal{B}}_{L,i} = \frac{g_{ij}^-\, f_{L,j}(b) - g_{ij}^+\, f_{L,j}(\bar{b})}{\det C_j}\,, \qquad \tilde{\mathcal{B}}_{U,i} = \frac{g_{ij}^-\, f_{U,j}(\bar{b}) - g_{ij}^+\, f_{U,j}(b)}{\det C_j}\,, \tag{3.18}$$

where $g_{ij}^- = k_{ij}^-/n_j - (\tilde{\beta} + \delta)_{ij}$ and $g_{ij}^+ = k_{ij}^+/n_j - (\tilde{\beta} + \delta)_{ij}$. The "true" branching fractions can be then calculated as $\mathcal{B}_{L,i} = \tilde{\mathcal{B}}_{L,i}/\eta_i$ and $\mathcal{B}_U = \tilde{\mathcal{B}}_{U,i}/\eta_i$.

Theoretically, neither the $\mathcal{B}_L$ nor the $\mathcal{B}_U$ should be negative. In practice, however, the (algebraic) solutions (3.18) can have negative values, which is either due to the statistical fluctuations of the right hand side of the expressions, or is due to the bias brought to the equations by the $\tilde{\beta} + \delta$ term, which is estimated separately.

Under the assumption that $\tilde{\beta}+\delta$ is unbiased, the confidence regions for the $\tilde{\mathcal{B}_L}$ and $\tilde{\mathcal{B}_U}$ are obtained as the highest probability density region of the likelihood function (3.17) with the constraint that both $\tilde{\mathcal{B}_L}$ and $\tilde{\mathcal{B}_U}$ must be nonnegative. It will be shown in the supplement that no bias is introduces by such constraints (again, this is only the case when $\tilde{\beta}+\delta$ is unbiased).

The likelihood function in (3.17) takes into account only one single value of $f_j(b)$. For all bins in j "continuously" the likelihood function in a *single momentum bin* i reads

$$\begin{aligned}\mathcal{L}_i &= \prod_j \mathcal{L}_{ij} \\ &= \prod_j \left(\begin{array}{c} n_j \\ k_{ij}^- \end{array} \right) (f_{ij}^-)^{k_{ij}^-} (1-f_{ij}^-)^{n_j-k_{ij}^-} \times \left(\begin{array}{c} n_j \\ k_{ij}^+ \end{array} \right) (f_{ij}^+)^{k_{ij}^+} (1-f_{ij}^+)^{n_j-k_{ij}^+} .\end{aligned} \tag{3.19}$$

The net effect of using the likelihood function (3.19) is that implicitly events with $f(b \mid \mathrm{B})$ near 0 or 1 have the highest weight, while events with $f(b \mid \mathrm{B})$ near 0.5 are weighted 0. The unbiased analyzing power factors can be obtained directly from (3.19). These derivations are lengthly and are skipped here. In practice, we by-pass this messy weighting scheme by simply proving that the estimators based on (3.19) have minimum variance and are unbiased, Sec. 3.4.

The above general derivations somewhat obscure the otherwise very simple nature of our method. We show what is it all about by going into two limits.

Near diagonal system In this limit $f(b)$ is either 1 or 0. Assuming, for a moment, that $\chi = 0$, and to simplify the discussion that $f(b\bar{b}) = 1$ and $\tilde{\beta}+\delta = 0$, one gets the master equation reduced to

$$\left(\begin{array}{c} f^- \\ f^+ \end{array} \right) = \left(\begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array} \right) \left(\begin{array}{c} \tilde{\mathcal{B}_L} \\ \tilde{\mathcal{B}_U} \end{array} \right) \qquad \text{when } f(b)=1 \text{ and } f(\bar{b})=0\,,$$

and

$$\left(\begin{array}{c} f^- \\ f^+ \end{array} \right) = \left(\begin{array}{cc} 0 & 1 \\ 1 & 0 \end{array} \right) \left(\begin{array}{c} \tilde{\mathcal{B}_L} \\ \tilde{\mathcal{B}_U} \end{array} \right) \qquad \text{when } f(b)=0 \text{ and } f(\bar{b})=1\,,$$

or jointly, when the the final state leptons are grouped into like and unlike sign with respect to the sign of the initial state b quark:

$$\begin{pmatrix} f_L \\ f_U \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \tilde{\mathcal{B}}_L \\ \tilde{\mathcal{B}}_U \end{pmatrix} \qquad \text{or} \qquad \begin{pmatrix} f_L \\ f_U \end{pmatrix} = \begin{pmatrix} 1-\chi & \chi \\ \chi & 1-\chi \end{pmatrix} \begin{pmatrix} \tilde{\mathcal{B}}_L \\ \tilde{\mathcal{B}}_U \end{pmatrix}$$

when the mixing is restored. The last equation very much resembles one that CLEO uses in its measurement of $\mathcal{B}(b\to l)$ [2], although they are fundamentally different as explained in Sec. 1.4.

Near singular system This is the opposite limit which we want to avoid as much as we can. In this limit $f(b) = f(\bar{b}) = 0.5$, and the correlation matrix is singular:

$$\begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}.$$

The same limit would have been achieved if χ were 0.5. In either case, the information about the flavor of the parent b quark would be completely scrambled, and we would have to resort to a different method to separate $\mathcal{B}_L$ from $\mathcal{B}_U$.

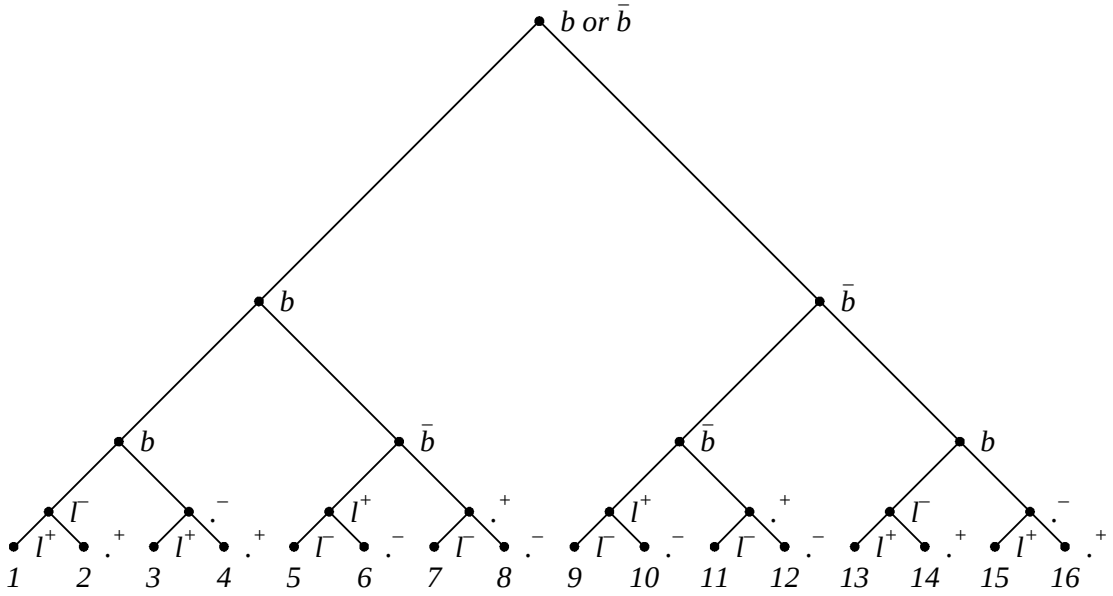
3.4 Supplement

Two results will be derived here in more detail: 1) the master equation when only events with a specified number of leptons per hemisphere are counted, and 2) proof that the estimator based on (3.19) has minimum variance and is unbiased.

3.4.1 Alternative derivation of the probability relations

The master equation (3.7) has more complicated form when only events with a specified number of leptons per hemisphere are counted: one and only one, two and only two, etc. We will illustrate this for a simplified case when $f(b\bar{b}) = 1$ and when no backgrounds and misidentified leptons are present. The consequence of the simplification is that leptons can only originate from b hadron decays, therefore only events with 0, 1, or 2 leptons per hemisphere can be observed. The logical structure representing all possible processes with 0, 1, or 2 leptons in the final state is shown in Fig. 3.8, and is formally known as the *distributive lattice* (only the upper portion of the lattice is shown).

The lattice is simple to read. Probability for the path that ends with node 4, for example, can be calculated as the product of the following four probabilities: 1) probability that a b is created, $f(b \mid \mathrm{B})$, 2) probability that the initial state b



branch	bin i	bin j	lepton yield probability
1	l^-	l^+	$\mathcal{B}_{L,i}\mathcal{B}_{U,j}(1-\chi)f(b \mid \mathrm{B})$
2	l^-	$\cdot^+$	$\mathcal{B}_{L,i}(1-\mathcal{B}_{U,j})(1-\chi)f(b \mid \mathrm{B})$
3	$\cdot^-$	l^+	$(1-\mathcal{B}_{L,i})\mathcal{B}_{U,j}(1-\chi)f(b \mid \mathrm{B})$
4	$\cdot^-$	$\cdot^+$	$(1-\mathcal{B}_{L,i})(1-\mathcal{B}_{U,j})(1-\chi)f(b \mid \mathrm{B})$
5	l^+	l^-	$\mathcal{B}_{L,i}\mathcal{B}_{U,j}\chi f(b \mid \mathrm{B})$
6	l^+	$\cdot^-$	$\mathcal{B}_{L,i}(1-\mathcal{B}_{U,j})\chi f(b \mid \mathrm{B})$
7	$\cdot^+$	l^-	$(1-\mathcal{B}_{L,i})\mathcal{B}_{U,j}\chi f(b \mid \mathrm{B})$
8	$\cdot^+$	$\cdot^-$	$(1-\mathcal{B}_{L,i})(1-\mathcal{B}_{U,j})\chi f(b \mid \mathrm{B})$
9	l^+	l^-	$\mathcal{B}_{L,i}\mathcal{B}_{U,j}(1-\chi)f(b \mid \mathrm{B})$
10	l^+	$\cdot^-$	$\mathcal{B}_{L,i}(1-\mathcal{B}_{U,j})(1-\chi)f(\bar{b} \mid \mathrm{B})$
11	$\cdot^+$	l^-	$(1-\mathcal{B}_{L,i})\mathcal{B}_{U,j}(1-\chi)f(\bar{b} \mid \mathrm{B})$
12	$\cdot^+$	$\cdot^-$	$(1-\mathcal{B}_{L,i})(1-\mathcal{B}_{U,j})(1-\chi)f(\bar{b} \mid \mathrm{B})$
13	l^-	l^+	$\mathcal{B}_{L,i}\mathcal{B}_{U,j}\chi f(b \mid \mathrm{B})$
14	l^-	$\cdot^+$	$\mathcal{B}_{L,i}(1-\mathcal{B}_{U,j})\chi f(\bar{b} \mid \mathrm{B})$
15	$\cdot^-$	l^+	$(1-\mathcal{B}_{L,i})\mathcal{B}_{U,j}\chi f(\bar{b} \mid \mathrm{B})$
16	$\cdot^-$	$\cdot^+$	$(1-\mathcal{B}_{L,i})(1-\mathcal{B}_{U,j})\chi f(\bar{b} \mid \mathrm{B})$

Figure 3.8: An oversimplified probability hierarchy for a decay of a $b\bar{b}$ system in a hemisphere: the initial state can be either b or $\bar{b}$, b can oscillate into either b or $\bar{b}$, etc., until the chain ends with a certain combination of l^+ and l^-, or none of them, $\cdot^-$ and $\cdot^+$.

propagates into a parent b without mixing, $(1-\chi)$, 3) probability that no primary lepton is created in bin i, $(1-\mathcal{B}_{L,i})$, and 4) probability that no cascade lepton is created in momentum bin j, $(1-\mathcal{B}_{L,j})$. Along the same line of reasoning one can obtain probabilities for the remaining 15 processes.

The probability that one and only one track will be tagged as a negatively charged lepton in bin i, for example, involves probability that a second lepton will be created but will not be identified (because the tracking efficiency is < 1), plus the probability that the second lepton will not be created at all. When the probabilities are combined, one gets, based on entries from the table in Fig. 3.8,

$$
\begin{aligned}
f^{-}_{1,L,i} &= \sum_j \{ \underbrace{\mathcal{B}_{L,i}\mathcal{B}_{U,j}(1-\chi)f(b \mid \mathrm{B})}_{\text{process 1}}\, \eta_i(1-\eta_j) + \underbrace{\mathcal{B}_{L,i}(1-\mathcal{B}_{U,j})(1-\chi)f(b \mid \mathrm{B})}_{\text{process 2}}\, \eta_i \\
&\quad + \underbrace{\mathcal{B}_{L,i}\mathcal{B}_{U,j}\chi f(\bar{b} \mid \mathrm{B})}_{\text{process 13}}\, \eta_i(1-\eta_j) + \underbrace{\mathcal{B}_{L,i}(1-\mathcal{B}_{U,j})\chi f(\bar{b} \mid \mathrm{B})}_{\text{process 14}}\, \eta_i \,\} \\
&= (1-\chi)f(b \mid \mathrm{B})\tilde{\mathcal{B}}_{L,i}(\mathcal{B}_U - \tilde{\mathcal{B}}_U) + (1-\chi)f(b \mid \mathrm{B})\tilde{\mathcal{B}}_{L,i}(1-\mathcal{B}_U) \\
&\quad + \chi f(\bar{b} \mid \mathrm{B})\tilde{\mathcal{B}}_{L,i}(\mathcal{B}_U - \tilde{\mathcal{B}}_U) + \chi f(\bar{b} \mid \mathrm{B})\tilde{\mathcal{B}}_{L,i}(1-\mathcal{B}_U) \\
&= (1-\chi)f(b \mid \mathrm{B})\tilde{\mathcal{B}}_{L,i}(1-\tilde{\mathcal{B}}_U) + \chi f(\bar{b})\tilde{\mathcal{B}}_{L,i}(1-\tilde{\mathcal{B}}_U) \\
&\equiv f_p(b \mid \mathrm{B})(1-\tilde{\mathcal{B}}_U)\tilde{\mathcal{B}}_{L,i}
\end{aligned}
$$

where η_j is the efficiency in bin j, and $(1-\eta_j)$ is the probability that a true lepton in bin j will not be identified. The summation over j is because the probabilities for the second lepton are for the entire momentum range. This can be repeated for processes grouped into (5, 7, 9, 11), (5, 6, 9, 10), and (1, 3, 13, 15) to yield

$$
\begin{aligned}
f^{-}_{1,U,i} &= f_p(\bar{b} \mid \mathrm{B})(1-\tilde{\mathcal{B}}_L)\tilde{\mathcal{B}}_{U,i}\,, \\
f^{+}_{1,L,i} &= f_p(\bar{b} \mid \mathrm{B})(1-\tilde{\mathcal{B}}_U)\tilde{\mathcal{B}}_{L,i}\,, \\
f^{+}_{1,U,i} &= f_p(b \mid \mathrm{B})(1-\tilde{\mathcal{B}}_L)\tilde{\mathcal{B}}_{U,i}\,,
\end{aligned}
$$

respectively. In the matrix form the above equations can be jointly written as

$$\begin{pmatrix} f^-_{1,i} \\ f^+_{1,i} \end{pmatrix} = \begin{pmatrix} f_p(b \mid \mathrm{B})(1-\tilde{\mathcal{B}}_U) & f_p(\bar{b} \mid \mathrm{B})(1-\tilde{\mathcal{B}}_L) \\ f_p(\bar{b} \mid \mathrm{B})(1-\tilde{\mathcal{B}}_U) & f_p(b \mid \mathrm{B})(1-\tilde{\mathcal{B}}_L) \end{pmatrix} \begin{pmatrix} \tilde{\mathcal{B}}_{L,i} \\ \tilde{\mathcal{B}}_{U,i} \end{pmatrix}, \tag{3.20}$$

where $f^-_{1,i} = f^-_{1,L,i} + f^-_{1,U,i}$ and $f^+_{1,i} = f^+_{1,L,i} + f^+_{1,U,i}$. Subscript 1 indicates that only events with a single tracks per hemisphere are counted.

It is easy to show that for events with two (and only two) leptons per hemisphere, which in our simplified approach can only be of the unlike signs,

$$\begin{pmatrix} f^-_{2,i} \\ f^+_{2,i} \end{pmatrix} = \begin{pmatrix} f_p(b \mid \mathrm{B})\tilde{\mathcal{B}}_U & f_p(\bar{b} \mid \mathrm{B})\tilde{\mathcal{B}}_L \\ f_p(\bar{b} \mid \mathrm{B})\tilde{\mathcal{B}}_U & f_p(b \mid \mathrm{B})\tilde{\mathcal{B}}_L \end{pmatrix} \begin{pmatrix} \tilde{\mathcal{B}}_{L,i} \\ \tilde{\mathcal{B}}_{U,i} \end{pmatrix}, \tag{3.21}$$

equivalent to (3.8) and (3.9) when $f(b\bar{b})$ is set to 1 and $\eta_i\beta_i + \delta_i$ to 0.

The marginal probability for a lepton in bin i is then obtained by summing (3.20) and (3.21):

$$\begin{pmatrix} f^-_i \\ f^+_i \end{pmatrix} = \begin{pmatrix} f_p(b \mid \mathrm{B}) & f_p(\bar{b} \mid \mathrm{B}) \\ f_p(\bar{b} \mid \mathrm{B}) & f_p(b \mid \mathrm{B}) \end{pmatrix} \begin{pmatrix} \tilde{\mathcal{B}}_{L,i} \\ \tilde{\mathcal{B}}_{U,i} \end{pmatrix}, \tag{3.22}$$

which restores the master equation (3.7).

The above derivations are needed to explain a potential source of error in measurements of $\mathcal{B}(b{\to}l)$ that have been performed at $\Upsilon(4S)$, that also use the charge correlation method to separate $b{\to}l$ from $b{\to}c{\to}l$, but involve isotropic topologies in which the directions of the initial state B mesons and the final state leptons are little correlated.

At the Z, directions of the final state leptons and the initial state b hadrons are highly correlated and no special cuts are needed to establish their parent-child connection.

At $\Upsilon(4S)$ on the other hand, various cuts are needed [2] to isolate leptons which tag one B meson, from the final state leptons produced in the decay of the other B meson in the pair. As a consequence, only events with a single lepton per "hemisphere" are counted, and the probability relations should read more like one in (3.20), with factors $(1-\tilde{\mathcal{B}}_L)$ and $(1-\tilde{\mathcal{B}}_U)$ included, while in practice they look more like (3.7), in which $f_i(L^- \mid \mathrm{B})$ and $f_i(L^+ \mid \mathrm{B})$ are marginal probabilities, obtained from a somewhat larger count of *all* leptons. If such a mismatch occurs, the values of the branching fractions are lowered by factors roughly equal to the values of $(1-\tilde{\mathcal{B}}_L)$ or $(1-\tilde{\mathcal{B}}_U)$, around $6-10\%$ depending on the lepton identification efficiencies.

3.4.2 Analytical properties of the estimator

The goal is to show that the estimator based on (3.19) is the best unbiased estimator of $\mathcal{B}_L$ and $\mathcal{B}_U$ given the constraints that the two must be non-negative. We again simplify the discussion by taking $\tilde{\beta}+\delta = 0$ and $\eta = 1$. The results are straightforward to generalize.

The problem is modeled in terms of the following variables

- $q_1 \equiv f_p(b \mid \mathrm{B})$ and $q_2 \equiv f_p(\bar{b} \mid \mathrm{B})$. $q_1 + q_2 = f(b\bar{b})$.
- n_1 and n_2: number of Bernoulli trials corresponding to q_1 and q_2. In terms of single trials $q_1 = f(n_1 = 1, n_2 = 0)$ and $q_2 = f(n_1 = 0, n_2 = 1)$.
- $p_1 \equiv \tilde{\mathcal{B}}_L$ and $p_2 \equiv \tilde{\mathcal{B}}_U$.
- k_1 and k_2: number of successes corresponding to p_1 and p_2.
- $p \equiv f_i(L^- \mid \mathrm{B})$.
- $n \equiv n(\mathrm{B})$.
- k: number of successes corresponding to p.

Clearly, $n_1 + n_2 \leq n$, $k = k_1 + k_2$, $k_1 \leq n_1$ and $k_2 \leq n_2$. Our hierarchical model consists of the probability distributions for the number of outcomes k_1 and k_2 given the number of trials n_1 and n_2, which are themselves outcomes of the number of trials n:

$$
\begin{aligned}
f(k_1|n_1) &= \left(\begin{array}{c} n_1 \\ k_1 \end{array} \right) p_1^{k_1} (1-p_1)^{n_1-k_1} , \\
f(k_2|n_2) &= \left(\begin{array}{c} n_2 \\ k_2 \end{array} \right) p_2^{k_2} (1-p_2)^{n_2-k_2} , \\
f(n_1, n_2) &= \frac{n!}{n_1! n_2! (n-n_1-n_2)!} q_1^{n_1} q_2^{n_2} (1-q_1-q_2)^{n-n_1-n_2} ,
\end{aligned}
$$

which serve as the likelihood functions when the number of counts is fixed and the probabilities have to be determined. The values of k_1, k_2, n_1, and n_2 are constrained as follows 1) $0 \leq k_1 \leq k$, 2) $n_2 \geq k_2$ and $n_1 + n_2 \leq n \Rightarrow n_1 \leq n - n_2 \Rightarrow n_1 \leq n - k_2 = n - k + k_1$ since $k_1 + k_2 = k$, and 3) $n_2 \geq k_2 = k - k_1$ and $n_2 \leq n - n_1$.

For n Bernoulli trials with k successes the best unbiased estimator of the probability $p = p_1 q_1 + p_2 q_2$ is k/n. On the other hand, the best unbiased estimators of p_1

and p_2 are just k_1/n_1 and k_2/n_2. Since the product of the two likelihood functions for two mutually independent random variables and a linear combination of the products of the likelihood functions is a likelihood function that will again produce the best unbiased estimator for the two variables, all one has to do is to prove that

$$f(k) = \sum_{k_1=0}^{k} \sum_{n_1=k_1}^{n-k+k_1} \sum_{n_2=k-k_1}^{n-n_1} f(k_1|n_1)\, f(k_2|n_2)\, f(n_1,n_1) \;=\; \left(\begin{array}{c} n \\ k \end{array} \right) p^k\, (1-p)^{n-k}\,.$$

Constraints $p_1 \geq 0$ and $p_2 \geq 0$ are implicit. The derivation follows:

$$\begin{aligned}
f(k) \;=\; & \sum_{k_1=0}^{k} \sum_{n_1=k_1}^{n-k+k_1} \sum_{n_2=k-k_1}^{n-n_1} \frac{n_1!}{k_1!(n_1-k_1)!}\, p_1^{k_1}\, (1-p_1)^{n_1-k_1} \;\times \\
& \frac{n_2!}{(k-k_1)!(n_2-k+k_1)!}\, p_2^{k-k_1}\, (1-p_2)^{n_2-k+k_1} \;\times \\
& \frac{n!}{n_1!n_2!(n-n_1-n_2)!} q_1^{n_1} q_2^{n_2} (1-q_1-q_2)^{n-n_1-n_2} \\
= \; & \sum_{k_1=0}^{k} \sum_{n_1=k_1}^{n-k+k_1} \frac{1}{k_1!(n_1-k_1)!}\, p_1^{k_1}\, (1-p_1)^{n_1-k_1} \frac{1}{(k-k_1)!}\, p_2^{k-k_1}\, (1-p_2)^{-k+k_1} \;\times \\
& n!\, q_1^{n_1}\, (1-q_1-q_2)^{n-n_1} \;\times \\
& \underbrace{\sum_{n_2=k-k_1}^{n-n_1} \frac{1}{\underbrace{(n_2-k+k_1)!}_{t!}\underbrace{(n-n_1-n_2)!}_{(n-n_1-k+k_1-t)!}} \left[\frac{(1-p_2)q_2}{1-q_1-q_2}\right]^{n_2} \frac{(n-n_1-k+k_1)!}{(n-n_1-k+k_1)!}}_{1} \\
& \stackrel{1}{=} \frac{1}{(n-n_1-k+k_1)!} \left[\frac{(1-p_2)q_2}{1-q_1-q_2}\right]^{k-k_1} \left[1+\frac{(1-p_2)q_2}{1-q_1-q_2}\right]^{n-n_1-k+k_1} \\
& \stackrel{1}{=} \frac{1}{(n-n_1-k+k_1)!} \left[\frac{(1-p_2)q_2}{1-q_1-q_2}\right]^{k-k_1} \left[\frac{1-q_1-q_2}{1-q_1(1-p_2)}\right]^{k-k_1-n+n_1} \\
& \stackrel{1}{=} \frac{1}{(n-n_1-k+k_1)!} \left[\frac{(1-p_2)q_2}{1-q_1-q_2p_2}\right]^{k-k_1} (1-q_1-p_2q_2)^{n-n_1} \\
= \; & \sum_{k_1=0}^{k} \frac{1}{k!}\, p_1^{k_1}\, (1-p_1)^{-k_1} \frac{1}{(k-k_1)!}\, p_2^{k-k_1}\, (1-p_2)^{-k+k_1}\, n! \left[\frac{(1-p_2)q_2}{1-q_1-q_2p_2}\right]^{k-k_1} \;\times
\end{aligned}$$

$$
\begin{aligned}
& (1-q_1-p_2q_2)^n \;\times \\
& \underbrace{\sum_{n_1=k_1}^{n-k+k_1} \frac{1}{(n_1-k_1)!(n-n_1-k+k_1)!}\,(1-p_1)^{n_1}\,q_1^{n_1}\,(1-q_1-q_2p_2)^{-n_1}}_{2} \\
& \quad \stackrel{2}{=} \frac{1}{(n-k)!}\left(\frac{q_1(1-p_1)}{1-q_1-q_2p_2}\right)^{k_1}\left(1+\frac{q_1(1-p_1)}{1-q_1-q_2p_2}\right)^{n-k} \\
& \quad \stackrel{2}{=} \frac{1}{(n-k)!}\left(\frac{q_1(1-p_1)}{1-q_1-q_2p_2}\right)^{k_1}\left(\frac{1-q_1p_1-q_2p_2)}{1-q_1-q_2p_2}\right)^{n-k} \\
= \;& \sum_{k_1=0}^{k}\frac{1}{k_1!}\left(\frac{p_1}{1-p_1}\right)^{k_1}\frac{1}{(k-k_1)!}\,n!\left(\frac{p_2q_2}{1-q_1-p_2q_2}\right)^{k-k_1}(1-q_1-p_2q_2)^n \\
& \frac{1}{(n-k)!}\left(\frac{q_1(1-p_1)}{1-q_1-q_2p_2}\right)^{k_1}\left(\frac{1-q_1p_1-q_2p_2}{1-q_1-q_2p_2}\right)^{n-k} \\
= \;& \frac{n!}{(n-k)!}\,(1-q_1p_1-q_2p_2)^{n-k}\,(p_2q_2)^k\,\underbrace{\sum_{k_1=0}^{k}\frac{1}{k_1!(k-k_1)!}\,p_1^{k_1}\,\frac{1}{(p_2q_2)^{k_1}}\,q_1^{k_1}\,\frac{k!}{k!}}_{3} \\
& \quad \stackrel{3}{=} \frac{1}{k!}\sum_{k_1=0}^{k}\binom{k}{k_1}\left(\frac{p_1q_1}{p_2q_2}\right)^{k_1} \;=\; \frac{1}{k!}\left(1+\frac{p_1q_1}{p_2q_2}\right)^{k} \;=\; \frac{1}{k!}\left(\frac{p_1q_1+p_2q_2}{p_2q_2}\right)^{k} \\
= \;& \frac{n!}{k!(n-k)!}\,(1-q_1p_1-q_2p_2)^{n-k}\,(p_2q_2)^k\left(\frac{p_1q_1+p_2q_2}{p_2q_2}\right)^{k} \\
= \;& \binom{n}{k}\,(p_1q_1+p_2q_2)^k\,(1-p_1q_1-p_2q_2)^{n-k} \;\equiv\; \binom{n}{k}\,p^k\,(1-p)^{n-k}\,.
\end{aligned}
$$

(In practice derivations like this are done using symbolic manipulation languages like *Mathematica* or *Maple*, especially in more complicated cases.)

4

Identification of electrons and muons

Probability relations from the previous chapter contain terms that have been identified as lepton efficiency η, and mis-ID probability ε. To determine the two, one needs a model that implements our idea of "what electron or muon is" in terms of their characteristic signatures, in either the fundamental processes in the events, or from the variables describing their interactions with the detector materials. In this analysis we use a model that is implemented in the Monte Carlo, which we shall calibrate to the data to extract the correct values of η and ε. Given that a test for a single particle type can only have two outcomes - particle *is*, say an electron, or *is not* - the theoretical framework of particle identification is naturally that of *simple* hypothesis testing (when the population parameter only takes a discrete value).

4.1 Basics of simple hypothesis testing

The most powerful hypothesis test for a pair of alternative simple hypotheses is given by famous *Neyman-Pearson lemma* [66]. The statement of the theorem and a corollary follow after some basic statistical concepts are put into the context.

• The populations will be expressed as $f(\mathbf{x} \mid \theta)$ where θ is the population parameter with only two discrete values representing the particle type, and $\mathbf{x}$ is a vector in the sample space - the sample vector. The sample spaces used in particle identification here will be spanned by the detector or the event variables. Some of the detector variables are $EM1$, $EM2$: energies deposited in two electromagnetic LAC modules, $EM1_T = EM1_{1\times1}/EM1_{3\times3}$, $EM2_T = EM2_{2\times2}/EM2_{3\times3}$: ratios of the 1×1 to 3×3 and the 2×2 to 3×3 energy matrices of the clusters in two electromagnetic modules (serving as a measure of the lateral distributions of the electromagnetic showers), etc. The event variables are track momentum, polar angle, impact parameter, invariant mass of pairs of tracks, etc. Examples of the sample vectors are $\mathbf{x} = (EM1, EM2)$, $\mathbf{x} = (EM1, EM2, EM1_T, EM2_T)$, etc.

• The two complementary hypotheses in a hypothesis testing are called the *null hypothesis*, $H_0 : \theta = \theta_0$, and the *alternative hypothesis*, $H_1 : \theta = \theta_1$. Here, the alternative hypotheses will be called e hypothesis, μ hypothesis, etc., their null hypotheses e null hypothesis, μ null hypothesis, etc.

• The region of the sample space for which H_0 is rejected is called the *rejection region* of the null hypothesis and will be denoted A, its conjugate region A^c. For the reminder of this document the rejection region of the null hypothesis is equated to the *acceptance region* of the alternative hypothesis.

• If $\theta = \theta_0$ but the test rejects H_0 as incorrect, then the test has made *type I error*. If $\theta = \theta_1$ but the test accepts H_0 as correct, then the test has made *type II error*. The size of the type I error is given by $\alpha = \int_A f(\mathbf{x} \mid \theta_0) d\mathbf{x}$, and the size of the type II error by $\beta = \int_{A^c} f(\mathbf{x} \mid \theta_0) d\mathbf{x}$. We call a hypothesis test *the best* if for a given size of type I error the size of type II error has a minimum.

Neyman-Pearson lemma *Consider testing null hypothesis H_0: $\theta = \theta_0$ versus alternative hypothesis H_1: $\theta = \theta_1$ where the corresponding distributions are given by $f(\mathbf{x} \mid \theta_0)$ and $f(\mathbf{x} \mid \theta_1)$ respectively. If for the null hypothesis there exists rejection region A of size $\alpha = \int_A f(\mathbf{x} \mid \theta_0) d\mathbf{x}$ and a constant k such that*

$$\frac{f(\mathbf{x} \mid \theta_1)}{f(\mathbf{x} \mid \theta_0)} \geq k \text{ inside } A\,, \qquad \frac{f(\mathbf{x} \mid \theta_1)}{f(\mathbf{x} \mid \theta_0)} < k \text{ outside } A\,, \tag{4.1}$$

then A is the best size α rejection region of the null hypothesis (For a simple proof see Chapter 9 of [67]).

[The two distributions are sometimes spelled out in terms of likelihood functions $\mathcal{L}(\theta_1 \mid \mathbf{x}) \equiv f(\mathbf{x} \mid \theta_1)$ and $\mathcal{L}(\theta_0 \mid \mathbf{x}) \equiv f(\mathbf{x} \mid \theta_0)$, whose only purpose is to suggest that when the test is actually applied $\mathbf{x}$ is fixed (from the experiment) while θ becomes a random variable. The ratio of the two distributions is then called the *likelihood ratio.*]

Corollary For any two rejection regions A_1 and A_2 of the null hypothesis and their respective constants k_1 and k_2 satisfying the condition of Neyman-Pearson lemma,

$$k_1 \leq k_2 \;\Rightarrow\; A_1 \supseteq A_2 \;\Rightarrow\; \int_{A_1} f(\mathbf{x} \mid \theta)\, d\mathbf{x} \geq \int_{A_2} f(\mathbf{x} \mid \theta)\, d\mathbf{x}\,. \tag{4.2}$$

In other words k and A are one-to-one and α (β) is monotonically decreasing (increasing) function of k.

4.2 Particle identification as a simple hypothesis test

An electron hypothesis will be denoted e, its null hypothesis non-e, and similarly for muons, μ and non-μ. Positive outcomes of the tests will be denoted with capital letters like E, H, etc., loosely indicating which subsystem is used by the test.

The acceptance region for electrons, for example, as determined by the condition of the Neyman-Pearson lemma, is given by

$$A_E = \{\mathbf{x} : f(\mathbf{x} \mid e) - k_E f(\mathbf{x} \mid \text{non-}e) \geq 0\}\,, \tag{4.3}$$

the corresponding type I error by

$$\alpha = \int_{A_E} f(\mathbf{x} \mid \text{non-}e)\, d\mathbf{x} = f(E \mid \text{non-}e) \equiv \text{ mis-ID probability}\,, \tag{4.4}$$

and the corresponding type II error by

$$\beta = \int_{A_E^c} f(\mathbf{x} \mid e)\, d\mathbf{x} \;\Rightarrow\; 1-\beta = \int_{A_E} f(\mathbf{x} \mid e)\, d\mathbf{x} = f(E \mid e) \equiv \text{ efficiency}\,. \tag{4.5}$$

Therefore, for a given mis-ID probability and a constant k_E the best acceptance region is set by the statement of the theorem. Examples of acceptance and rejection regions using $\mathbf{x} = (EM1, EM2)$ for electrons and $\mathbf{x} = (HAD1, HAD2)$ for muons are shown in Fig. 4.1 and Fig. 4.2.

[Note that the efficiency and the mis-ID probability are conjugate variables in probability relations (3.7), as well as in the statement of the Neyman-Pearson lemma, (4.4) and (4.5).]

Clearly, in order to determine the acceptance regions one first has to know the $f(\mathbf{x} \mid \theta_1)$ and $f(\mathbf{x} \mid \theta_0)$. When the sample space is spanned by one or two variables the distributions can be obtained directly from the Monte Carlo. When the number of variables is larger, which *is* the case for electrons, and somewhat less for muons, the $f(\mathbf{x} \mid \theta)$ cannot be obtained directly, but has to be obtained from the lower dimensional distributions $f(\mathbf{x_i} \mid \theta)$ instead, in which $\mathbf{x_i}$ is one or two dimensional. In general, $f(\mathbf{x} \mid \theta)$ is *not* equal to $\prod_i f(\mathbf{x_i} \mid \theta)$ since $\mathbf{x_i}$ are generally not uncorrelated.

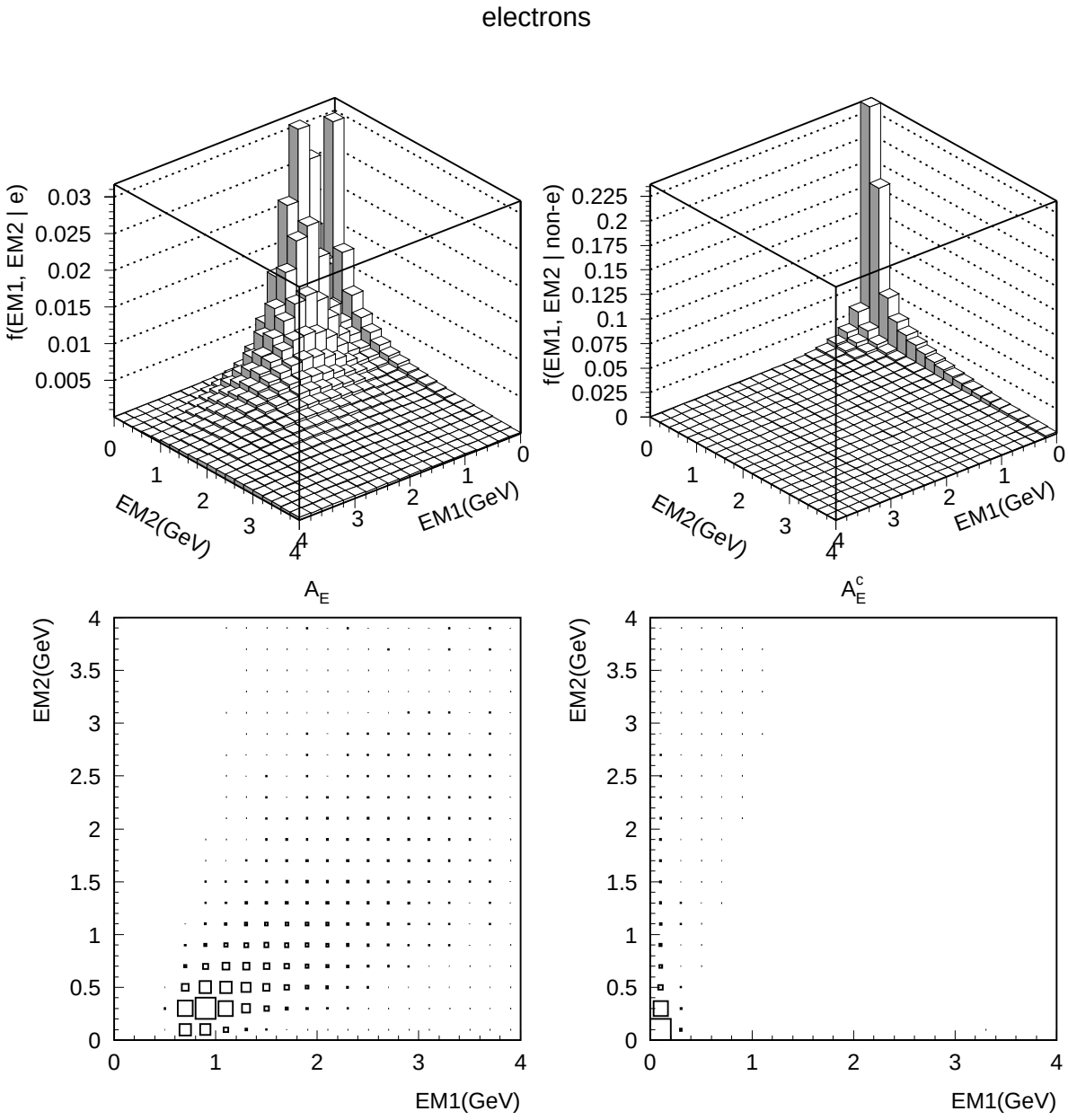


Figure 4.1: Electron "fingerprint" using $\mathbf{x} = (EM1, EM2)$. The acceptance region is defined as $A_E = \{\mathbf{x} : f(\mathbf{x} \mid e) - k_E f(\mathbf{x} \mid \text{non-}e) \geq 0\}$ with k_E set to 1. The four plots feature $f(\mathbf{x} \mid e)$ and $f(\mathbf{x} \mid \text{non-}e)$ in the top row, the distribution of $f(\mathbf{x} \mid e) - k_E f(\mathbf{x} \mid \text{non-}e)$ in the electron acceptance region (bottom left), and the distribution of $-(f(\mathbf{x} \mid e) - k_E f(\mathbf{x} \mid \text{non-}e))$ in the electron rejection region A_E^c (bottom right).

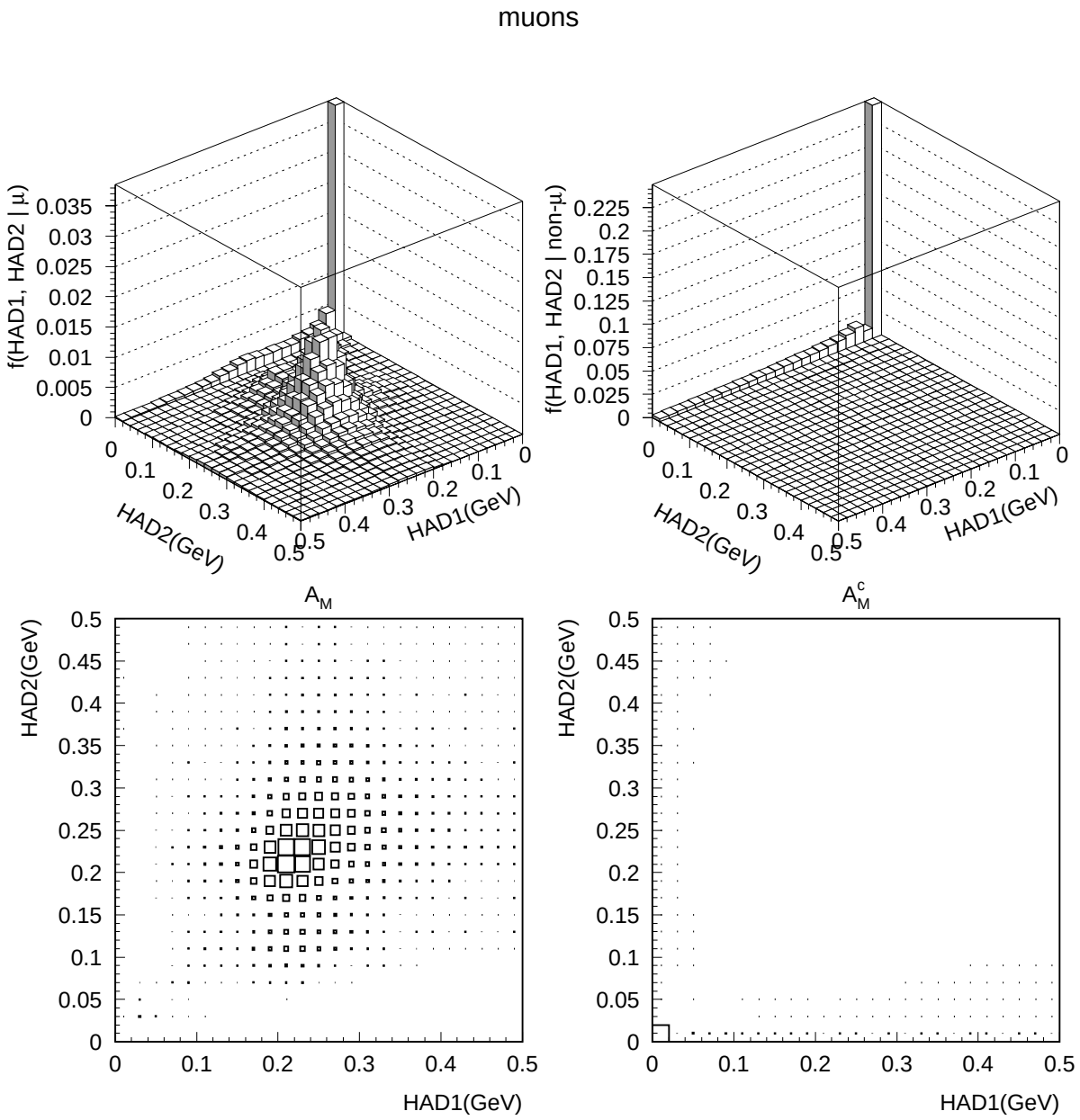


Figure 4.2: Muon "fingerprint" using $\mathbf{x} = (HAD1, HAD2)$. The acceptance region is defined as $A_M = \{\mathbf{x} : f(\mathbf{x} \mid \mu) - k_M f(\mathbf{x} \mid \text{non-}\mu) \geq 0\}$ with k_M set to 1. The four plots feature $f(\mathbf{x} \mid \mu)$ and $f(\mathbf{x} \mid \text{non-}\mu)$ in the top row, the distribution of $f(\mathbf{x} \mid \mu) - k_M f(\mathbf{x} \mid \text{non-}\mu)$ in the muon acceptance region (bottom left), and the distribution of $-(f(\mathbf{x} \mid \mu) - k_M f(\mathbf{x} \mid \text{non-}\mu))$ in the muon rejection region A_M^c (bottom right).

Actually most of the variables *are* mutually correlated as will be shown in Sec. 4.3. In this analysis we have adopted a simplified approach by using $\prod_i f(\mathbf{x_i} \mid \theta)$ as our multidimensional distribution and calling it $f(\mathbf{x} \mid \theta)$. The likelihood ratio from the distributions obtained in such a way is not "the best"; in order to get one, one has to find a transformation of $\{\mathbf{x_i}\}$ into a set of uncorrelated variables, $\{\mathbf{y_i}\}$, for which $f(\mathbf{y} \mid \theta) = \prod_i f(\mathbf{y_i} \mid \theta)$ holds. Finding these transformations in general is not a trivial task, which we have by-passed by making an educated guess of the "not so much correlated" initial set of variables $\{\mathbf{x_i}\}$.

It will be shown in Sec. 4.4 that the accuracy of the efficiency calibration can be improved when a hypothesis test has a form of a direct product of two mutually independent tests. As a result, the set of variables $\{\mathbf{x_i}\}$ has to be partitioned into two subsets, each subset corresponding to one of the two tests. If A_1 and A_2 are two best acceptance regions in the sample (sub)spaces spanned by the respective variables of the two tests, then $A_1 \cap A_2$ is not necessarily the best acceptance region, one that would have been obtained if the entire set of variables $\{\mathbf{x_i}\}$ has been used to create a single test from the scratch. Therefore some information is deliberately lost for the calibration purpose.

Also, when the electron and the muon tests are independently applied some of the tracks are labeled by both tests. In calculations in which both particle types have to be simultaneously determined, for example pairs of leptons either of which could be an electron or a muon, a particle type arbitration has to be done for tracks that are labeled by both tests. Precisely for that reason we simplified the analysis somewhat (again at expense of some information loss) by working with electrons and muons separately, in which case it does not matter whether a track is labeled by both tests or not.

4.3 LAC, WIC and CRID variables

Variables from LAC, WIC, and CRID are used as a means of identifying electrons, muons, and hadrons. Electrons are mainly distinguished from heavier particles by their characteristic longitudinal and lateral electromagnetic shower profiles which are observed by the two finely segmented LAC EM modules. Combined efficiency and mis-ID probability for electrons is improved somewhat by additionally using gas CRID and LAC HAD1 information. Muons are mainly identified by means of variables from WIC and two LAC HAD modules. Combined efficiency and mis-ID probability for muons is slightly improved by additionally using gas CRID and LAC EM information. The hadron varieties, π, K, and p, can only be mutually distinguished using the CRID.

4.3.1 Passage of heavy particles through matter

The rate of energy loss per unit path length dE/dx for heavy particles (heavier than electron) passing through matter is commonly described by the well known *Bethe-Bloch formula* [52, 7]. For most of materials and energies ranging from ≈ 10 MeV up to about 100 GeV it is accurate to within few percents. Its correctness is based on observations that heavy particles lose their energy mostly through inelastic collisions with the atomic electrons in the material, and that deflection of particles per scatterer is small. Deflections are mostly due to elastic scattering of particles with atomic nuclei which are also responsible for energy loss arising from the acceleration of the particle in the process (*bremsstrahlung*). In general, heavy particles transfer very little energy through the elastic collisions with nuclei. As a consequence the total length of the zigzag path is good approximation of the projected range to the initial direction of the particle, and the energy is deposited rather narrowly along the path. The dependence of dE/dx as a function of the penetration depth, known as *Bragg curve* [52], shows that most of the energy is deposited near the end of the trajectory which gives rise to a characteristic longitudinal energy deposition profile.

Most of the deflections of heavy particles in a medium is due to Coulomb scattering from nuclei. Strong interactions make the total cross section (and thus deflections) for hadrons somewhat larger as compared to muons, thus increasing the lateral and decreasing the longitudinal projection of the track trajectory. Muons are thus distinguished by their long penetration depth.

4.3.2 Passage of electrons through matter

The behavior of electrons is essentially different from that of heavy particles for several reasons. Due to their small mass bremsstrahlung becomes the dominant mechanism of energy loss above the *critical energy* (energy at which the radiation loss equals collision loss), which for most materials does not exceed 100 MeV, well below the SLD cutoff energy of 1 GeV. At the lowest order the bremsstrahlung matrix element is similar to that of Compton scattering except that the incident photon is that from the static field of a nucleus [14, 52]. In momentum range well above the critical energy the cross section varies quadratically with Z/m, $\sigma \propto (Z/m)^2$, while the energy loss varies linearly with energy. Another feature of the radiation loss processes is that unlike the ionization loss, which is quasi-continuous along the trajectory, almost all radiation energy can be emitted in one or two photons. Combined with the pair production by radiated photons the deposition of energy by electrons and positrons has a form of electromagnetic showers with broad lateral distributions and small penetration depth [53].

4.3.3 LAC variables

The following LAC variables are used to span the sample spaces of $f(\mathbf{x} \mid e)$ and $f(\mathbf{x} \mid \text{non-}e)$ [54]:

- $\underline{EM1,\ EM2,\ HAD1,\ HAD2}$
  Cluster energies deposited by a particle in two electromagnetic and two hadronic LAC modules. Longitudinal distribution of deposited energy is sometimes characterized by $EM_{\parallel} = EM2/EM1$.

- $\underline{EM1S = EM1\ \sin\theta,\ EM2S = EM2\ \sin\theta}$
  $EM1$ and $EM2$ scaled by the $\sin\theta$ of the polar angle. These variables take into account the change in the material thickness with the polar angle. Compared to bare $EM1$ and $EM2$ they slightly improve the electron resolution.

- $\underline{EM1_{1\times1},\ EM1_{2\times2},\ EM1_{3\times3}}$
  Energies of 1×1, 2×2, and 3×3 matrices in EM1 clusters. In our studies here $EM1_{\perp} = EM1_{1\times1}/EM1_{3\times3}$ is used as a measure of lateral distributions of em showers in EM1.

- $\underline{EM2_{2\times2},\ EM2_{3\times3}}$
  Energies of 2×2, and 3×3 matrices in EM2 clusters. Similarly as $EM1_{\perp}$, $EM2_{\perp} = EM2_{2\times2}/EM2_{3\times3}$ is used as a measure of lateral distributions of em showers in EM2.

- $\underline{HAD1/p}$
  Momentum scaled HAD1 energy. Used to additionally separate electrons from other particles.

- $\underline{\sigma_{\phi},\ \sigma_{\theta}}$
  Energy widths in ϕ and θ. These variables poorly discriminate the particle species and are not used in our studies.

- $\underline{\cos\theta,\ p}$
  These are kinematical, not the detector variables. In principle all of the above LAC variables depend on polar angle and particle momentum.

The cluster variables (all of the above except $\cos\theta$ and p) are only useful when the clusters are well separated between each other and are unambiguously associated to the tracks. In $Z\rightarrow b\overline{b}$ events with relatively high track multiplicity and narrow back-to-back jet topologies, the clusters, especially in the EM modules, do overlap. Moreover, the track-cluster association is more ambiguous in high multiplicity back-to-back events than in events with fewer tracks or isotropic events. The net effect observed is lower efficiency in hadronic events compared to τ events by about 20%.

In that regard, muons are effectively better separated from other particles than electrons. Due to their long penetration depth, they are observed by the outer calorimetric subsystems, like the LAC HAD modules and the WIC, where the tracks are better separated between each other and where the track multiplicity is lower. The total thickness of the two EM modules of about 0.8 absorption lengths is enough to absorb about 40% of the hadron energy. With the thickness of each of the two HAD modules of about 1 absorption length, the hadron energy absorbed by the EM1, EM2, and HAD1 modules combined is $70-80\%$ [51]. Muons are therefore easy to observe by means of $HAD1$ and $HAD2$ variables, Fig. 4.2, which are used as an alternative way of their identification, and are also important for the calibration purposes. To illustrate this we anticipate some of the results of Sec. 4.8.2. The lower right plot in Fig. 4.38, for example, shows that the the purity of muons selected by the LAC (in combination with an impact parameter test) is around 35% in the entire momentum range.

It will be argued in Sec. 4.4 (and confirmed with the studies of Sec. 4.8) that the accuracy of the efficiency calibration can be improved when two mutually independent tests are used for a single particle type hypothesis. We obtain first of the two electron tests by using $\mathbf{x} = \{EM1S, EM2S, EM1_\perp, EM2_\perp, p\}$ as

$$f(\mathbf{x} \mid e)_E = f(EM1S, p \mid e)\, f(EM2S, p \mid e)\, f(EM1_\perp, p \mid e)\, f(EM2_\perp, p \mid e)\,, \quad (4.6)$$

and similarly for the null hypothesis. The individual terms are obtained directly from the MC, Fig. 4.3 and 4.4.

The second electron test is obtained by combining

$$\mathcal{L}_{H,e} \equiv f(\mathbf{x} \mid e)_H = f(HAD1/p, \cos\theta \mid e) \quad (4.7)$$

Fig. 4.5, with the CRID likelihood function (see CRID tests below).

One of the two tests for muons uses

$$\mathcal{L}_{L,\mu} \equiv f(\mathbf{y} \mid \mu)_L = f(EM1, EM2 \mid \mu)\, f(HAD1, HAD2 \mid \mu) \quad (4.8)$$

in combination with the CRID likelihood function (see CRID tests below). This is actually a secondary test; the main one is obtained by using WIC variables (see WIC variables). Subscript L in (4.8) indicates that the test mostly relies on LAC variables (the contribution of CRID in this test is minimal).

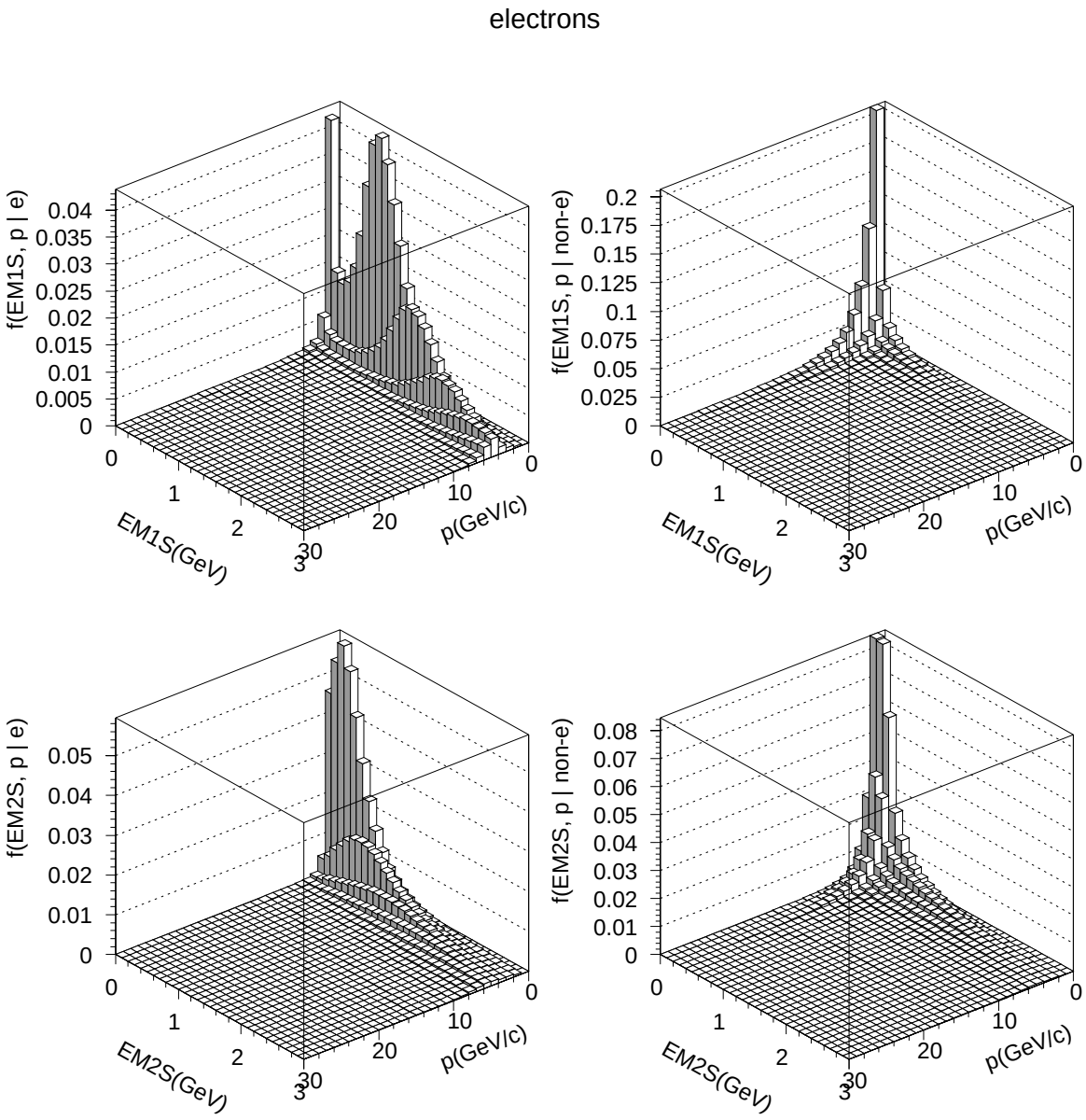


Figure 4.3: Distributions $f(EM1S, p \mid e)$ and $f(EM2S, p \mid e)$ for electron hypothesis are shown in the left column. The right column contains corresponding distributions for electron null hypothesis. Only portions of the matrices are shown for clarity. The momentum extends up to 50 GeV (divided into 50 bins), while the $EM1S$ and $EM2S$ extend up to 20 GeV (divided into 200 bins).

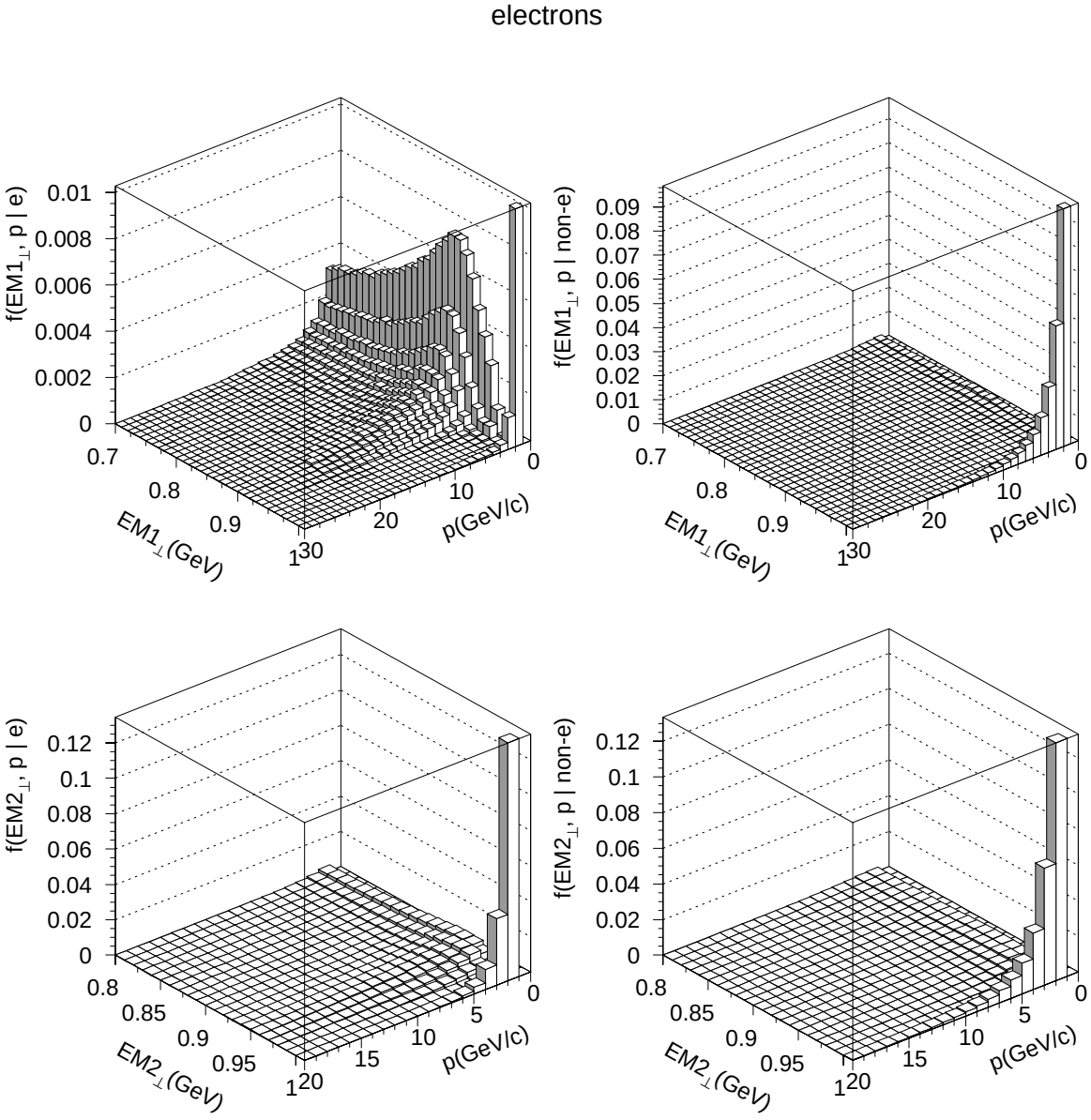


Figure 4.4: Distributions $f(EM1_\perp, p \mid e)$ and $f(EM2_\perp, p \mid e)$ for electron hypothesis are shown in the left column. The right column contains corresponding distributions for electron null hypothesis. Only portions of the matrices are shown for clarity. The momentum extends up to 50 GeV (divided into 50 bins). Lower portions of $EM1_\perp$ and $EM2_\perp$ are not shown. Both $EM1_\perp$ and $EM2_\perp$ extend from 0 to 1.01, and are divided into 101 bins. The extra (101) bin is used to handle exceptional conditions.

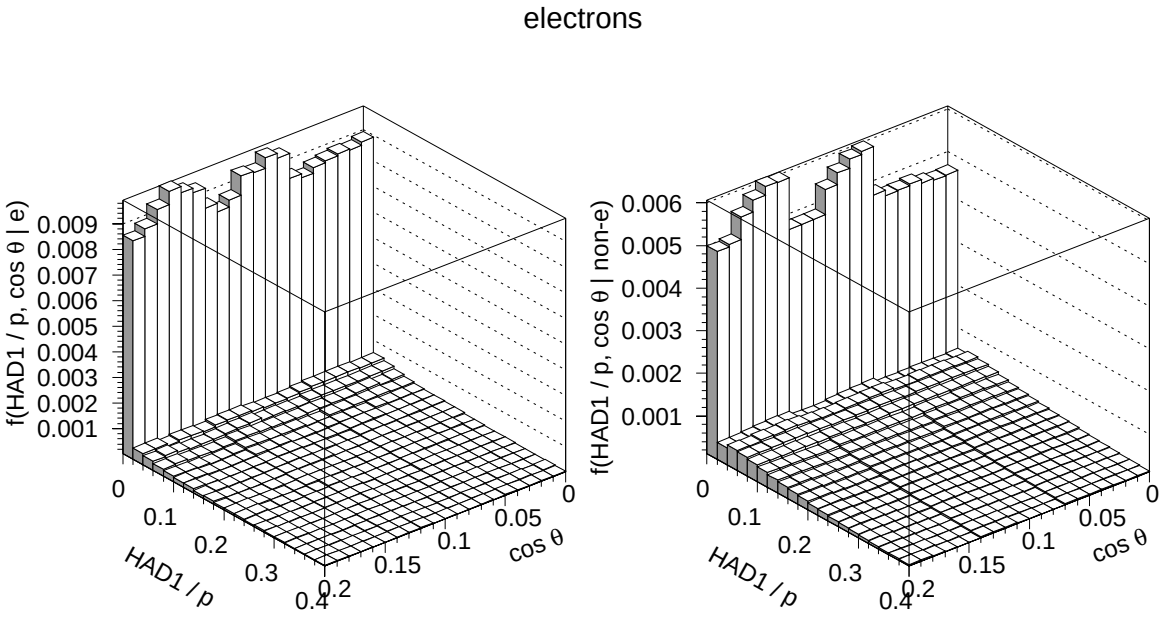


Figure 4.5: Distributions $f(HAD1/p, \cos\theta \mid e)$ and $f(HAD1/p, \cos\theta \mid \text{non-}e)$. From the plots the test does not appear to be highly discriminating. However, it brings substantial improvements in momentum region above 8 GeV. The acceptance region for electrons is a very narrow band near $HAD1/p = 0$, while the rest of the area is the acceptance region for hadrons.

4.3.4 WIC variables

The following two WIC variables were used to build the muon hypothesis test:

- wflag
  WIC flag takes values on $\{0, 1\}$, indicating whether a particle has been registered by the WIC (requires deposited energy above the electronic noise threshold).

- nlaybey
  Number of layers intersected by a track beyond preset interaction length limit.

There is a large number of other WIC variables that discriminate muons on the basis of their long penetration depth. The main effect of the WIC thickness of about 4 hadronic absorption lengths, is that the residual hadrons are more likely to be absorbed in the inner layers while the muons are more punch-through. Considerable

improvements, estimated to about 20% in efficiency for the given mis-ID probability, can be achieved using other WIC variables [69].

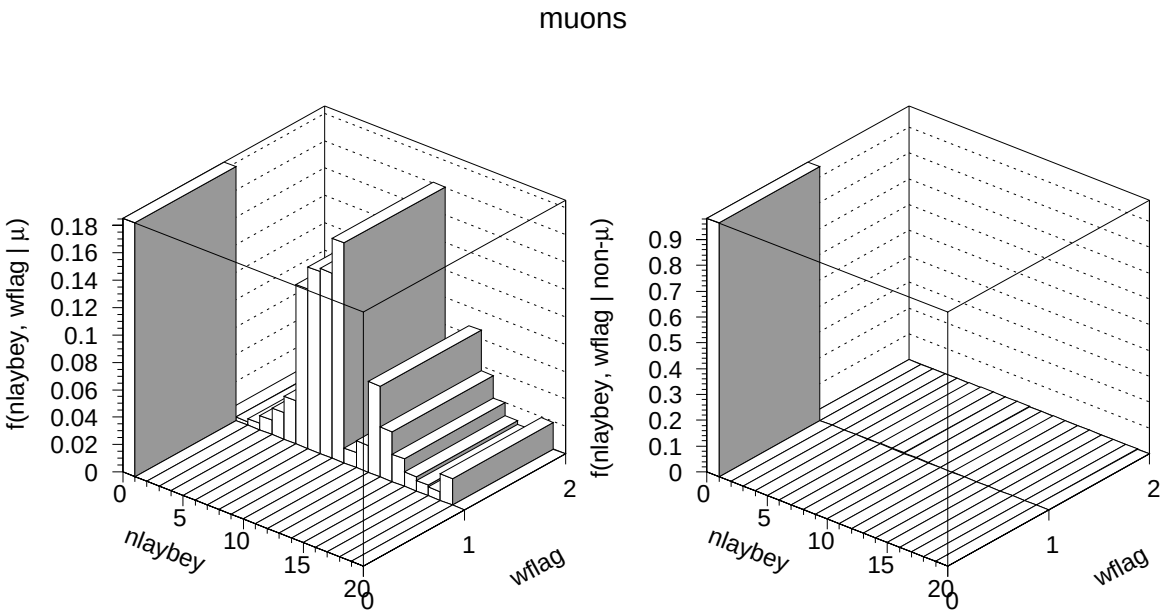


Figure 4.6: Distributions $f(\text{nlaybey}, \text{wflag} \mid \mu)$ and $f(\text{nlaybey}, \text{wflag} \mid \text{non-}\mu)$. This rather primitive test does not take into account a whole host of other WIC variables, which can be used to build much more efficient test in a manner similar to the electron test.

In terms of the wflag and nlaybey variables, the likelihood function for the muon hypothesis is given by

$$f(\mathbf{y} \mid \mu)_W \equiv f(\text{nlaybey}, \text{wflag} \mid \mu)\,, \tag{4.9}$$

and similarly for the muon null hypothesis, Fig. 4.6.

4.3.5 CRID tests

A very nice particle type test using CRID has been developed at SLD by members of the CRID group [70]. The outcomes of the test are in the form of logarithms of likelihood functions for each particle hypothesis separately: $\ln \mathcal{L}_e$, $\ln \mathcal{L}_\mu$, $\ln \mathcal{L}_\pi$, $\ln \mathcal{L}_K$, and $\ln \mathcal{L}_P$, for either the LIQUID or the GAS CRID subsystem. For electron

identification only GAS CRID information is useful, in the momentum range $p > 1$ GeV (determined by $\beta n(\omega) > 1$), up to the saturation point for pions of about 5 GeV, Fig. 2.7. Improvements in muon identification using GAS CRID are very minor.

By the book, the null hypothesis for electrons (and similarly for muons) is obtained from the above likelihood functions as

$$\mathcal{L}_{C,non-e} = \mathcal{L}_\mu f_\mu + \mathcal{L}_\pi f_\pi + \mathcal{L}_K f_K + \mathcal{L}_P f_P$$

where f_μ, f_π, f_K, f_P are fractions of particle types other than electron in the composition. In this analysis we used pion hypothesis as a null hypothesis for both electrons and muons. Plots of $\ln \mathcal{L}_e - \ln \mathcal{L}_\pi$ and $\ln \mathcal{L}_\mu - \ln \mathcal{L}_\pi$ for electron, muon, and pion hypotheses are shown in Fig. 4.7.

Also, the CRID test has not been used alone, but rather in conjunction with the LAC HAD test for electrons, Eq. (4.7), and the LAC EM and HAD test for muons, Eq. (4.8). Following the prescription of the Neyman-Pearson lemma, acceptance regions for the "best" combined tests are determined from

$$\frac{\mathcal{L}_{H,e}\,\mathcal{L}_{C,e}}{\mathcal{L}_{H,non-e}\,\mathcal{L}_{C,non-e}} \geq k_{HC} \tag{4.10}$$

for electrons, and

$$\frac{\mathcal{L}_{L,\mu}\,\mathcal{L}_{C,\mu}}{\mathcal{L}_{L,non-\mu}\,\mathcal{L}_{C,non-e}} \geq k_{LC} \tag{4.11}$$

for muons. The CRID variable (gas Cherenkov ring radius) is implicitly taken into account there, and is assumed uncorrelated to the LAC HAD variables. In a more "CRID-like fashion" this can be written as

$$(\ln \mathcal{L}_{H,e} + \ln \mathcal{L}_{C,e}) - (\ln \mathcal{L}_{H,non-e} + \ln \mathcal{L}_{C,non-e}) \geq \ln k_{HC}\ ,$$

for electrons and similarly for muons.

* * *

To summarize, we have two electron tests, E and H, Eq. (4.6) and (4.10), and two muon tests, L and W, Eq. (4.9) and (4.11). The property of mutual independence of two tests (for a single hypothesis) can be used to calibrate the efficiency in cases when no pure reference samples are available.

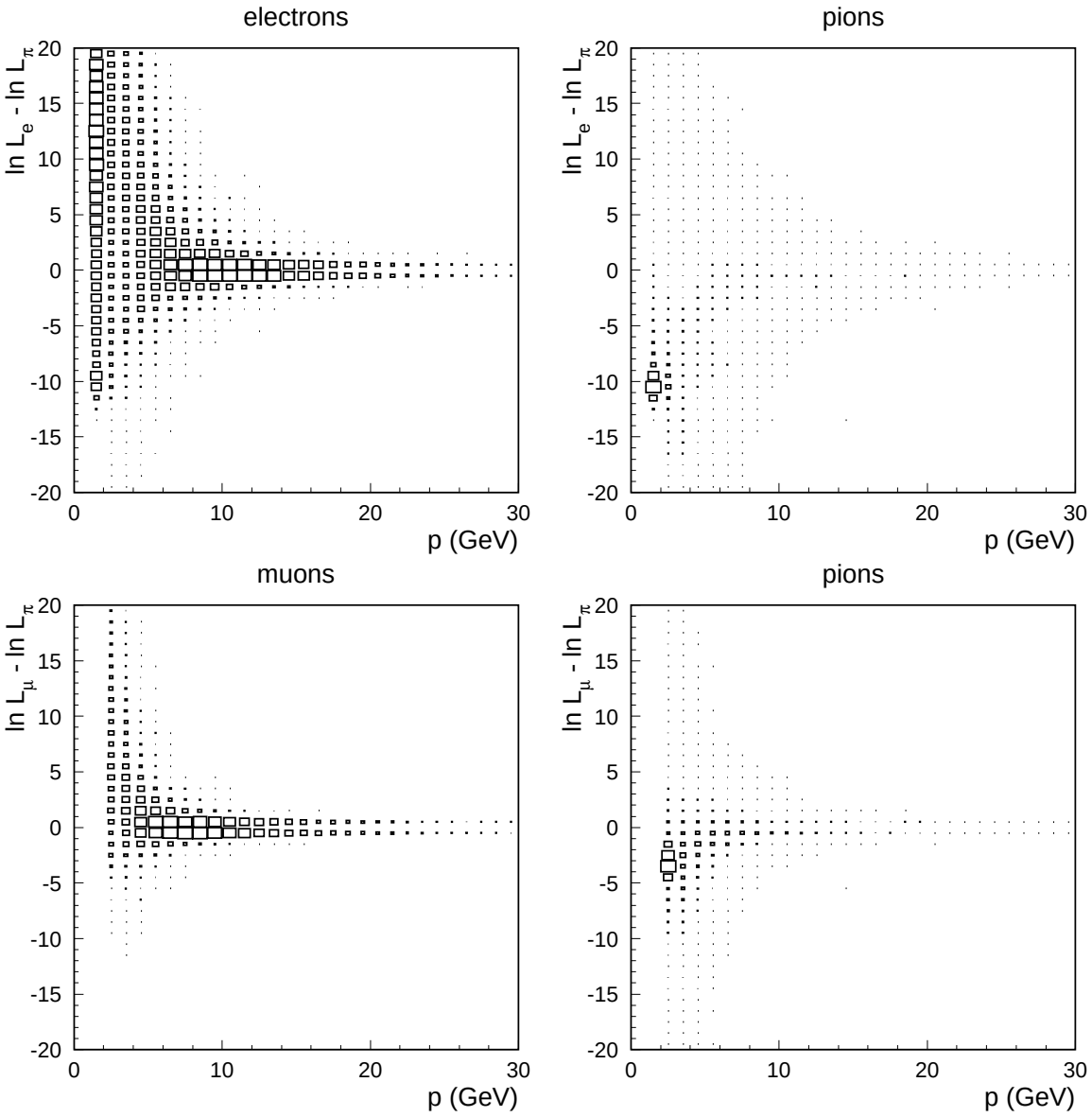


Figure 4.7: Top row: $\ln \mathcal{L}_e - \ln \mathcal{L}_\pi$ versus momentum for electron and pion hypotheses. Bottom row: $\ln \mathcal{L}_\mu - \ln \mathcal{L}_\pi$ versus momentum for muon and pion hypothesis. Pion hypothesis is taken to be the null hypothesis for both electrons and muons.

4.4 Mutually independent tests and the calibration technique

Ideally, one thinks of calibrating the efficiency of a particle counter by observing what fraction of particles is registered by the counter from a population produced by a pure source with accurately known yield. Another scheme, somewhat more realistic for collider environments, is to use two counters: one to preselect (tag) a sample of pure particles from a larger population of many particle types, which then serves as a "source" for the second counter, one that is being calibrated. In that arrangement one has to ensure that the two counters produce mutually independent outputs. Consider, on the contrary, two maximally correlated counters (their 2×2 contingency table of yes-no outputs diagonal). If the first counter preselects, say 70 out of 100 particles, then the second one will count exactly the same 70 particles (because they are maximally correlated). The naively calculated efficiency is then $70/70 = 1$, which is incorrect because it is 0.7.

In collider experiments there are practically no sources of particles other than the events themselves except, perhaps, in the low momentum region. Therefore the bulk of the calibration has to be done using real events. In busy environments like the hadronic Z decays, few tests can produce pure sources of reference particles in large momentum intervals. Here we present a technique for efficiency and mis-ID probability calibration of particle counters that makes use of the concept of mutual independence of two tests. Considerations will be restricted to tests using LAC, CRID, and WIC variables, although the idea can easily be extended to other similar tests.

4.4.1 Calibrating the efficiency

For the sake of definiteness let L and W denote two μ tests, with acceptance regions A_L and A_W in their respective sample spaces. The L and W should suggest that the two tests use LAC and WIC variables respectively, although for the general considerations here the only important thing is that the entire set of variables is partitioned into two subsets, each spanning the sample (sub)space of the corresponding test. According to our notation $f(L \mid \mu)$ is probability that the muon hypothesis is accepted by the L test (similarly for the W and the combined LW test).

The idea is as follows. Call the two tests *mutually independent* if

$$f(LW \mid \mu) = f(L \mid \mu)\, f(W \mid \mu)\,, \tag{4.12}$$

where

$$f(L \mid \mu) = \int_{A_L} f(\mathbf{x} \mid \mu)\, d\mathbf{x} \qquad \text{and} \qquad f(W \mid \mu) = \int_{A_W} f(\mathbf{y} \mid \mu)\, d\mathbf{y}\,. \tag{4.13}$$

The fraction of "true" muons preselected by the L test that are also tagged by the W test is then given by

$$f(W \mid \mu L) = \frac{f(\mu \mid WL)}{f(\mu \mid L)} f(W \mid L) \ , \tag{4.14}$$

where $f(W \mid L)$ can be evaluated in the MC and in the real data alike. The key observation is that

$$f(LW \mid \mu) = f(L \mid \mu) f(W \mid \mu) \quad \Leftrightarrow \quad f(W \mid \mu L) = f(W \mid \mu) \ . \tag{4.15}$$

Therefore, the requirement of mutual independence of the two tests (according to definition (4.12) of the mutual independence), turns the left hand side of Eq. (4.14) into $f(W \mid \mu)$: *the efficiency*. Given that the Monte Carlo value of $f(W \mid L)$ can always be corrected to satisfy

$$f(W \mid L)_{MC} = f(W \mid L)_{data} \ ,$$

which is what we refer to as the *calibration* here, the only condition that has to be fulfilled in order to calculate the efficiency in the data is that

$$\left. \frac{f(\mu \mid WL)}{f(\mu \mid L)} \right|_{data} = \left. \frac{f(\mu \mid WL)}{f(\mu \mid L)} \right|_{MC} \quad \text{(assumption at this stage)} \tag{4.16}$$

In assuming this we took advantage of a common trick, that often the ratio of two quantities can be better determined than either quantity alone (in a more formal language the ratio of $f(\mu \mid L)$ and $f(\mu \mid WL)$ is a weaker quantity than each of the two quantities alone). The correctness of the assumption can be qualitatively confirmed by observing that uncertainties in the numerator and the denominator of (4.16) are correlated, which is guaranteed partly because $A_L \cap A_W \subseteq A_L$. Our objective, therefore, is to determine how accurate the assumption of Eq. (4.16) is. The nature of the problem can be better understood by first going into two limits.

High purity limit This limit corresponds to the usual notion of the calibration and possesses special simplicity: $f(\mu \mid L) = 1$ (pure source), $\Rightarrow$ $f(\mu \mid WL) = 1$, $\Rightarrow$

$$\frac{f(\mu \mid WL)}{f(\mu \mid L)} = 1 \quad \Rightarrow \quad f(W \mid \mu) \equiv f(W \mid L)\,, \tag{4.17}$$

i.e., the W test is directly calibrated by using the outcome of the pure L test as a reference sample, and by assuming that L and W are mutually independent according to the definition of the mutual independence of two tests in Eq. (4.12). In this limit $A_L \cap A_W$ is a large subset of A_L. As a consequence $f(\mu \mid WL)$ and $f(\mu \mid L)$ are highly correlated, leading to the "correctness" of assumption (4.16).

This can be illustrated with the following simple example: let $f(\mu \mid WL) = 0.95 \pm 0.03$ and $f(\mu \mid L) = 0.90 \pm 0.06$. Then one approximately has

$$\frac{f(\mu \mid WL)}{f(\mu \mid L)} = \begin{cases} 1.06^{+0.08}_{-0.07} & f(\mu \mid WL) \text{ and } f(\mu \mid L) \text{ uncorrelated}\,, \\ \\ 1.06^{+0.04}_{-0.03} & f(\mu \mid WL) \text{ and } f(\mu \mid L) \text{ maximally correlated}\,. \end{cases}$$

[We thank Dave Muller here for this illustration as well as for a number of other revisions in this and the previous chapter.]

Low purity limit In this limit $f(\mu \mid L)$, $f(W \mid L)$ is scaled by a large purity ratio, $f(\mu \mid WL)/f(\mu \mid L)$, which in terms of the acceptance regions translates into a statement that $A_L \cap A_W$ is a small subset of A_L. As a consequence, $f(\mu \mid WL)$ is defined locally within A_L, and is susceptible to the local fluctuations in the Monte Carlo yield of muons. In other words, $f(\mu \mid WL)$ and $f(\mu \mid L)$ are little correlated in this limit, which increases the uncertainty of their ratio (such as in the example above).

[Note that the requirement of mutual independence of W and L is a primary requirement. Assume, on the contrary, that the two are maximally correlated. Then $f(\mu \mid WL) = f(\mu \mid L) = f(\mu \mid W)$, $\Rightarrow f(\mu \mid WL)/f(\mu \mid L) = 1$ and $f(W \mid L\,\mu) = f(WL\,\mu)/f(L\,\mu) = 1$, regardless of what the purity of the L test is.]

The goal is therefore to either get as closer as possible to the limit of Eq. (4.17), without losing the statistics, or to independently verify the correctness of assumption (4.16). The latter can in principle be done by using results of the mis-ID probability calibration, Sec. 4.4.4.

Clearly, the W and the L in Eq. (4.14) can be swapped to get

$$f(L \mid \mu W) = \frac{f(\mu \mid WL)}{f(\mu \mid W)}\, f(L \mid W)\,. \tag{4.18}$$

Moreover, this can be combined with (4.14) into a single test as

$$\begin{aligned} f(LW \mid \mu) &= f(L \mid \mu W)\, f(W \mid \mu L) \\ &= \frac{f(\mu \mid WL)}{f(\mu \mid L)}\, \frac{f(\mu \mid WL)}{f(\mu \mid W)}\, f(W \mid L)\, f(L \mid W)\,. \end{aligned} \tag{4.19}$$

On the other hand $f(LW \mid \mu)$ can be calculated directly as

$$f(LW \mid \mu) = \frac{n(LW\mu)}{n(\mu)} = \frac{f(\mu \mid WL)}{f(\mu)}\, f(WL)\,, \tag{4.20}$$

where $n(LW\mu)$ is the number of tagged true muons, $n(\mu)$ total number of true muons, $f(WL)$ the fraction of tracks tagged by the combined test, and $f(\mu)$ fraction of muons in all tracks. So if the two tests are designed to satisfy (4.15), then what is the advantage of using (4.19) over (4.20)?

The question has been partly answered in the discussions of high and low purity limits above: for actual fraction of electrons or muons of about 3% among all tracks, one has

$$\frac{f(\mu \mid LW)}{f(\mu)} \approx 25 \qquad \text{and} \qquad \frac{f(\mu \mid WL)}{f(\mu \mid L)}\, \frac{f(\mu \mid WL)}{f(\mu \mid W)} \approx 7 \tag{4.21}$$

for (4.19) and (4.20) respectively. The second ratio can be additionally lowered by using auxiliary tests to increase the purities of the reference samples. The ratio goes down to about 4 when the impact parameter test is used, and down to near 1 when the γ conversion test is used for electrons, Sec. 4.8.

A simple way of seeing the difference between (4.20) and (4.19) is pictured in Fig. 4.8. The entire set of tracks in the b tagged hemisphere is represented by a square, which is divided into halves representing tracks accepted by the two tests, $L = 1$ and $W = 1$, and tracks rejected by the two tests, $L = 0$ and $W = 0$. In the

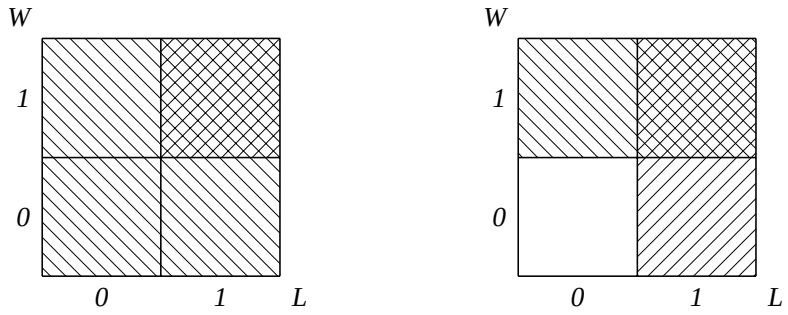


Figure 4.8: A schematic illustration of the difference in the calculation of the ratios of the purities using Eq. (4.20) and Eq. (4.19) respectively.

first schematic, the muon count in $A_L \cap A_W$ is divided with the total muon count, while in the second schematic, the muon count in the $A_L \cap A_W$ is divided with the muon counts in the A_L and the A_W respectively, and then the product of the two taken. Note that in the second approach tracks that are rejected by both tests, $L = 0$ and $W = 0$, are excluded from the calculations.

But apart from having a simple calibration recipe, what else does one get from the idea? For one thing a natural procedure for removing the correlation between the two tests while preserving their discriminating capabilities.

4.4.2 Removing the correlation between the two tests

It is very unlikely that for a given set of variables two tests will be completely uncorrelated regardless of how the set of variables is partitioned between the two tests. This is particularly true for LAC variables which are all mutually correlated to some degree.

There is an elegant way to use the parameterization of the acceptance region A in terms of the parameter k of the likelihood ratio test, Corollary (4.2), to correct two individual tests so as to make them mutually independent. In essence, the procedure introduces a new correlation to compensate for the existing one.

The yes-no outcomes of the L and W tests can be represented in a more functional form as

$$L = \begin{cases} 1 & \text{on } A_L \\ 0 & \text{on } A_L^c \end{cases} \quad \text{and} \quad W = \begin{cases} 1 & \text{on } A_W \\ 0 & \text{on } A_W^c \end{cases} .$$

Since A_L and k_L are one-to-one (4.2), one can write $L = L(k_L)$. Similarly $W = W(k_W)$. Therefore (4.14) turns into

$$f(L(k_L), W(k_W) \mid \mu) = f(L(k_L) \mid \mu)\, f(W(k_W) \mid \mu)\,. \tag{4.22}$$

For $L = 0$ one can write $f(\bar{L}(k_L) \mid \mu)$, etc.

We define the correlation function of the two tests as

$$\begin{aligned} g(L, W) &\equiv g(L(k_L), W(k_W)) \\ &= f(L(k_L), W(k_W) \mid \mu) - f(L(k_L) \mid \mu)\, f(W(k_W) \mid \mu)\,, \end{aligned} \tag{4.23}$$

in terms of which Eq. (4.12) becomes $g(LW) = 0$. [It is easy to show that in general $g(LW) = g(\bar{L}\bar{W}) = -g(L\bar{W}) = -g(\bar{L}W)$.] Note that the solution of $g(LW) = 0$, assuming it does exist, is not a single pair of k_L and k_W, but rather a collection of such pairs, each uniquely corresponding to a pair of acceptance regions A_L and A_W.

It turns out in practice that for the entire range of interest of k_L and k_W, $g(LW) = 0$ may either not exist at all, or the $g(LW)$ may be such a slowly varying function of k_L and k_W that the values of $f(L(k_L), W(k_W) \mid \mu)$ at which $g(LW) = 0$ is of no practical interest (either too small, or large enough but accompanied with unacceptably large mis-ID probability).

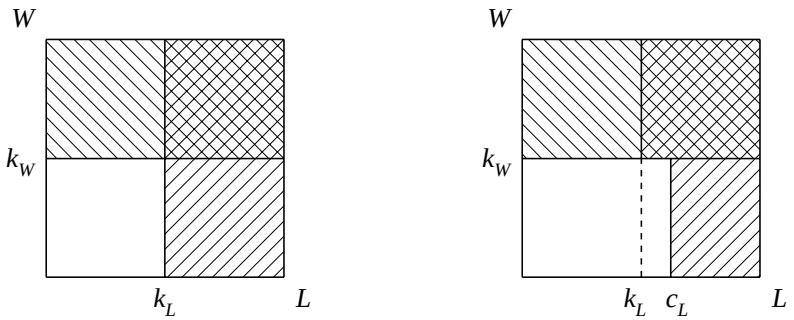


Figure 4.9: A schematic illustration of the difference between the original calibration scheme, Eq. (4.3), and the calibration scheme in which the correlation between the two tests is completely removed, Eq. (4.24).

Alternatively, one can use a simple trick that allows the $g(LW)$ to change more rapidly. Introduce a new constant c_L with exactly the same properties as k_L regarding

the L test (just think of c_L as one particular value of k_L). The trick is then to redefine the acceptance region of the L test in terms of k_L and c_L as

$$A_L = \begin{cases} \{\mathbf{x} : f(\mathbf{x} \mid \mu) - k_L f(\mathbf{x} \mid \text{non-}\mu) \geq 0\} & \text{when } W = 1\,, \\ \{\mathbf{x} : f(\mathbf{x} \mid \mu) - c_L f(\mathbf{x} \mid \text{non-}\mu) \geq 0\} & \text{when } W = 0\,. \end{cases} \tag{4.24}$$

Therefore L turns into a "function" of two parameters, k_L and c_L, and can be written as

$$f(L(k_L, c_L) \mid \mu) = f(L(k_L), W(k_W) \mid \mu) + f(L(c_L), \bar{W}(k_W) \mid \mu)\,.$$

The new correlation function

$$g(LW) = f(L(k_L, c_L), W(k_W) \mid \mu) - f(L(k_L, c_L) \mid \mu)\, f(W(k_W) \mid \mu)$$

is now a rapidly changing function of c_L, which can easily be chosen to satisfy $g(LW) = 0$ for any values of k_L and k_W. The modified L test is depicted in the right schematic of Fig. 4.9. In principle, one can symmetrically work with the modified W test parameterized in terms of k_W and c_W.

4.4.3 Calibrating the mis-ID probability

The *mis-ID probability* is probability that a particle other than a lepton is identified as such, and is calibrated using $> 99.9\%$ pure source of hadrons from $K^0_s \to \pi^+\pi^-$ decays. The hadron test based on these decays will be called the K test.

We continue to use muons as an example. By expanding on the idea of mutual independence of the two tests, which holds very strongly here because the K test involves only CDC and the L test uses no CDC, one has

$$f(L \mid \text{non-}\mu, K) = \frac{f(\text{non-}\mu \mid K, L)}{f(\text{non-}\mu \mid K)}\, f(L \mid K)\,, \tag{4.25}$$

where

$$f(KL \mid \text{non-}\mu) = f(K \mid \text{non-}\mu) f(L \mid \text{non-}\mu) \;\Leftrightarrow\; f(L \mid \text{non-}\mu K) = f(L \mid \text{non-}\mu)\,, \tag{4.26}$$

turns the left hand side of Eq. (4.25) into $f(L \mid \text{non-}\mu)$: the *mis-ID probability*.

The ratio of the two purities in Eq. (4.25) is similar quantity as that in Eq. (4.14), with one important distinction: in (4.14) both tests are *muon* tests, targeting regions in their respective sample spaces that are rich in muons. Therefore purity of the combined LW test, $f(\mu \mid LW)$, is higher than the purity of the L test alone, $f(\mu \mid L)$. Hence $f(\mu \mid LW)/f(\mu \mid L)$ is always > 1.

In Eq. (4.25) on the contrary, the L and the K are the muon and the hadron tests respectively, targeting mutually disjoint regions of the sample space. In a relatively pure set of hadrons selected by the K test, the L test will point to a subset of tracks that are more likely to be muons than hadrons, therefore making the *hadron purity* of the combined KL test, $f(\text{non-}\mu \mid K, L)$, lower than the purity of the K test alone, $f(\text{non-}\mu \mid K)$. Hence $f(\text{non-}\mu \mid KL)/f(\text{non-}\mu \mid K)$ is always < 1.

4.4.4 Relation between efficiency and mis-ID probability

Calibrated mis-ID probability from Eq. (4.25) can be used to calibrate the ratio of the purities in Eq. (4.14), and hence to calibrate the efficiency on the left hand side of the same equation. Since $f(\mu \mid L) + f(\text{non-}\mu \mid L) = 1$, etc., one has

$$r \equiv \frac{f(\mu \mid LW)}{f(\mu \mid L)} = \frac{1 - f(\text{non-}\mu \mid LW)}{1 - f(\text{non-}\mu \mid L)}\,. \tag{4.27}$$

The denominator (and in the same way the numerator) can be transformed as

$$f(\text{non-}\mu \mid L) = f(L \mid \text{non-}\mu)\,\frac{f(\text{non-}\mu)}{f(L)} = f(L \mid \text{non-}\mu\, K)\,\frac{f(\text{non-}\mu)}{f(L)}\,,$$

where the second relation has been obtained by taking (4.26) into account. Clearly, to evaluate $f(\text{non-}\mu \mid L)$, the fraction of non-muons $f(\text{non-}\mu)$ among all tracks in hadronic events has to be estimated in the first place. This can be done by using

$$f(\mu) + f(\text{non-}\mu) = 1\,,$$

where $f(\mu) = \eta\tilde{f}(\mu)$, η efficiency estimated as described in Sec. 4.4.1, and $\tilde{f}(\mu)$ the yield of muons labeled by the muon test (obtained by subtracting the mis-ID rate from all the tracks labeled as muons).

A really nice thing about $f(\text{non-}\mu)$ is that r from (4.27) is not very sensitive to it. This can be seen as follows:

$$r = \frac{1 \; - \; f(\text{non-}\mu)\,\dfrac{f(LW \mid \text{non-}\mu\, K)}{f(LW)}}{1 \; - \; f(\text{non-}\mu)\,\dfrac{f(L \mid \text{non-}\mu\, K)}{f(L)}} \equiv \frac{1-\lambda x}{1-\lambda y}\,, \tag{4.28}$$

where $\lambda \equiv f(\text{non-}\mu)$, $x \equiv f(LW \mid \text{non-}\mu\, K)/f(LW)$, and $y \equiv f(L \mid \text{non-}\mu\, K)/f(L)$ have been introduced to simplify the notation. Then

$$\frac{dr}{d\lambda} = \frac{y-x}{(1-\lambda y)^2}$$

which goes to 0 in the limit of pure L and LW tests ($x = y = 0$).

4.5 Track classes and the parameterizations of the tests

Two classes of tracks are used to study the efficiency and the mis-ID probability of lepton identification: the VXD and the VEE class.

The VXD class of tracks

- At least 40 vector hits in the CDC.
- At least one hit in the VXD.
- 3D impact parameter < 0.5 cm.
- Helix parameter $1/p_\perp < 4$ GeV^{-1} is formally used although it is redundant because we work with total momentum cut of 1 GeV.

The VEE class of tracks

- At least 40 vector hits in the CDC.
- $\rho = \sqrt{x^2+y^2}$ of the first fit point < 60 cm.

- 3D impact parameter > 0.01 cm.
- $\chi^2/\mathrm{ndf} < 5$ for the drift time fit per degree of freedom.

The VXD class is used for efficiency and mis-ID studies that do not rely on tracks originating in the decays of long-lived particles like K_s^0 and Λ, or on γ conversions. The VXD class is also used in the actual calculations of the branching fractions. The assumptions is that all tracks of interest in this class originate either in the fragmentation processes, or in the decays of the short-lived particles.

The VEE class is designed to select tracks from $\gamma \to e^+e^-$ conversions and $K_s^0 \to \pi^+\pi^-$ and $\Lambda \to p\pi^-$ decays with high efficiency. Many of these processes occur well beyond the outermost layer of the VXD, which is the primary reason why no VXD hits are required. This class is used only for the calibration purposes.

In a complex dynamical system like the hadronic Z decays, probability relations like (4.14) and (4.25) depend on a number of variables associated to individual tracks: p, $p_\perp$, polar angle, opening angle between track direction and jet axis, impact parameter, and others. The efficiencies vary substantially in p, while in $\cos\theta$ they are almost flat up to $|\cos\theta|$ of about 0.7. The dependencies on other variables turn out to be at the level of small corrections. For example, position of a track within the jet has to be taken into account due to the large anisotropy of the event topologies at the Z, which is reflected in the way the tracks are linked to the LAC clusters: in the middle of a narrow jet, the track-cluster association is more ambiguous than outside the jet core. Moreover, the LAC clusters tend to overlap more in the core of a jet than on its periphery, which lowers cluster resolution in almost all its attributes.

Another problem related to the anisotropies of events at the Z has to do with the difference between the topologies of the uds and the heavy flavor events, which affects translation of results obtained using all hadronic events (which are needed to increase the statistics), into results specific to $b\bar{b}$ events. The results will be shown in Sec. 4.7.1 and Sec. 4.7.2 for the mis-ID probability studies.

As for the efficiencies, the net effect of all variables other than p is that the efficiency as a function of p is slightly different for leptons from $b \to l$ and $b \to c \to l$ decays, due to the differences in their distributions as a function of the opening angle between the track and the jet direction. This is more the case for muons due to their mass, which is comparable to the average value of the $p_\perp$.

Studies in this chapter will be restricted to the subset of all *reconstructed* tracks, not all true charged particles, with restriction imposed on their polar angle that $|\cos\theta| < 0.7$. The main correction to these numbers will come from the tracking efficiency, and somewhat smaller from the extensions in the polar angle. All other corrections taken together turn out to be at the level of $1 - 2\%$. All corrections

to the efficiencies calculated in this chapter will be jointly presented in Ch. 6 as a preliminary to the calculations of the branching fractions.

4.6 The calibration procedure

The framework of the hypothesis testing and the calibration technique has been outlined in sections 4.1 through 4.4. In this section we show how is the calibration done in practice.

For each momentum bin, a 2×2 contingency table of the yes-no outcomes of the two tests is formed. For tests E and H, for example, the contingency table consists of the number of counts $n(EH)$, $n(E\bar{H})$, $n(\bar{E}H)$, and $n(\bar{E}\bar{H})$. One takes the contingency table for the data and loops through the MC sets until

$$n(EH)_{MC} = n(EH)_{data}\,, \quad n(E\bar{H})_{MC} = n(E\bar{H})_{data}\,, \quad \cdots\,, \tag{4.29}$$

in each momentum bin separately, while simultaneously incrementing elements of the 2×2 Monte Carlo contingency tables for electron hypothesis $n(eEH)$, $n(eE\bar{H})$, ..., and electron null hypothesis $n(\text{non-}eEH)$, $n(\text{non-}eE\bar{H})$, etc. Other counts are easily obtained from these as

$$n(eH) = n(eEH) + n(eE\bar{H})\,,$$

etc. Various probabilities are also easy to obtain directly as

$$f(E \mid eH) = \frac{n(eEH)}{n(eH)}\,, \quad f(e \mid EH) = \frac{n(eEH)}{n(EH)}\,, \quad \cdots\,.$$

A MC sample obtained that way is called a *calibrated* MC sample. It corresponds in size to a single (and only one) real data sample. Since the total number of generated MC events is roughly 12 times the size of the real data sample, one can continue looping through the MC data sets to generate 2nd, 3rd, 4th, ..., calibrated MC samples according to the procedure described above. Results of that work will be shown in the next section following a brief comment on the selection of the parameter k of the H test, which is more related to the reduction of errors in the calculation of the branching fractions than to the calibration procedure.

The efficiency contributes to the overall error in the measurement of the $\mathcal{B}(b\to l)$ and $\mathcal{B}(b\to c\to l)$ mostly through the scaling of the statistical errors of the directly observed branching fractions ($\mathcal{B}_L = \tilde{\mathcal{B}}_L/\eta \;\Rightarrow\; \sigma_{\mathcal{B}_L} = \sigma_{\tilde{\mathcal{B}}_L}/\eta$ and similarly for $\mathcal{B}_U$). Errors in $\mathcal{B}_L$ and $\mathcal{B}_U$ from the uncertainty in η are relatively small. On the other hand, the $\tilde{\beta}+\delta$ has to be subtracted from the total count of the tracks tagged as leptons, which also introduces an error into $\mathcal{B}_L$ and $\mathcal{B}_U$, this time due to the uncertainty in the $\tilde{\beta}+\delta$, which is large.

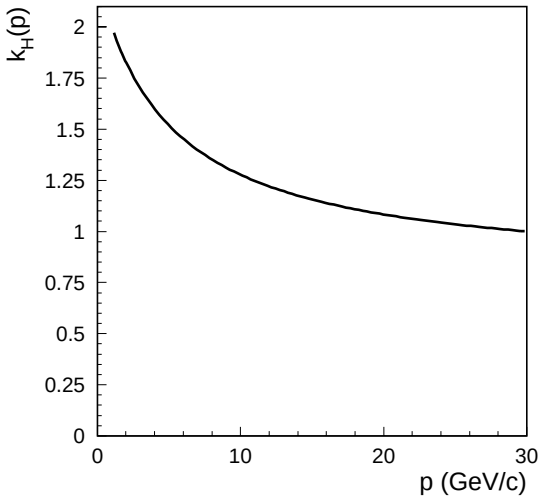


Figure 4.10: Momentum dependent parameter $k_H(p)$ for the electron H test, parameterized as in Eq. (4.30).

In practice, one does not get the smallest overall errors for $\mathcal{B}_L$ and $\mathcal{B}_U$ for fixed values for k in the entire momentum range. Instead, the k turns out to be varying function of momentum. After a number of trials with different momentum dependencies of the k in both electron tests, E and H, we managed to reduce the overall errors in $\mathcal{B}_L$ and $\mathcal{B}_U$ somewhat by using

$$k_H(p) = \frac{a}{p+b} + c \;=\; \frac{7.3259}{p+5.059} + 0.791\,, \tag{4.30}$$

Fig. 4.10. All other tests use fixed $k=1$ in entire momentum range. The a and b are in GeV while the c is dimensionless. Fig. 4.11 compares the efficiencies and the mis-ID probabilities for varying and four fixed values of k_H across the entire momentum range. The improvements are mostly visible in the lower momentum range, where the reduction of the mis-ID probability is roughly by factor $2-3$, while the efficiency is reduced by less than 30%.

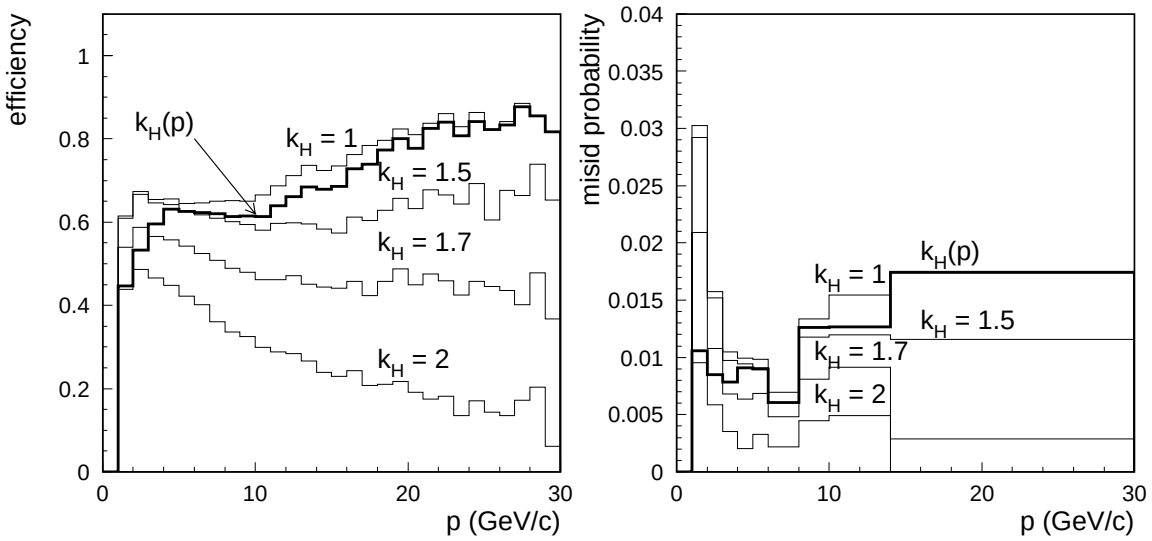


Figure 4.11: Effects of applying $k_H(p)$ from Fig. 4.10 compared against four fixed values of k_H. In the first bin the reduction in the mis-ID probability is by more than factor 3, while the loss in the efficiency is around 30%.

The summary of lepton tests used in this analysis follows:

- E: Electron test using (4.6).
- H: Electron test using (4.10).
- W: Muon test using (4.9).
- L: Muon test using (4.11).
- EH: Combined electron test.
- LW: Combined muon test.

Auxiliary tests are used to get purer reference samples of either hadrons or leptons, and will be discussed in the subsequent sections.

4.7 Results of the mis-ID probability calibrations

Hadron test based on K_s^0 decays will be called the K test.

The K test:

- Tracks from the VEE class are used.
- $|\cos\theta^*| < 0.8$, where θ^* is the opening angle between the pion candidates in the K_s^0 rest frame.
- $|m_{\pi\pi} - 0.497| < 0.02$ GeV, where $m_{\pi\pi}$ is the invariant mass of the pion candidates from $K_s^0 \to \pi^+\pi^-$ decays, for the mass of K_s^0 taken to be 497 MeV.
- Tracks that are also labeled as electrons from γ conversions are rejected. The γ conversion test uses only $m_{ee} < 0.03$ GeV, where m_{ee} is the invariant mass of the electron pair.

The box-cut in $\cos\theta^*$ and $m_{\pi\pi}$ is pictured in Fig. 4.12. The VEE tracks are selected using `ZXFIND` for vertex finding (Gary Gladding), in combination with `FINDK0` (Dave Muller), which calculates Lorentz transformations and $\cos\theta^*$ as a function of $m_{\pi\pi}$ in the K_s^0 rest frame. When tracks that are labeled as electrons from γ conversion are rejected, the hadron purity of the remaining tracks reaches 99.98% in all hadronic events.

4.7.1 Electrons

The goal is to determine a scaling factor between the mis-ID probability in the data and the MC:

$$\text{scaling factor} = \frac{f(EH \mid \text{non-}e)_{data}}{f(EH \mid \text{non-}e)_{MC}}\,, \tag{4.31}$$

which is used to scale the MC mis-ID rate in $b\bar{b}$ events. The ratio is expected to be $\mathcal{O}(1)$ regardless of the parameterization. We consider the mis-ID probability only function of p, and the variations in other variables to be small. Therefore, the variation of the *ratio* in (4.31) is taken to be of even lower order.

Since the scaling factor (4.31) is used to correct the mis-ID rate in $b\bar{b}$ events, we want to verify that it is unchanged in the larger classes, which are needed because of the larger statistics. The mis-ID probability is therefore estimated by applying the K test in the following classes of events:

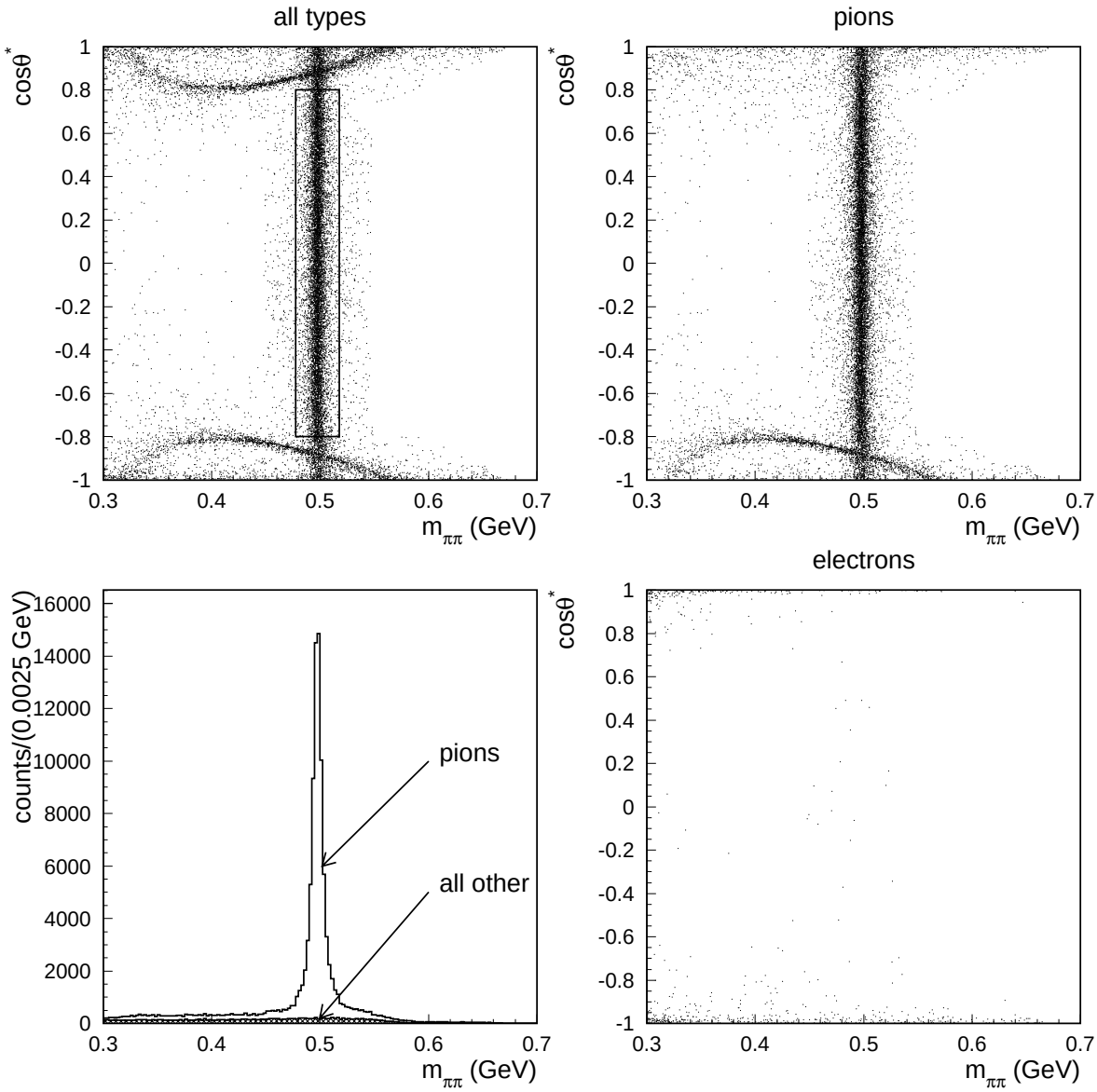


Figure 4.12: Scatter plots of $\cos\theta^*$ versus $m_{\pi\pi}$ for various particle types, and its projection on $m_{\pi\pi}$ (lower left plot). The box-cut in the top left plot selects pions with purity $> 99.9\%$. The overall purity of hadrons is even higher.

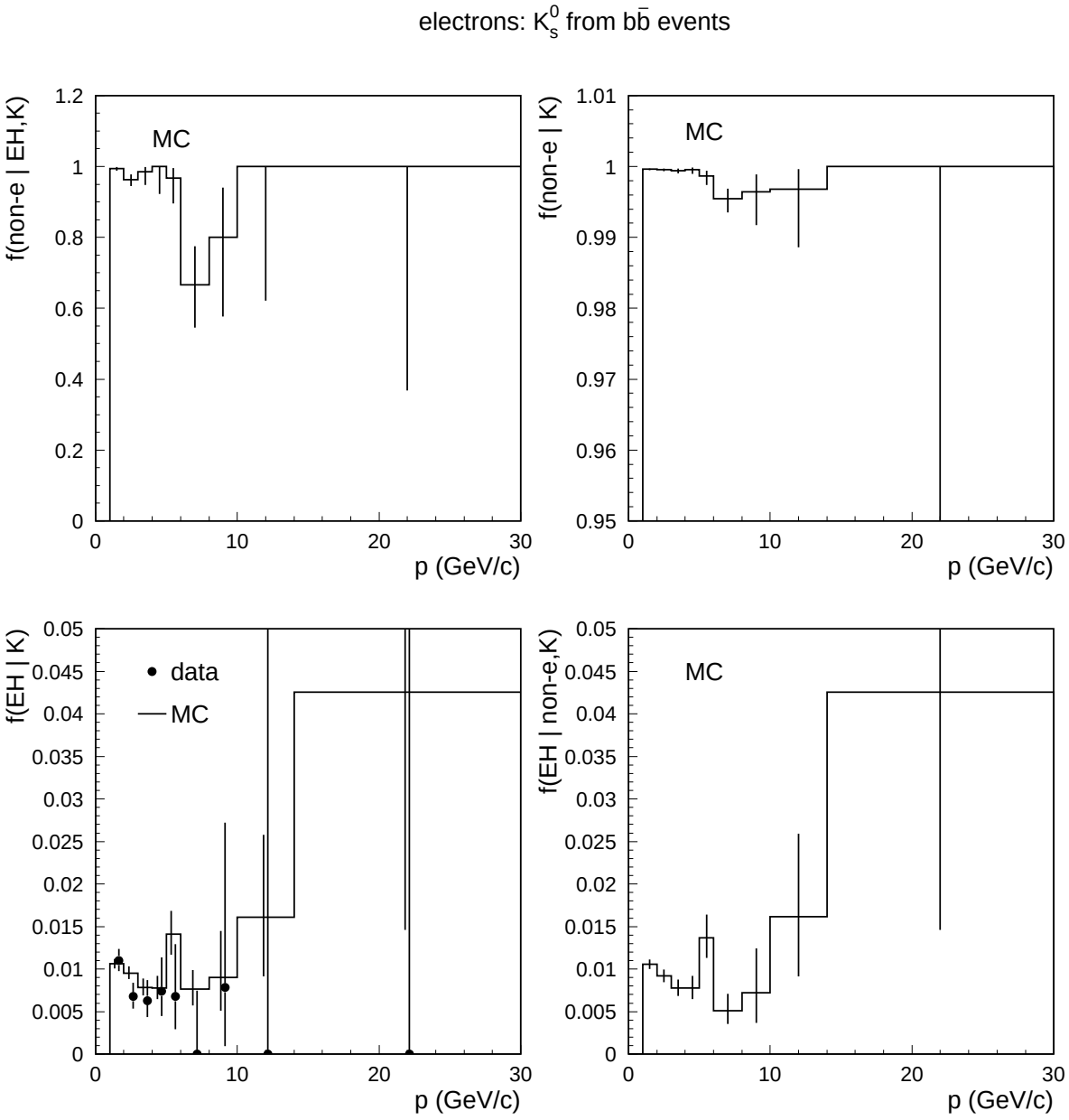


Figure 4.13: Individual terms of Eq. (4.25) with the L test replaced by the combined EH electron test. The sample of $K^0_s \to \pi^+\pi^-$ decays is taken from about 97% pure $b\bar{b}$ events (invariant mass cut of > 2 GeV in at least one hemisphere).

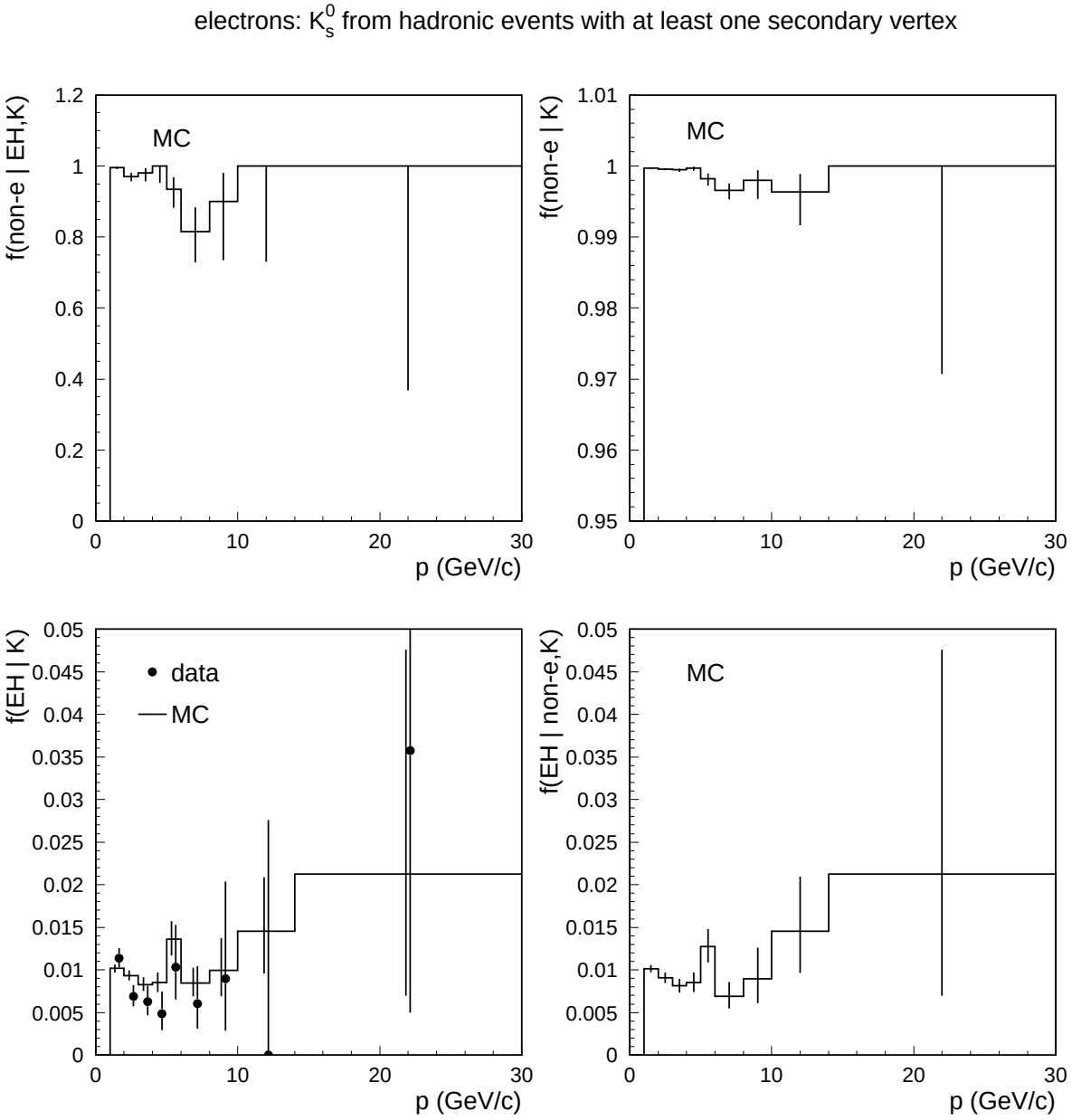


Figure 4.14: Same as in Fig. 4.13 but with somewhat larger size of the $K^0_s \to \pi^+\pi^-$ decay sample taken from all events with at least one displaced vertex. These events contain about 65% of $b\bar{b}$ events.

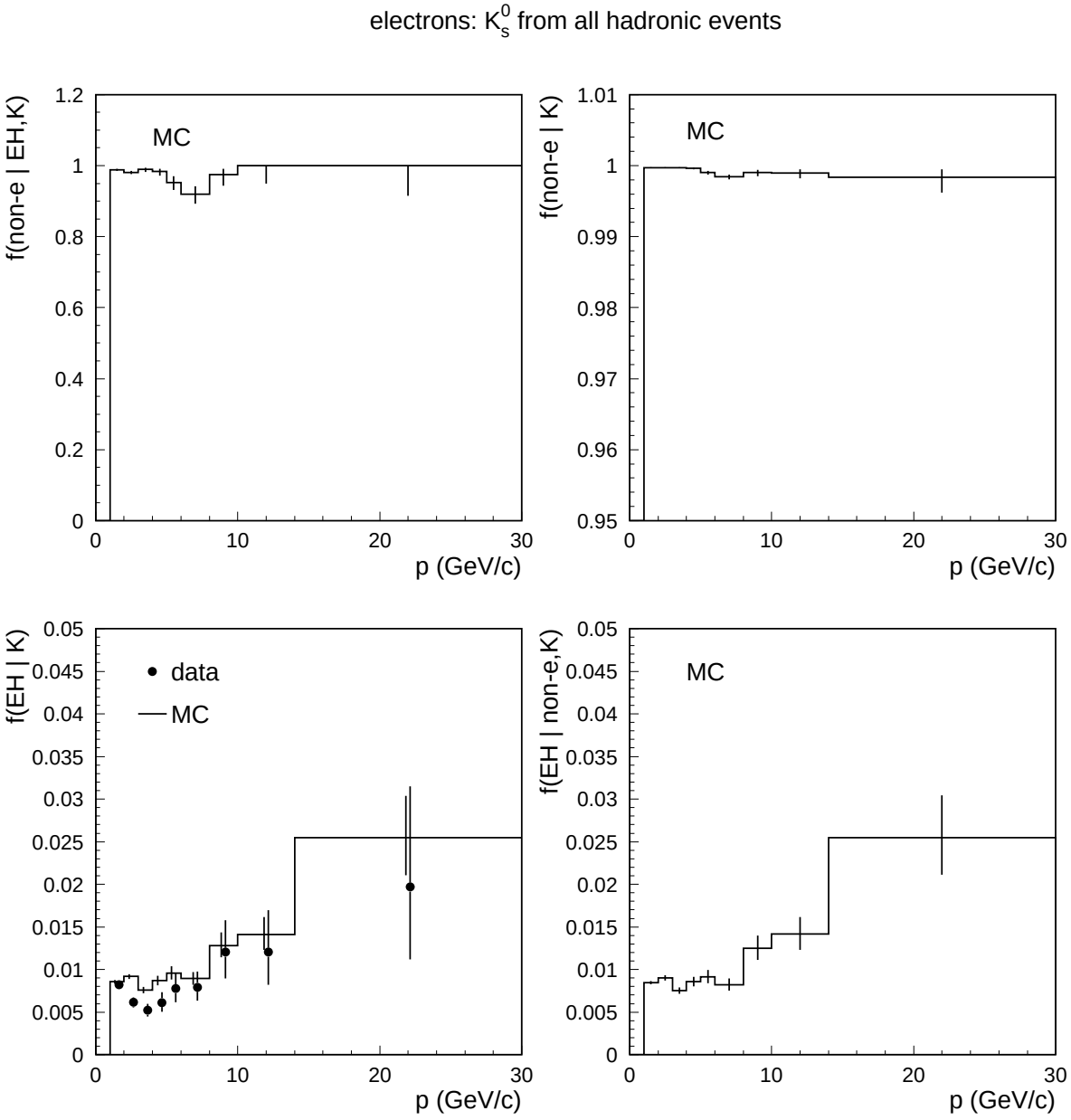


Figure 4.15: Same as in Fig. 4.13 and Fig. 4.14 but with about 10 times larger $K_s^0 \to \pi^+\pi^-$ decay sample taken from all hadronic events. A very good news here is that the trend of $f(EH \mid K)$ in the data relative to the MC is close to that of Fig. 4.13 and Fig. 4.14.

- $b\bar{b}$ events, Fig. 4.13,
- Hadronic events with at least one secondary vertex, Fig. 4.14,
- All hadronic events, Fig. 4.15.

Plots in Fig. 4.13 through Fig. 4.15 feature all ingredients of Eq. (4.25) except that the muon L test has been replaced with the combined electron EH test. Since no significant differences in the behavior of the mis-ID probability $f(EH \mid \text{non-}e, K)$ have been observed between the classes, we use the class of all hadronic events for the calibration, which has the statistics roughly 20 times higher than the class of $b\bar{b}$ events.

The calibration procedure is similar to the one outlined in Sec. 4.6 for the efficiency calibration: a single calibrated MC sample is obtained by filling 2×2 contingency tables of yes-no outcomes of the EH and K tests until

$$n(EH, K)_{MC} = n(EH, K)_{data}\,, \quad n(EH, \bar{K})_{MC} = n(EH, \bar{K})_{data}\,, \quad \cdots\,, \tag{4.32}$$

in each momentum bin separately, while simultaneously incrementing counts in the 2×2 Monte Carlo contingency tables for electron null hypothesis $n(EH, \text{non-}e, K)$, $n(\overline{EH}, \text{non-}e, K)$, etc. Other counts are easily obtained from these as

$$n(\text{non-}e, K) = n(EH, \text{non-}e, K) + n(\overline{EH}, \text{non-}eK)\,,$$

etc. The mis-ID probability is then calculated directly as

$$\begin{aligned} f(EH \mid \text{non-}e, K) \equiv f(EH \mid \text{non-}e) &= \frac{n(\text{non-}e, EH, K)}{n(\text{non-}e, K)} \\ &= \frac{f(\text{non-}e \mid EH, K)}{f(\text{non-}e \mid K)}\, f(EH \mid K)\,, \end{aligned}$$

which we write as $z = r\,x$ to simplify the notation somewhat. The ratio of the two purities r is not directly used in these calculations, but is needed as a controlling parameter (how close it is to 1), and for the calculation of the calibration errors.

It is the ratio r that contains our "definition" of electron in terms of the detector variables (LAC, WIC, CRID, etc.). The origins of the uncertainties in r are very

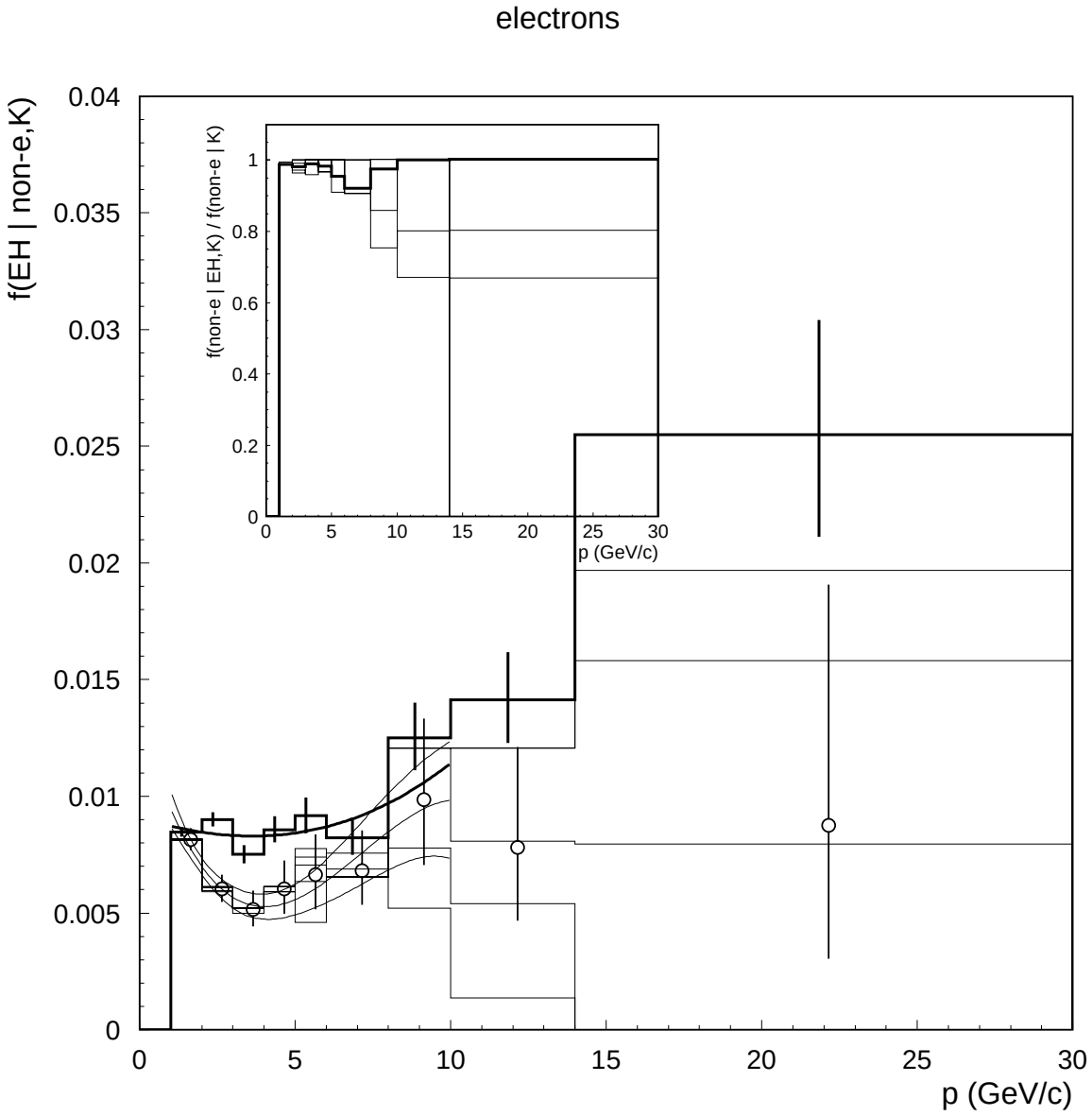


Figure 4.16: Results of the electron mis-ID probability calibration. The thick line histogram is the uncalibrated mis-ID probability. The thin line histograms are calibrated mis-ID probabilities from random samples equivalent in size to the $93-98$ data sample. Open circles represent average values of the calibrated distributions, with the error bars corresponding to only statistical uncertainties of the single $93-98$ data sample.

complicated to disentangle. They are partly statistical, partly due to the widths of the intrinsic distributions used by the event generators, and partly related to the cardinality of the space of objects generated by the Monte Carlo.

To avoid the possibility of underestimating the calibration error, we adopted the "worst case scenario" approach, in which the r and the x are considered completely uncorrelated. In reality, they *are* correlated but not maximally. Assuming then r and x Gaussian with variances σ_r^2 and σ_x^2 one has

$$\sigma_z = \sqrt{\sigma_r^2\,\sigma_x^2 \;+\; \sigma_r^2\,x^2 \;+\; r^2\,\sigma_x^2} \;\approx\; \sqrt{\sigma_r^2\,x^2 \;+\; r^2\,\sigma_x^2}\,, \tag{4.33}$$

as usual. The calibration results are shown in Fig. 4.16.

4.7.2 Muons

The muon mis-ID probability calibration is an exact replica of the procedure for electrons. The combined electron EH test is replaced with the combined muon LW test, while exactly the same K test is used relatively pure source of hadrons.

The goal is again to verify that scaling factor (4.31) remains unchanged when the mis-ID probability is estimated by applying the K test in the three classes of events:

- $b\bar{b}$ events, Fig. 4.17,
- Hadronic events with at least one secondary vertex, Fig. 4.18,
- All hadronic events, Fig. 4.19.

Plots in Fig. 4.17 through Fig. 4.19 have precisely the same meaning as plots in Fig. 4.14 through Fig. 4.15 for electrons. Comparison of the lower left plots in the figures shows that ratios of $f(LW \mid K)$ in the data and the MC remain unchanged to within statistics. The MC values of the mis-ID probability (the lower right plots) are also steady.

The final results of the muon mis-ID probability calibration are shown in Fig. 4.20. The only major difference between electrons and muons is in the purities of the K sample, which are somewhat higher for electrons, and translate into $r = f(\text{non-}\mu \mid LW, K)/f(\text{non-}\mu \mid K) \approx 0.9$ for muons, and $r = f(\text{non-}e \mid EH, K)/f(\text{non-}e \mid K) \approx 1$ for electrons.

* * *

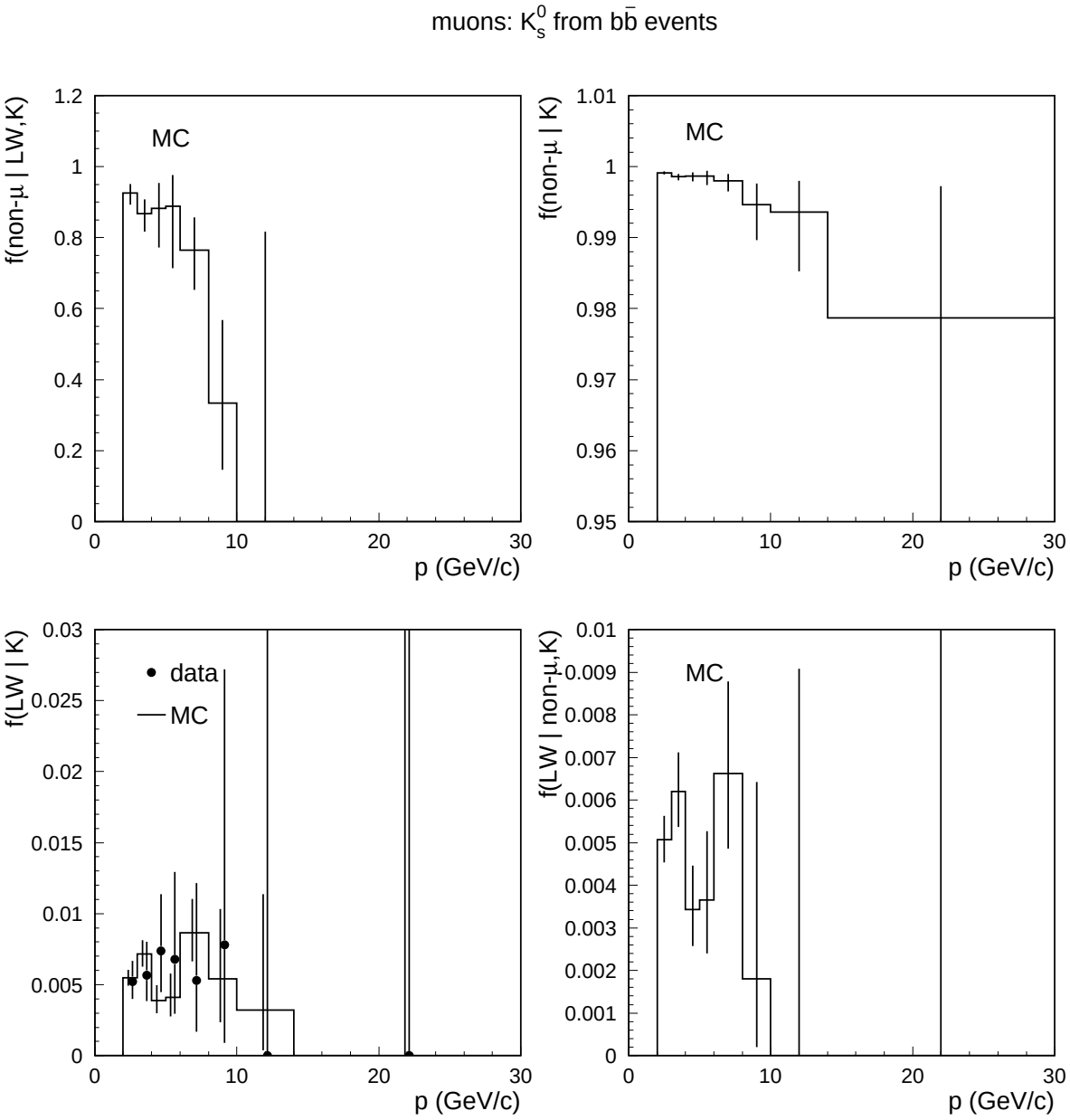


Figure 4.17: Same as in Fig. 4.13 but for muons: individual terms of Eq. (4.25) with the L test replaced by the combined LW muon test. The sample of $K^0_s \to \pi^+\pi^-$ decays is taken from about 97% pure $b\bar{b}$ events (invariant mass cut of > 2 GeV in at least one hemisphere).

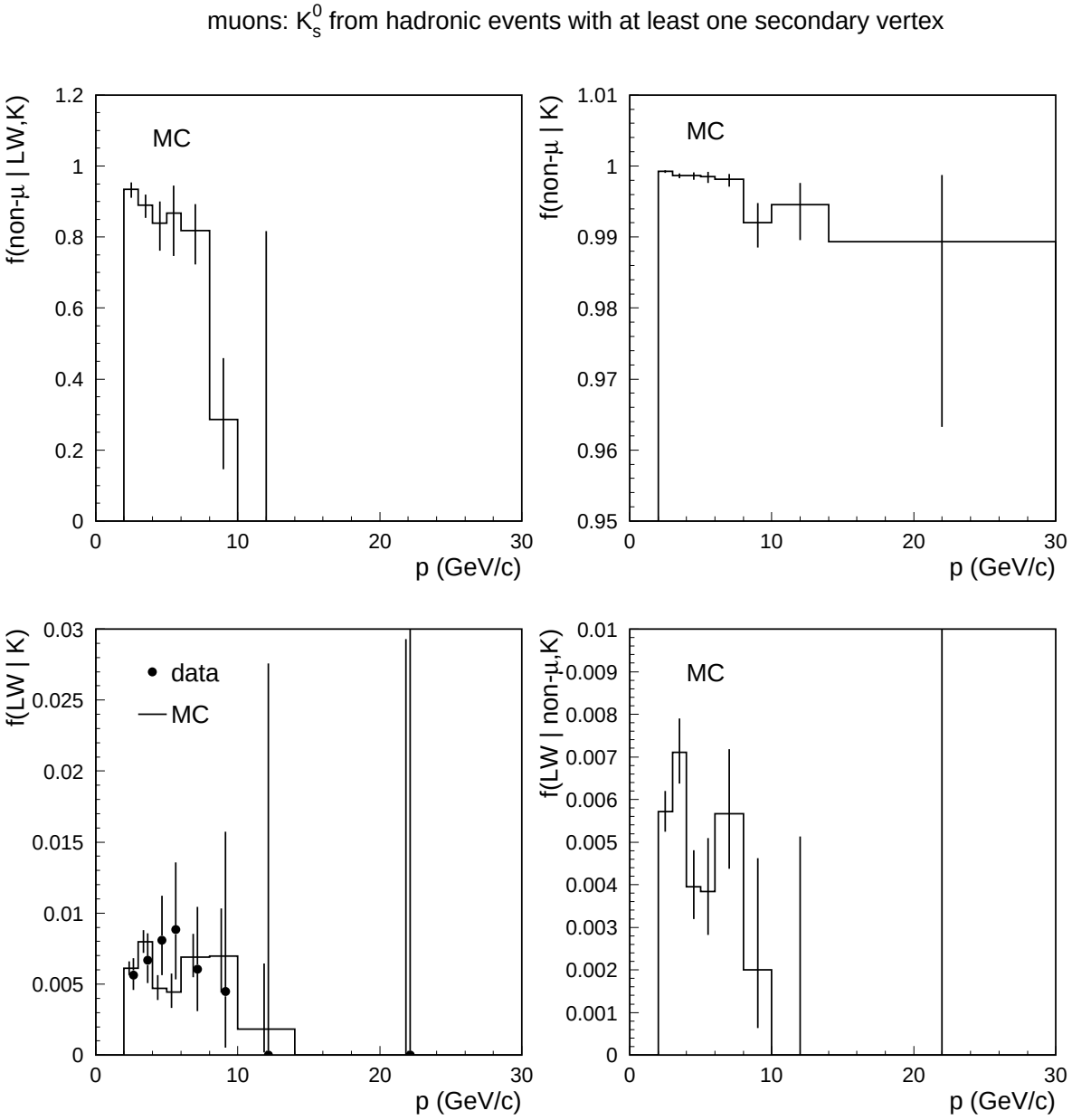


Figure 4.18: Same as in Fig. 4.17 but with somewhat larger size of the $K^0_s \to \pi^+\pi^-$ decay sample taken from all events with at least one displaced vertex. These events contain about 65% of $b\bar{b}$ events.

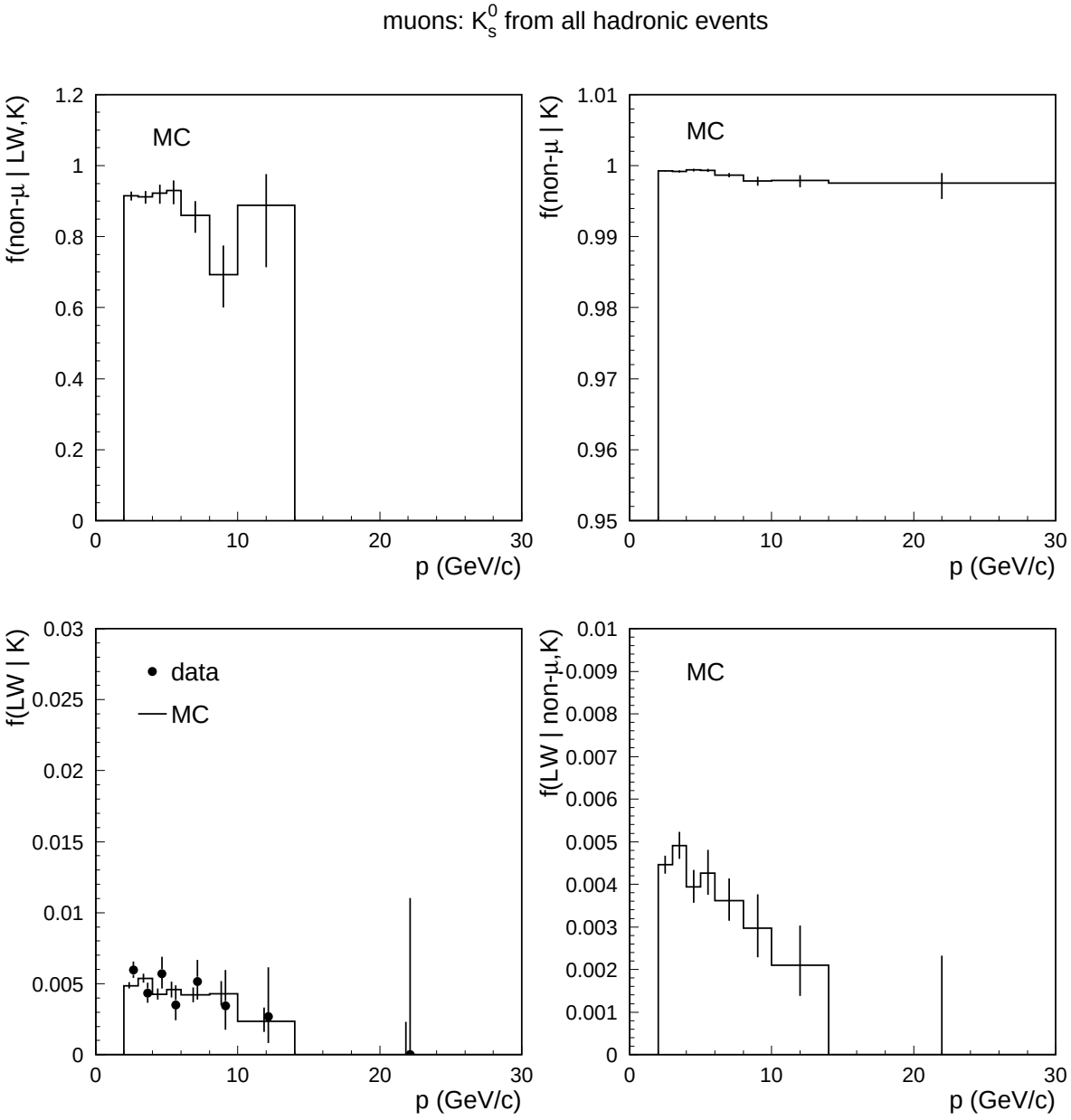


Figure 4.19: Same as in Fig. 4.17 and Fig. 4.18 but with about 10 times larger $K^0_s \to \pi^+\pi^-$ decay sample taken from all hadronic events. A very good news here is that the trend of $f(LW \mid K)$ in the data relative to the MC is close to that of Fig. 4.17 and Fig. 4.18.

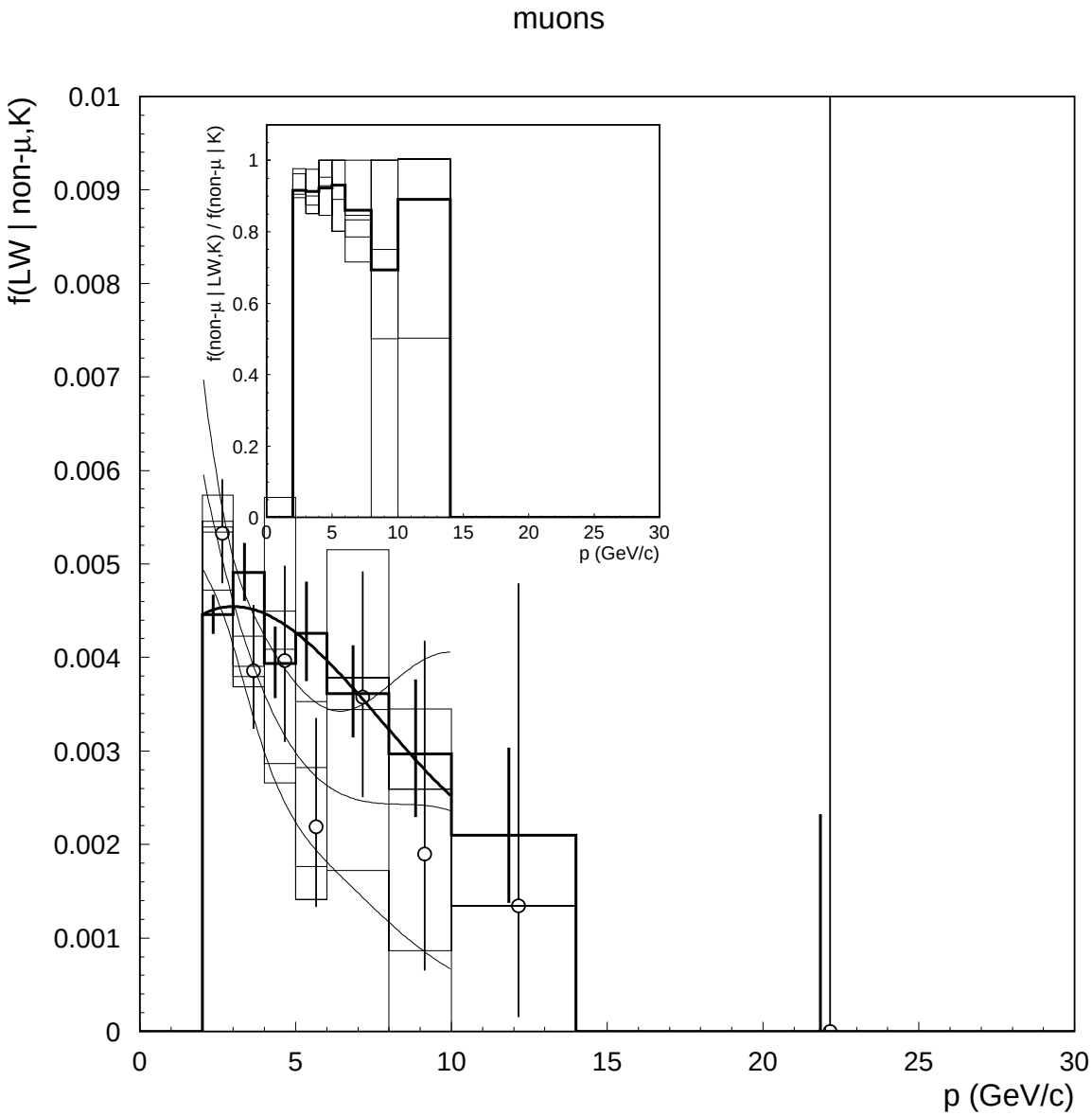


Figure 4.20: Results of the muon mis-ID probability calibration. The thick line histogram is the uncalibrated mis-ID probability. The thin line histograms are calibrated mis-ID probabilities from random samples equivalent in size to the $1993-98$ data sample. The open circles represent the average values of the calibrated distributions, with the error bars corresponding to only statistical uncertainties of the single $1993-98$ data sample.

Two more aspects of the hadron test based on K_s^0 decays have been considered in this analysis.

The first has to deal with the difference in the distributions of tracks as a function of the opening angle between the track direction and the thrust axis, which is different for a class of hadrons from $K_s^0 \to \pi^+\pi^-$ decays, and for a class containing all hadrons in $b\bar{b}$ events. Although the values of the mis-ID probability are different for the two classes of track, the ratio (4.31) between the values for the data and the MC is expected to be steady.

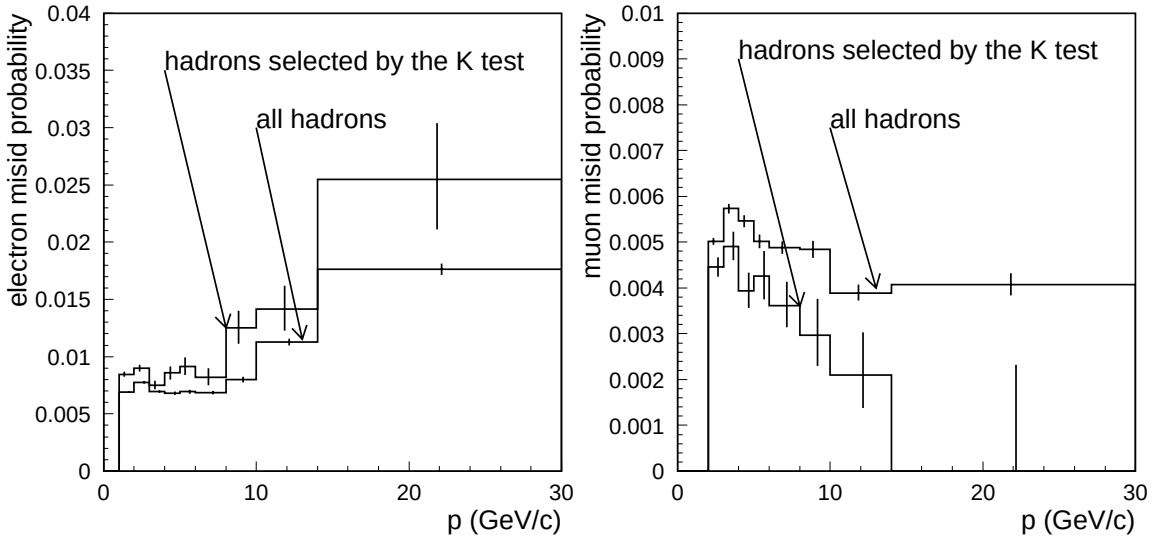


Figure 4.21: Left: comparison between $f(EH \mid \text{non-}e, K)$ and $f(EH \mid \text{non-}e)$. Right: comparison between $f(LW \mid \text{non-}\mu, K)$ and $f(LW \mid \text{non-}\mu)$.

The second aspect has to deal with the difference in the composition of hadrons within the two classes. The composition within the class containing all hadrons in $b\bar{b}$ events is roughly $\pi : K : p = 10 : 2 : 1$, while in the subset of candidates from K_s^0 decays it is more likely to be $100 : 1 : 1$. The true MC mis-ID probabilities for the two classes of tracks are compared in Fig. 4.21 for both electrons and muons.

4.8 Results of the efficiency calibrations

The efficiency calibration technique outlined in Sec. 4.4 has been specifically developed to handle the problem arising from our inability to obtain a pure reference sample of leptons in the entire momentum range. The problem is alleviated by using mutually independent tests, leading to equations like (4.19) in which

$$r = \frac{f(e \mid EH)}{f(e \mid H)} \frac{f(e \mid EH)}{f(e \mid E)} \equiv r_1 r_2$$

tend to be a relatively small number, Eq. (4.21). Since our calibration technique is based on the idea that as long as r is small the outcomes of the tests are reliable, the first thing to do is to somehow observe how the whole scheme works in practice. To do that, we repeat the calibration for the following classes of tracks:

- Tracks from events with at least one displaced vertex, with γ conversions excluded. About 70% of events with at least one displaced vertex are $b\bar{b}$ events. The remaining 30% are mostly $c\bar{c}$ events.
- Same as the class above but with the impact parameter cut > 1 mm applied, Fig. 4.22, to reduce the fraction of hadrons from the fragmentation processes. The test based on the impact parameter cut will be called the I test, and will be used in conjunction with the other tests, EI, HI, etc., to increase the purities of the reference samples.
- Tracks from γ conversions taken from all recorded events (this applies only to electrons).

The calibration results for the first two classes are highly correlated because the second class is the subset of the first. However the set of tracks from γ conversions is completely disjoint from the first two classes. The ratio of the calibrated to the uncalibrated MC efficiencies is then observed for the values of r corresponding to the three classes. If the ratio is flat, the assumptions (4.16) about the purity ratios is very likely to be correct. Here is how it works for electrons.

4.8.1 Electrons

Fig. 4.23 shows true MC efficiencies and purities for electrons. The efficiencies are calculated as conditional probabilities: $f(E \mid eH) = f(E \mid e)$ and $f(H \mid eE) = f(H \mid e)$, which hold when E and H are mutually independent. Therefore, bin-by-bin in momentum

$$\begin{aligned} f(EH \mid e) &= f(E \mid eH)\, f(H \mid eE) \\ &= \frac{f(e \mid EH)}{f(e \mid H)} \frac{f(e \mid EH)}{f(e \mid E)} f(E \mid H)\, f(H \mid E)\,, \end{aligned}$$

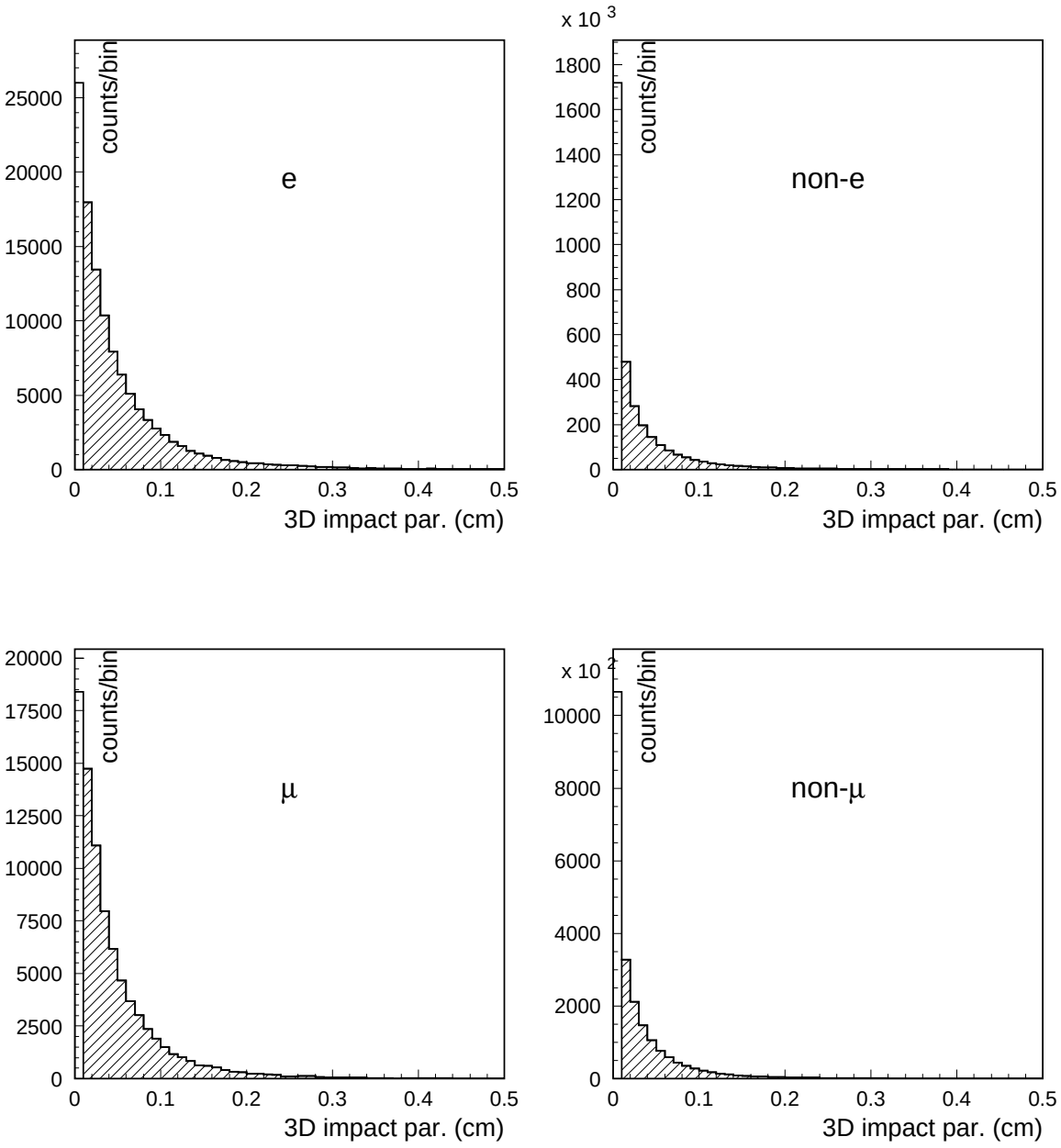


Figure 4.22: Top row: 3D impact parameter distributions for the generator level MC electrons and non-electrons with $p > 1$ GeV. Bottom row: the same for muons with $p > 2$ GeV.

$$\equiv \quad r\, f(E \mid H)\, f(H \mid E)\,. \tag{4.34}$$

Efficiency and purity of the combined EH test are shown in Fig. 4.24, together with $f(E \mid H)$ and $f(H \mid E)$ for both the MC and the real data.

The calibration results, Fig. 4.25, feature purity ratios r_1, r_2, and $r = r_1\, r_2$ in the top row. The key results are shown in the bottom two plots of the same figure. The plot on the left compares uncalibrated and calibrated efficiencies calculated using (4.20), with the large purity ratio $f(e \mid EH)/f(e)$. Plot on the right uses more robust procedure of Eq. (4.34). The difference between the central values of the calibrated to the uncalibrated efficiencies is about 10%. The fluctuations in the calibrated efficiencies in the left plot are also much larger than when our procedure is used. We confirm the correctness of our approach by redoing the calibration using purer reference samples based on impact parameter test I, and the electron test based on γ conversions.

The purities of the reference samples can be slightly improved by combining combining the E and H tests with the impact parameter test I, Fig. 4.22. The efficiencies of the E and the H tests are then

$$f(E \mid eHI) \equiv f(E \mid e) = \frac{f(e \mid EHI)}{f(e \mid HI)}\, f(E \mid HI)\,, \tag{4.35}$$

and

$$f(H \mid eEI) \equiv f(H \mid e) = \frac{f(e \mid EHI)}{f(e \mid EI)}\, f(H \mid EI)\,, \tag{4.36}$$

assuming E and HI, and H and EI mutually independent. The results are shown in Fig. 4.26 through Fig. 4.28. The ratio of the purities r in Fig. 4.28 is about 50% lower than that in Fig. 4.25. The fluctuations of the calibrated MC efficiencies, the lower right plot in Fig. 4.28 are somewhat lower compared to the fluctuations in Fig. 4.25. The ratio between the calibrated and uncalibrated MC efficiencies in Fig. 4.25 and Fig. 4.28 remains practically unchanged.

By far the most accurate way of calibrating the electron efficiency in the lower momentum bins (up to about 5 GeV) is by using electrons from γ conversions, which will be called the G test.

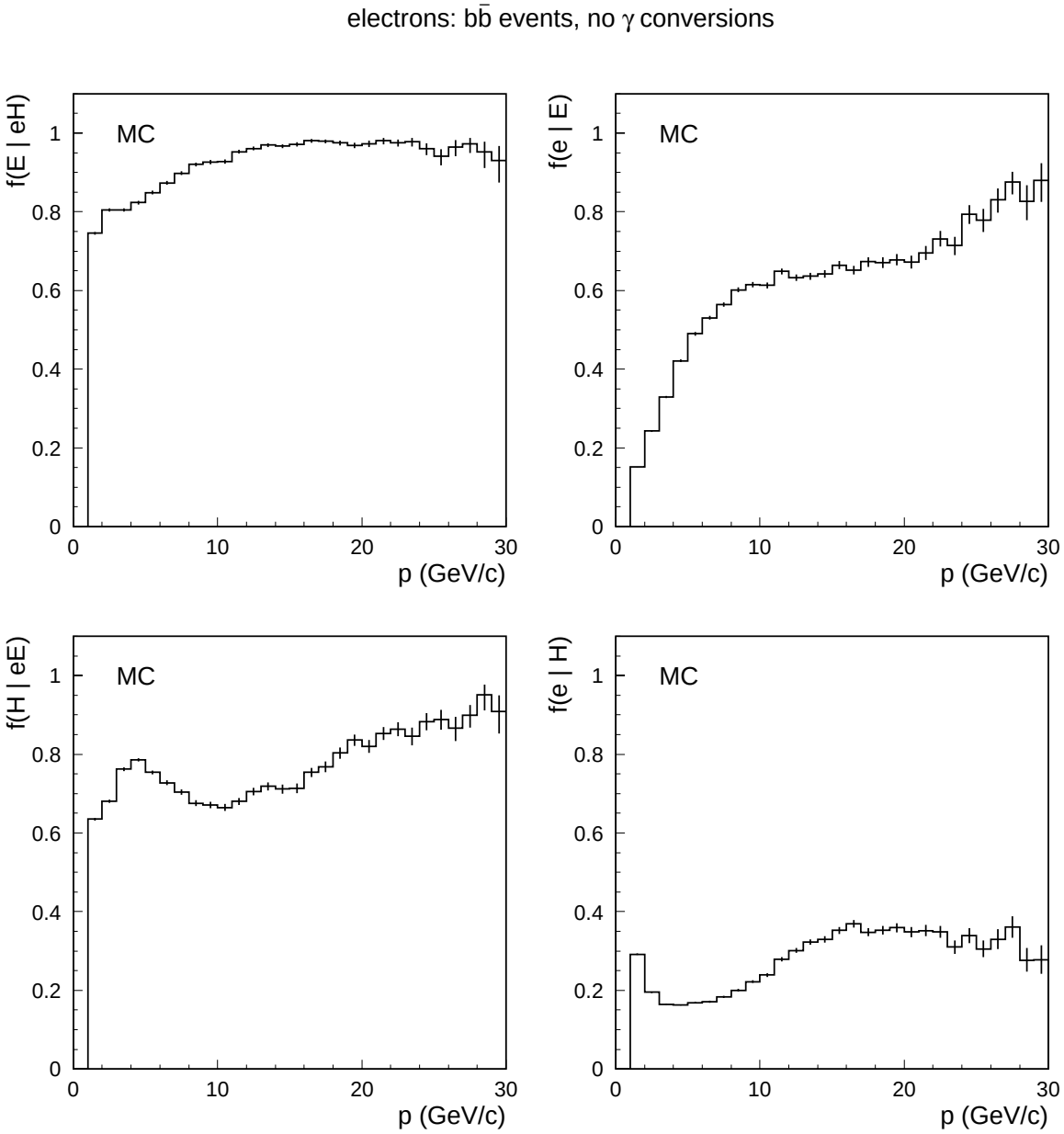


Figure 4.23: MC efficiencies (left column) and purities (right column) of the E electron test (top row) and the H electron test (bottom row).

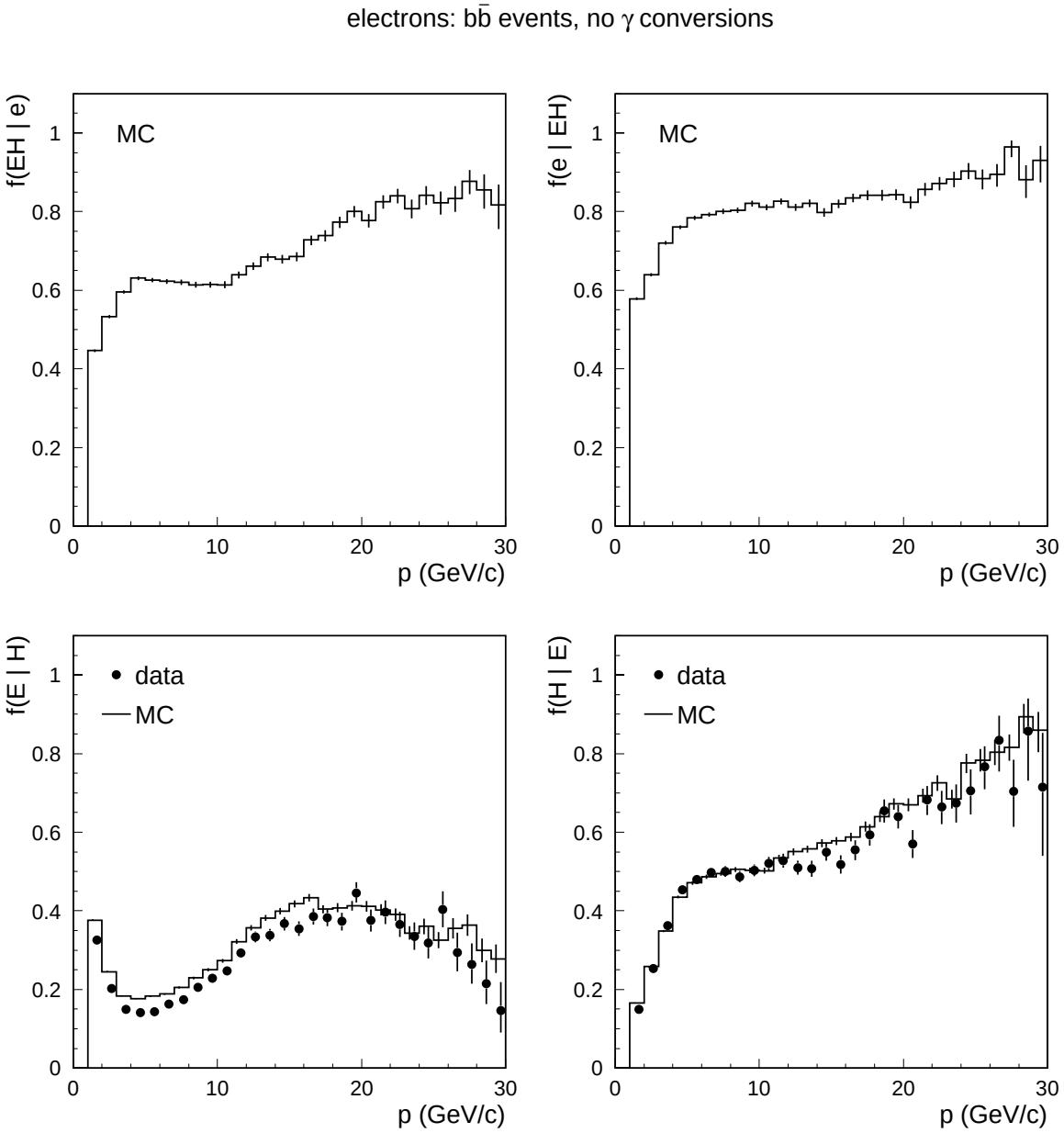


Figure 4.24: MC efficiency and purity of the combined electron EH test (top row), and mutual conditional probabilities between the E and the H tests (bottom row).

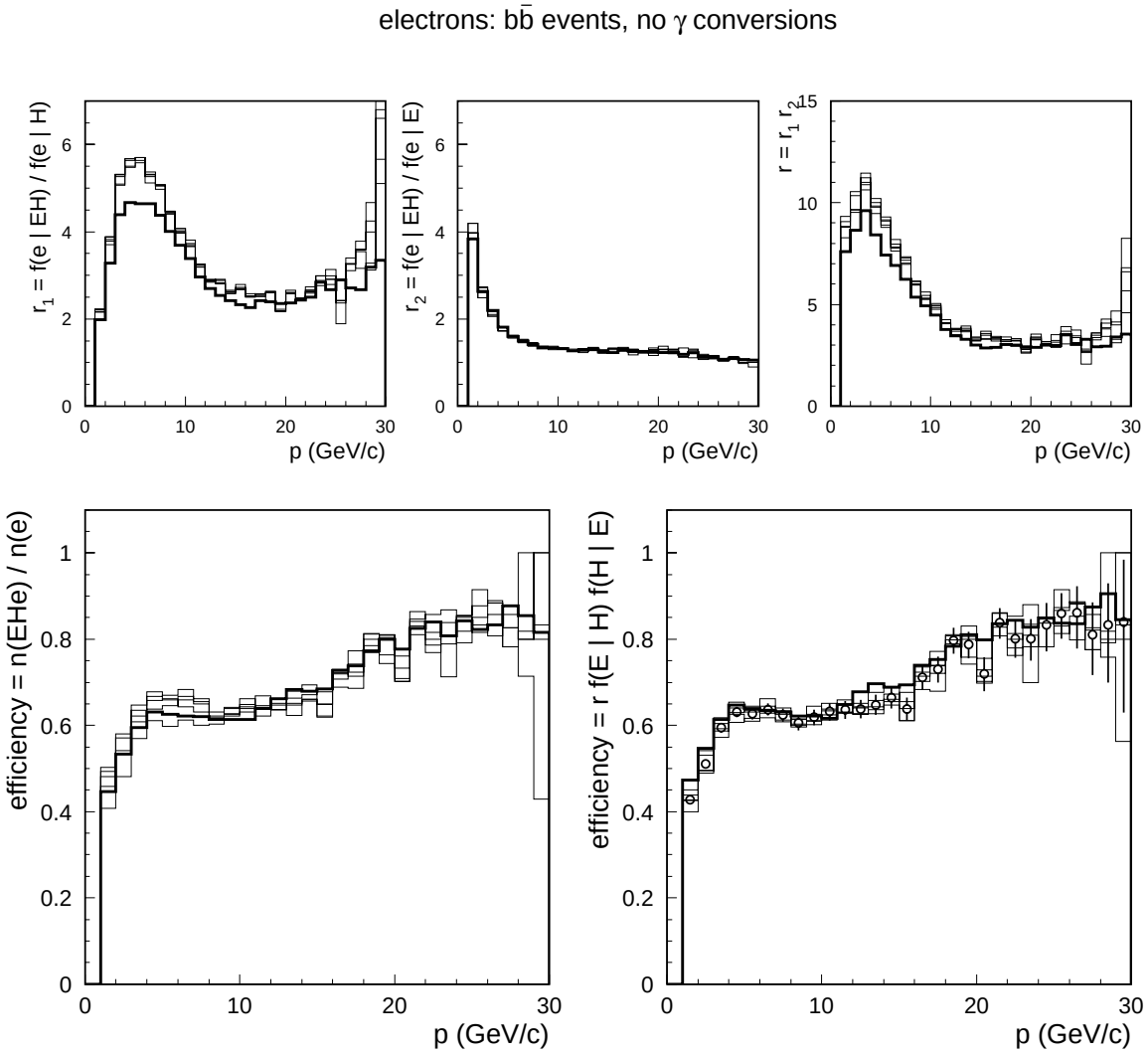


Figure 4.25: Top 3 plots show the two purity ratios $r_1 = f(e \mid EH)/f(e \mid H)$, $r_2 = f(e \mid EH)/f(e \mid E)$, and their product $r = r_1\, r_2$. The bottom left plot shows true MC efficiency (thick line) and calibrated efficiencies (collection of thin lines) using Eq. (4.20) with the $f(EH)$ calibrated to the data. The bottom right plot shows true MC efficiency (thick line) and calibrated efficiencies (collection of thin lines) using Eq. (4.19) instead, with the $f(E \mid H)\, f(H \mid E)$ calibrated to the data.

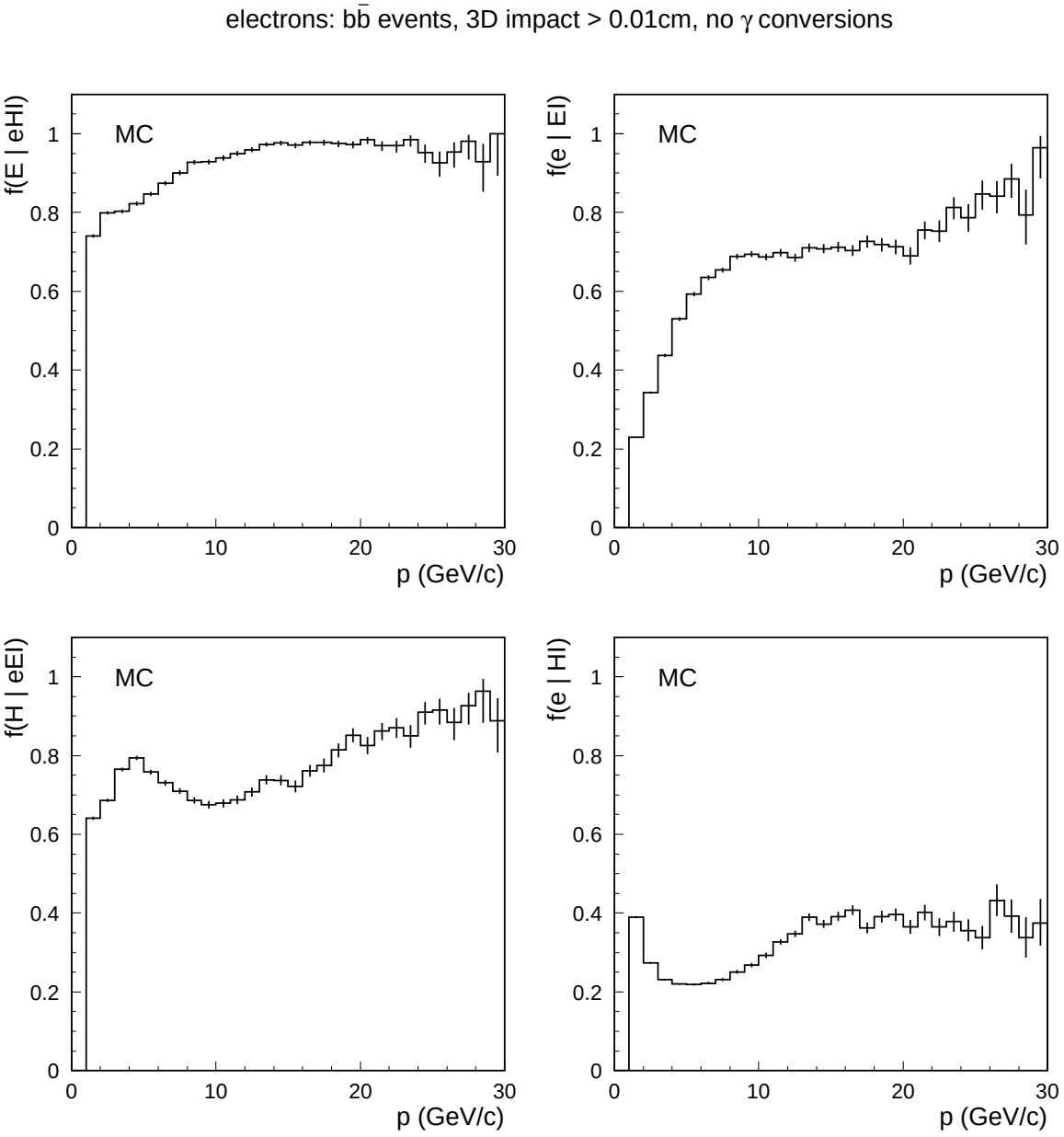


Figure 4.26: MC efficiencies of the electron E and H tests (left column) and their purities when they are combined with the impact parameter test I (right column). The efficiencies of the E and H remain practically unchanged when the H and E reference tests are replaced with the HI and EI tests.

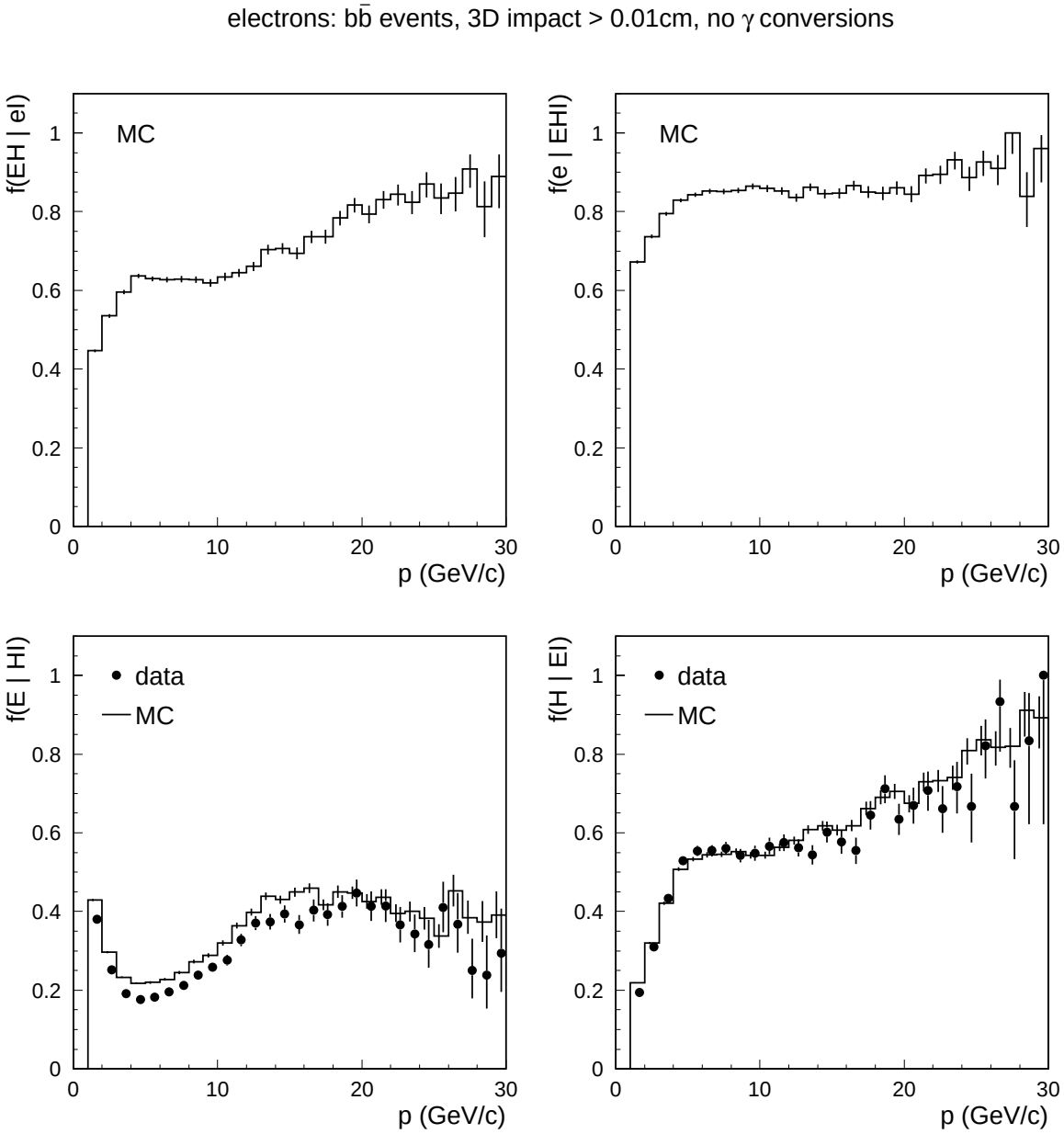


Figure 4.27: Efficiency of the EH test estimated using the reference set of electrons whose purity is increased by the impact parameter I tests (top left plot). Purity of the combined EHI test is shown in the top right plot. The yield of electrons selected by the E test, from the reference sample preselected by the HI test, is shown in the bottom left plot, and the yield of electrons selected by the H test, from the reference sample preselected by the EI test, at the bottom right plot.

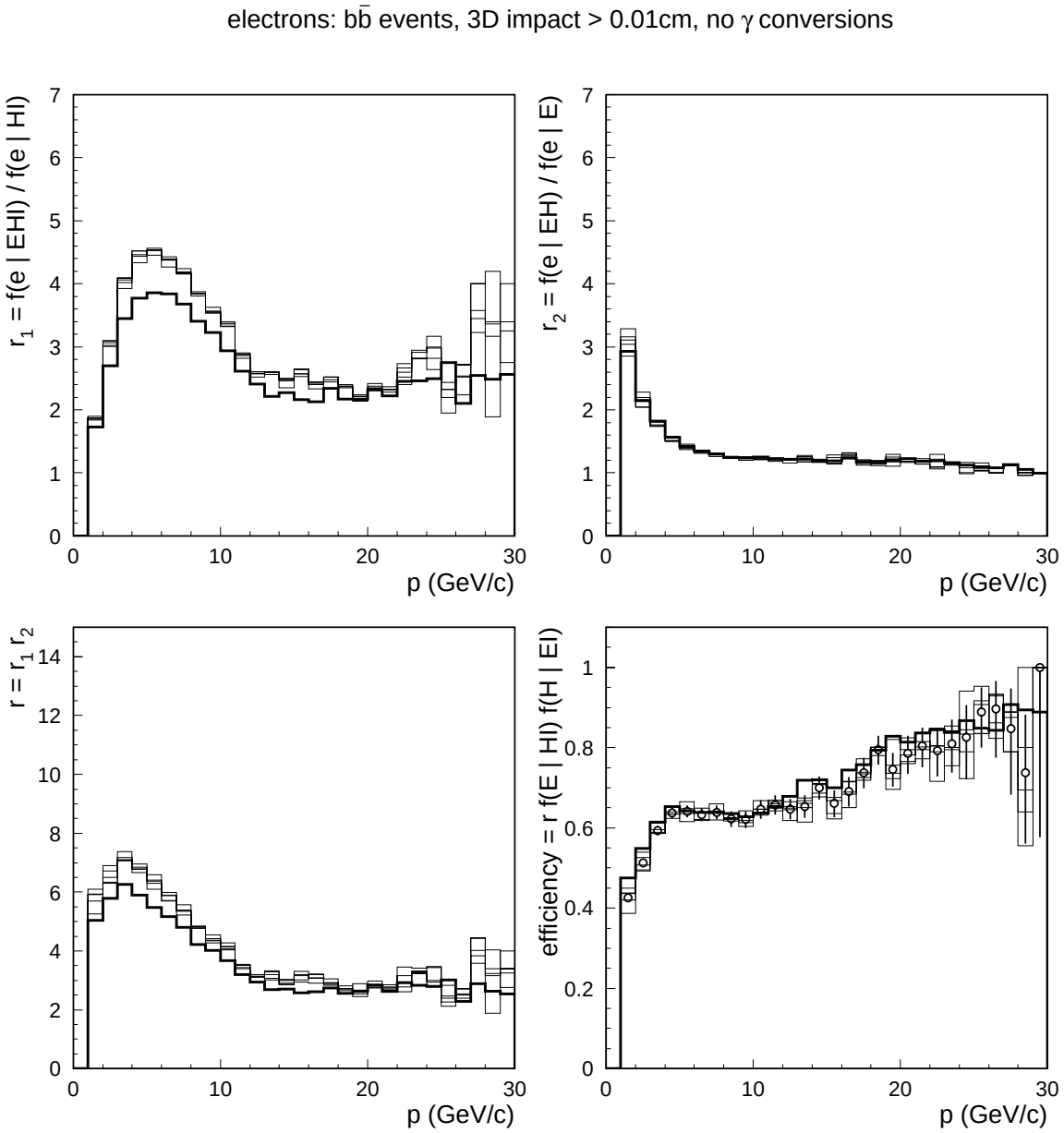


Figure 4.28: Same as in Fig. 4.25 but for a purer sample of electrons obtained by applying > 0.01 cm impact parameter cut from Fig. 4.22. The value of the scaling factor r at around 4 GeV is reduced compared to that in Fig. 4.25 (with no impact parameter cut) by about 50%.

The G test:

- $m_{ee} < 0.01$ GeV, where m_{ee} is the invariant mass of electron candidates from $\gamma \to e^+e^-$ conversion
- imp2 > 0.02 cm, cut on 2D impact parameter in the SLD xy plane, applied to both tracks in a pair
- $|\mathbf{p_1} + \mathbf{p_2}| < 20$ GeV/c, cut in the total momentum of the pairs of tracks
- $\rho = \sqrt{x^2 + y^2}$, for the γ conversion coordinates x and y, is required to fall in the following intervals: $2 - 6$ cm, $11 - 16$ cm, and $18 - 24$ cm. The intervals correspond to the regions of the higher density of the detector materials, where the conversions are more likely to occur.

[The above is "traditional" electron test using γ conversions. Large improvements can be obtained by using a likelihood ratio test similar to one described in sections 4.1 and 4.2. Distributions like $f(\mathbf{x} \mid \gamma)$ and $f(\mathbf{x} \mid \text{non-}\gamma)$, similar to those in Eq. 4.6 for example, can easily be obtained from the Monte Carlo. Some of the highly discriminating variables that can be used to span the sample space are: ρ, imp2, track momentum $p = \mid \mathbf{p} \mid$, difference between momenta of the two tracks with a common vertex $\Delta = \mid \mathbf{p} - \mathbf{q} \mid$, m_{ee}, and others.]

Pure reference samples of electrons using γ conversions are obtained from the regions of the higher material density in the detector, Fig. 4.29. The purity achieved is around 90%, lower right plot in Fig. 4.29.

The G test can be combined with the E and the H test to give

$$f(E \mid eHG) \equiv f(E \mid e) = \frac{f(e \mid EHG)}{f(e \mid HG)} f(E \mid HG)\,, \tag{4.37}$$

and

$$f(H \mid eEG) \equiv f(H \mid e) = \frac{f(e \mid EHG)}{f(e \mid EG)} f(H \mid EG)\,, \tag{4.38}$$

assuming E and HG, and H and EG mutually independent. The efficiency of the combined test is then $f(EH \mid e) = f(E \mid eHG)\, f(H \mid eEG)$.

The MC efficiencies, purities, etc., Fig. 4.30 through Fig. 4.32, correspond to Fig. 4.26 through Fig. 4.28, when the impact parameter test is used instead of the G

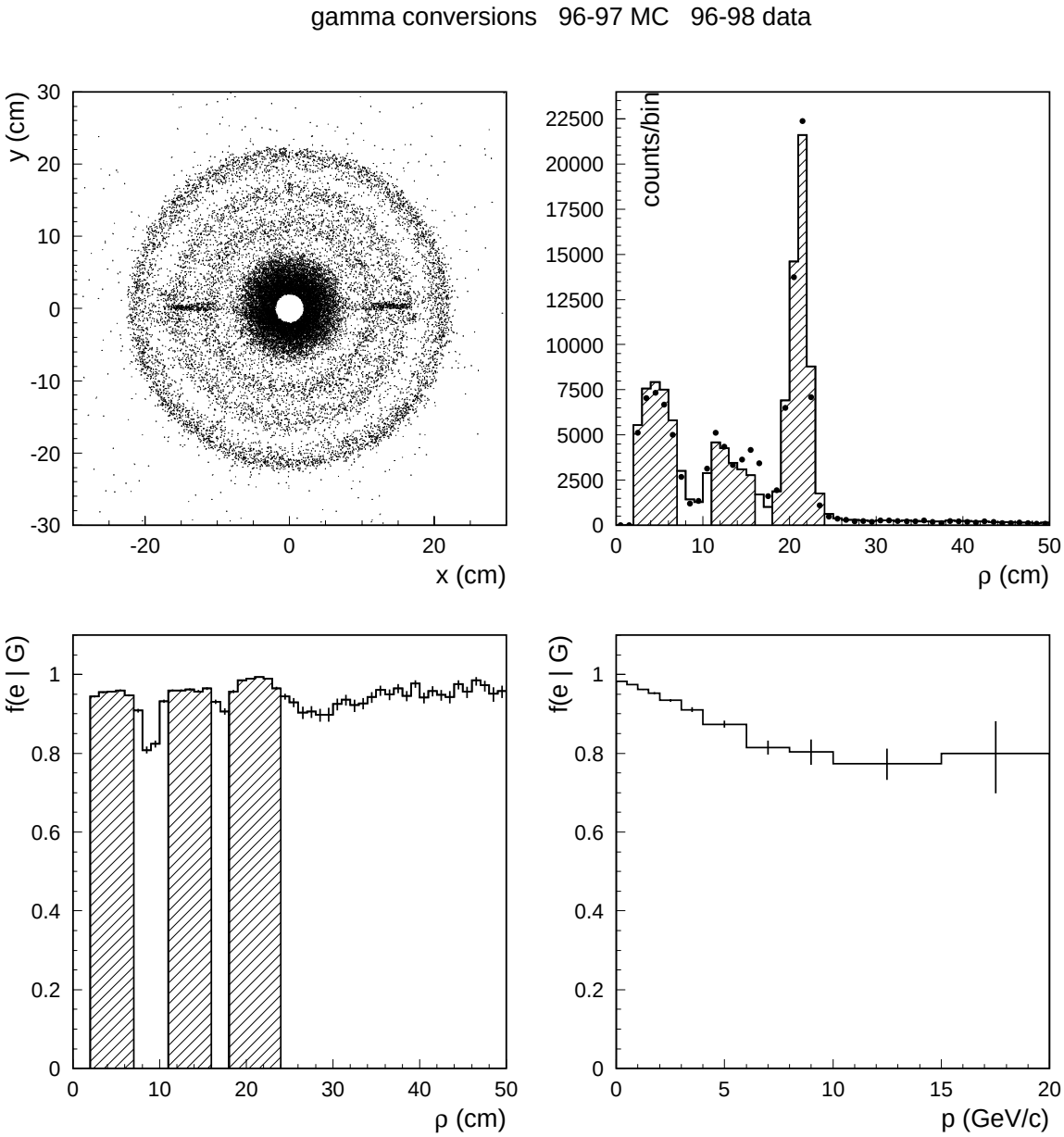


Figure 4.29: Top left: An "X ray" image of a 6 cm thick slab of the detector around the IP obtained by reconstructing vertices of electron candidates from $\gamma \to e^+e^-$ conversions in the detector materials. The outermost circle is image of the inner CDC wall; the inner circles are images of the cryostat walls and the VXD3. Top right: Distribution of the conversion vertices in $\rho = \sqrt{x^2 + y^2}$. Bottom left: Efficiency in ρ. Vertices from the hatched areas are used as a high purity source of γ conversion electrons, the purity of which is shown in the bottom right plot as a function of momentum.

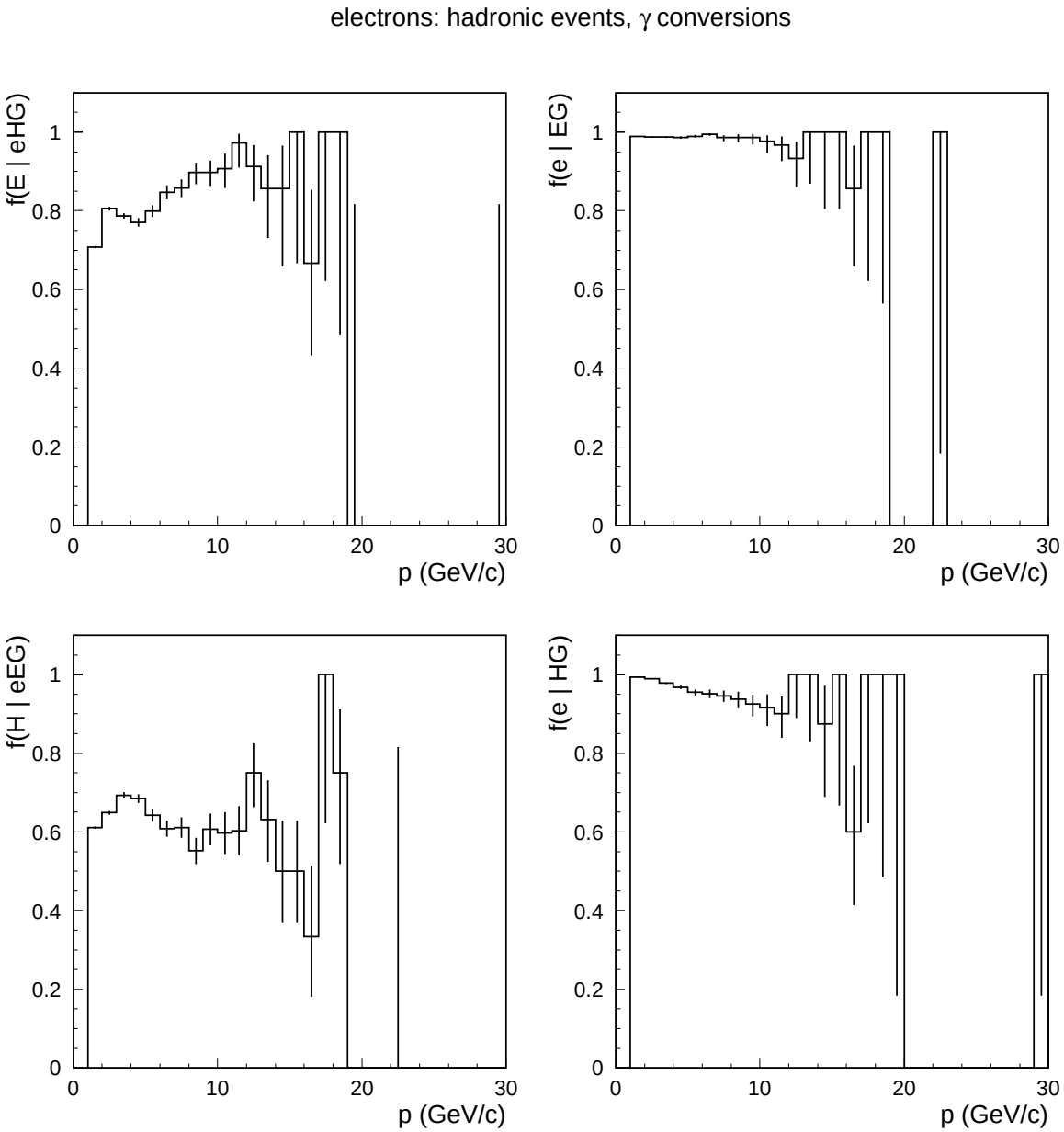


Figure 4.30: Same as in Fig. 4.26 but with the fraction of electrons in the reference samples increased by the γ conversion test G from Fig. 4.29.

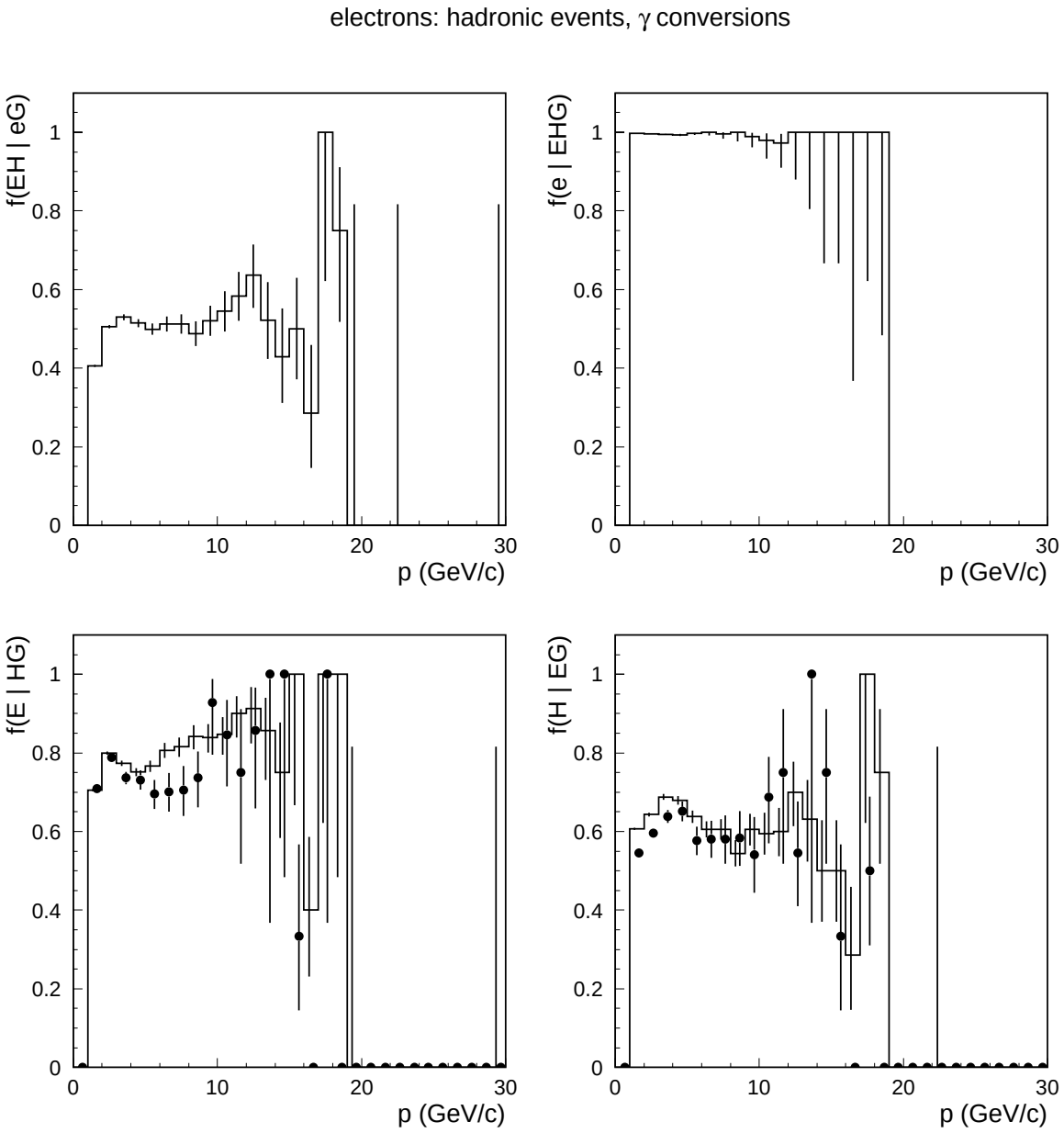


Figure 4.31: Same as in Fig. 4.27 but with the fraction of electrons in the reference samples increased by the γ conversion test G from Fig. 4.29. The true MC efficiency (top left plot) is different from that in Fig. 4.23 or Fig. 4.26 due to the substantially different distribution of electrons from γ conversions with respect to the jet topologies.

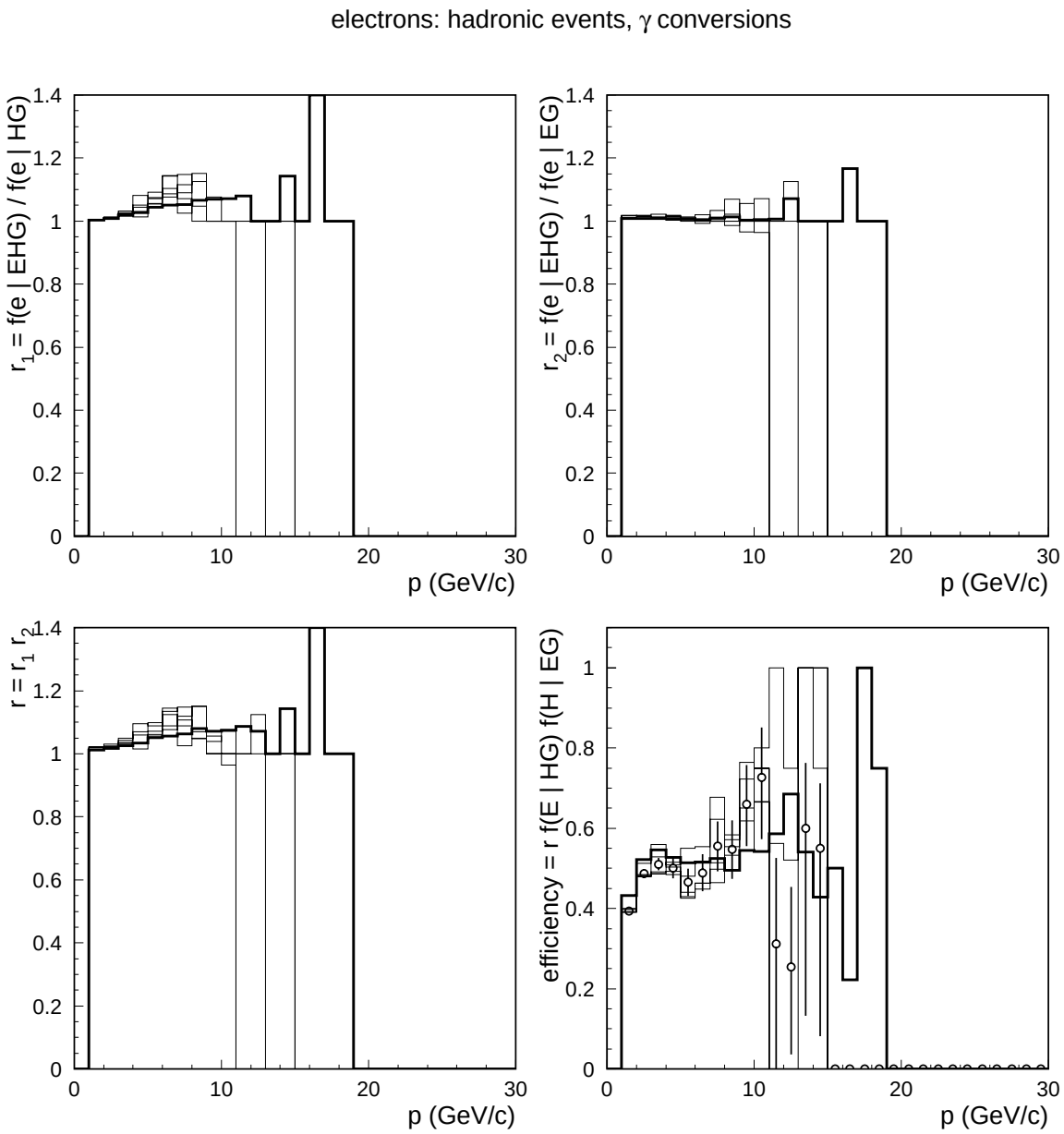


Figure 4.32: Same as in Fig. 4.25 but for a very pure reference sample of electrons from γ conversions from Fig. 4.29. The $f(E \mid HG)\, f(H \mid EG)$ is scaled by a factor $r \approx 1$.

test. As is the case with the previous two studies, the ratio of the efficiencies for the MC and the real data is practically unchanged (compare lower right plots in Fig. 4.32, Fig. 4.28, and Fig. 4.25).

This is a very good news: if one thinks of ratio of the two purities r in Eq. (4.34) as a controlling parameter of the calibration, which one tries to make as close to 1 as possible, then it turns out that the ratio of the calibrated to the uncalibrated efficiency is almost flat as r changes from ≈ 7 in Fig. 4.23 to ≈ 1.05 in Fig. 4.30.

The final result for the electron efficiency is obtained by averaging the calibration results from the $b\bar{b}$ events with 3D impact parameter cut, Fig. 4.28, with the γ conversion results, Fig. 4.32. Since the absolute efficiencies are different in the two classes, one first scales the γ conversion results as

$$\frac{\mathrm{MC}(b\bar{b}\text{ evts with 3D impact par cut})}{\mathrm{MC}(\gamma\text{ conversion candidates})} \times \mathrm{data}(\gamma\text{ conversion candidates})\,.$$

The ratio of the MC values in this expression varies roughly between 1.07 at 1 GeV and 1.25 at around 8 GeV (lower right plots in Fig. 4.31 and Fig. 4.27). Assuming the distributions in each bin Gaussian, and by taking the average of two independent Gaussian random variables x and y with variances σ_x^2 and σ_y^2 to be

$$z = \frac{\sigma_y^2}{\sigma_x^2+\sigma_y^2}\,x \,+\, \frac{\sigma_x^2}{\sigma_x^2+\sigma_y^2}\,y\,,$$

one has

$$\frac{1}{\sigma_z^2} \,=\, \frac{1}{\sigma_x^2} \,+\, \frac{1}{\sigma_y^2}\,. \tag{4.39}$$

When everything is put together, the result looks like in Fig. 4.33. Filled circles represent the average calibrated values, and the associated error bars combine all the errors: statistical and systematic, plus the combination of the errors as in (4.39). The histogram is fitted using 8-th degree Chebyshev polynomials. The variance of the fit is calculated from the variance matrix of the polynomial parameters as usual [71, 7].

4.8.2 Muons

The muon MC count is calibrated to the data in exactly the same manner as the electron count. The momentum range is $2-30$ GeV, and the mis-ID probability

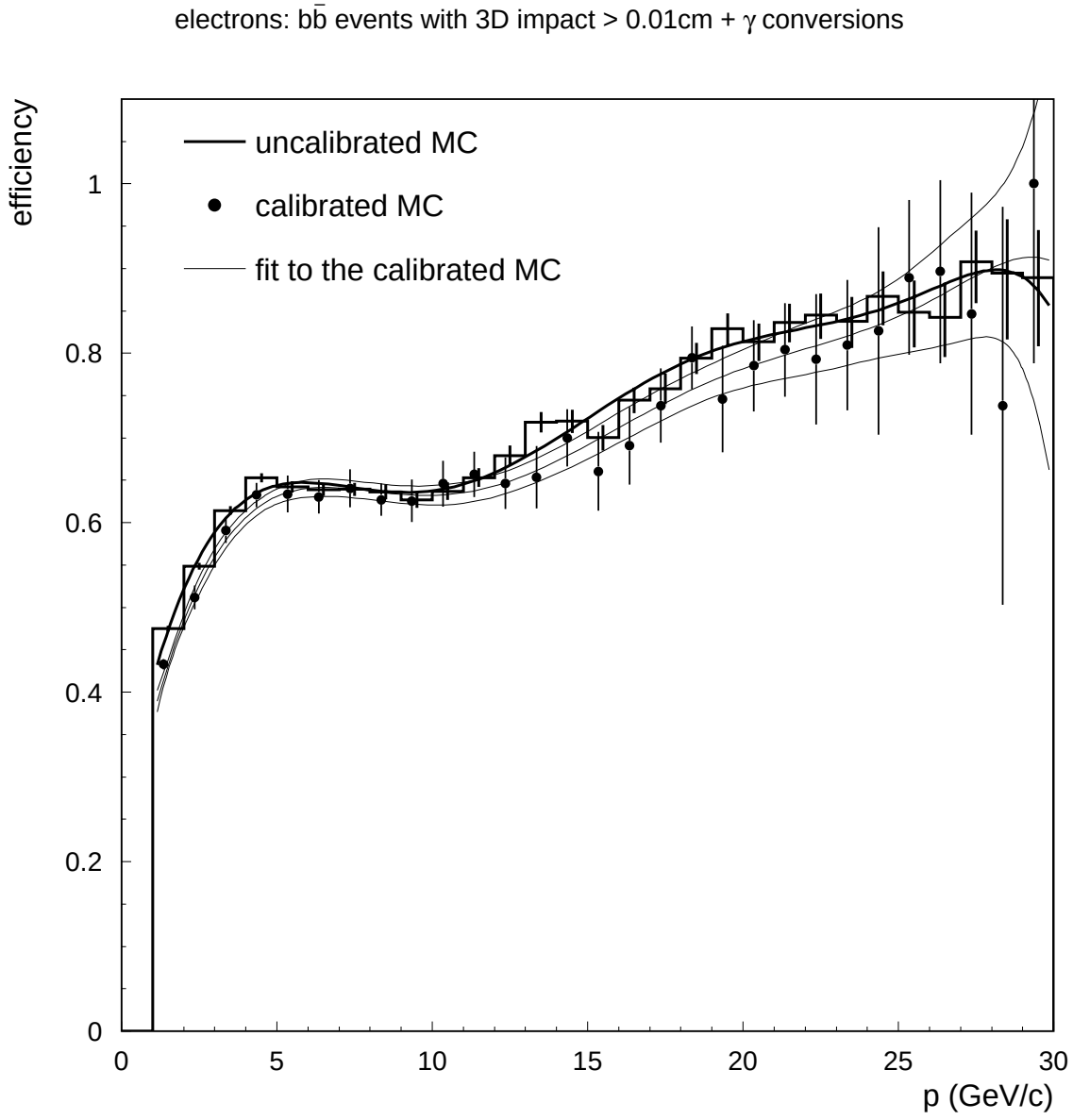


Figure 4.33: Calibrated electron efficiency by combining results for $b\bar{b}$ events with 3D impact parameter cut, Fig. 4.28, with results from γ conversions, Fig. 4.32.

somewhat lower than for electrons. Since no pure sources of muons are available in hadronic events, one can only use reference samples preselected by one of the two muon tests (L or W). As with electrons, the purities of the reference samples can be additionally enhanced by using the impact parameter test.

The calibration results, following precisely the same route as for electrons, are shown in Fig. 4.34 through Fig. 4.36, for the reference samples obtained using only the L and the W tests, and Fig. 4.37 through Fig. 4.39, when the purities of the reference samples are enhanced by the impact parameter I test.

Since our calibration procedure is based on very general assumptions about the nature of the purity ratios like $f(e \mid EH)/f(e \mid E)$ and the others, it is important to have an independent way of cross-checking the results. The only universal prescription on how to do that is to try to find as many disjoint reference samples as possible, and with the highest possible purity. In this analysis, a line of reasoning is to connect the efficiency and the mis-ID probability as described in Sec. 4.4, by means of which one can calculate the purity ratios using outcomes of the mis-ID probability calibrations, which use very pure samples of hadrons from K_s^0 decays, and are almost completely disjoint from the lepton samples.

The final result for muons is shown Fig. 4.36. Both the electrons and the muons estimators for the efficiencies as functions of p are obtained by minimizing the χ^2 of the 8-th degree Chebyshev polynomial fit (Chebyshev polynomials are used for convenience). The errors around the estimators are obtained from the covariance matrices as usual.

4.9 Supplement: Polynomial fit as an estimator

The estimators of the momentum dependent efficiencies $\eta = \eta(p)$ and mis-ID probabilities $\varepsilon = \varepsilon(p)$ are obtained by fitting the histograms with a function that is a linear combination of the first n Chebyshev polynomials. The "true" value of the mis-ID probability (for example) is then

$$\varepsilon(p) = a_0 T_0 + a_1 T_1 + \cdots + a_n T_n \ + \text{remainder}\,, \tag{4.40}$$

where $T_0, \cdots, T_n$ are the first n Chebyshev polynomials and the $a_0, \cdots, a_n$ are obtained from the least square fitting of the right hand side in (4.40) to the data. (Chebyshev polynomials are convenient for their property of being bounded between -1 and 1 [72], which is what makes them robust against roundoff errors in random summations up to a degree of around 50).

The argument in favor of the polynomial fit as an estimator is functional rather than statistical [73]. For arbitrary but fixed $p = p_0$ one can get the $\varepsilon(p_0)$ directly from

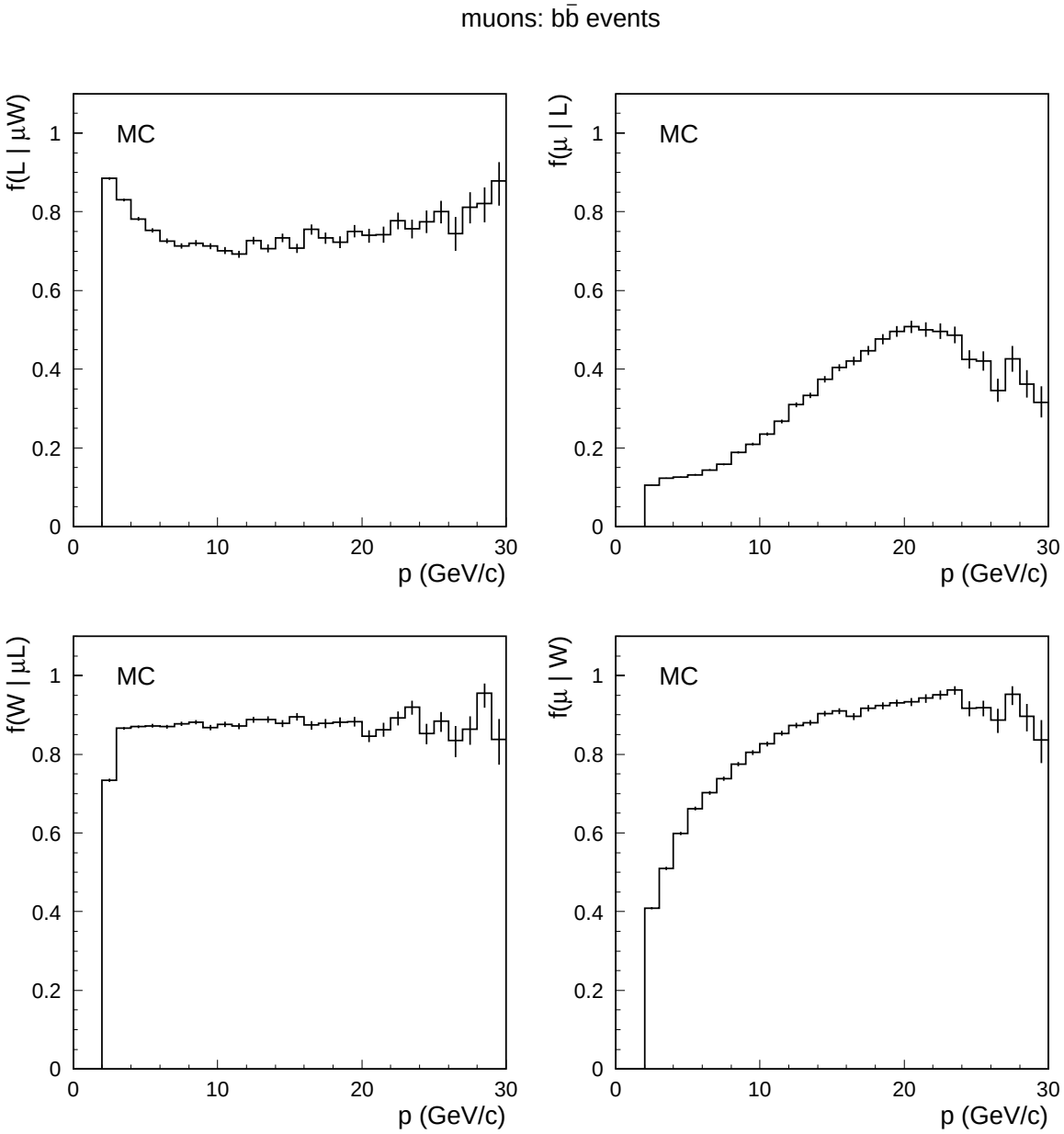


Figure 4.34: MC efficiencies (left column) and purities (right column) of the muon L test (top row) and the muon W test (bottom row).

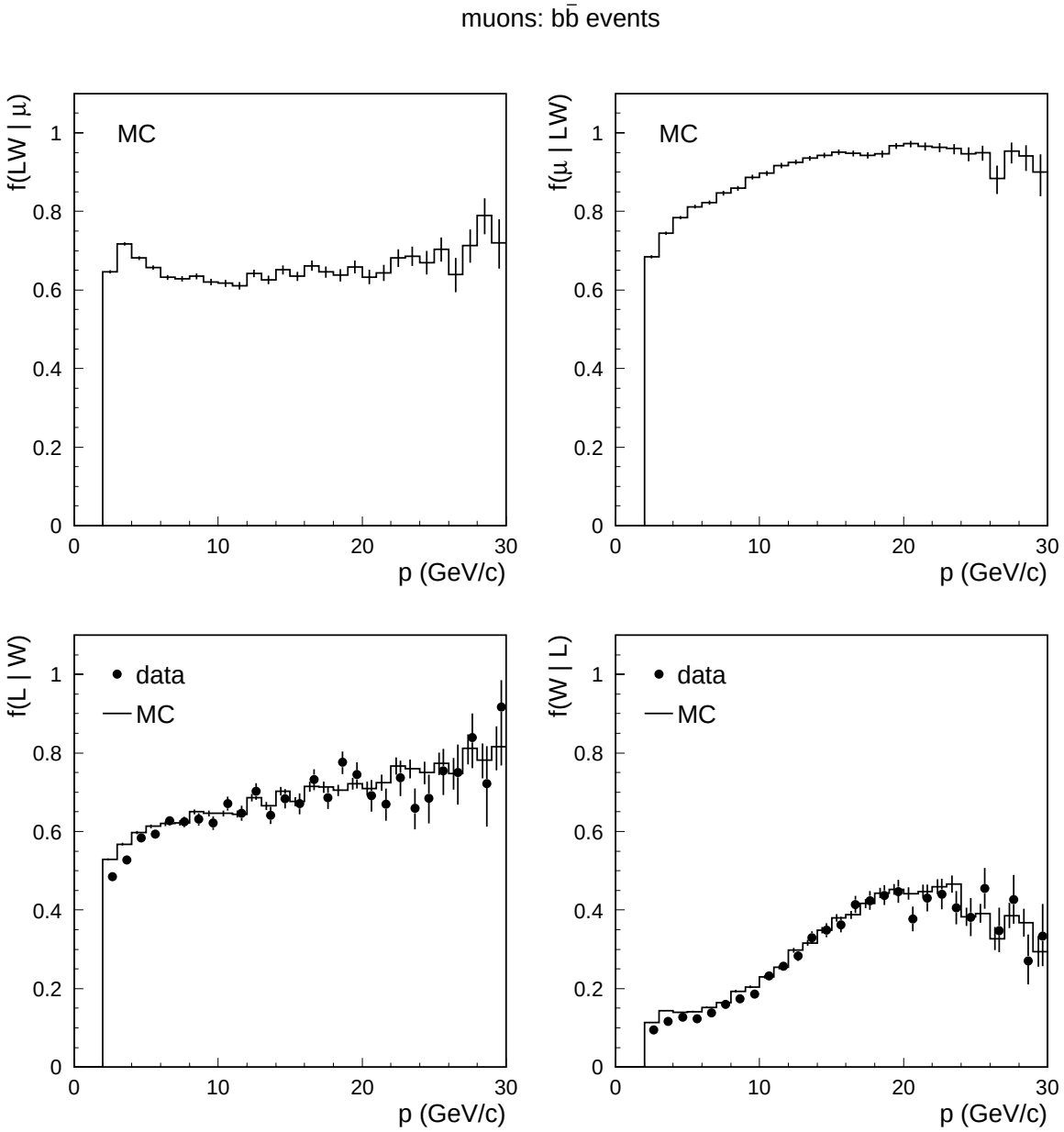


Figure 4.35: MC efficiency and purity of the combined muon LW test (top row), and mutual conditional probabilities of the L and the W tests (bottom row).

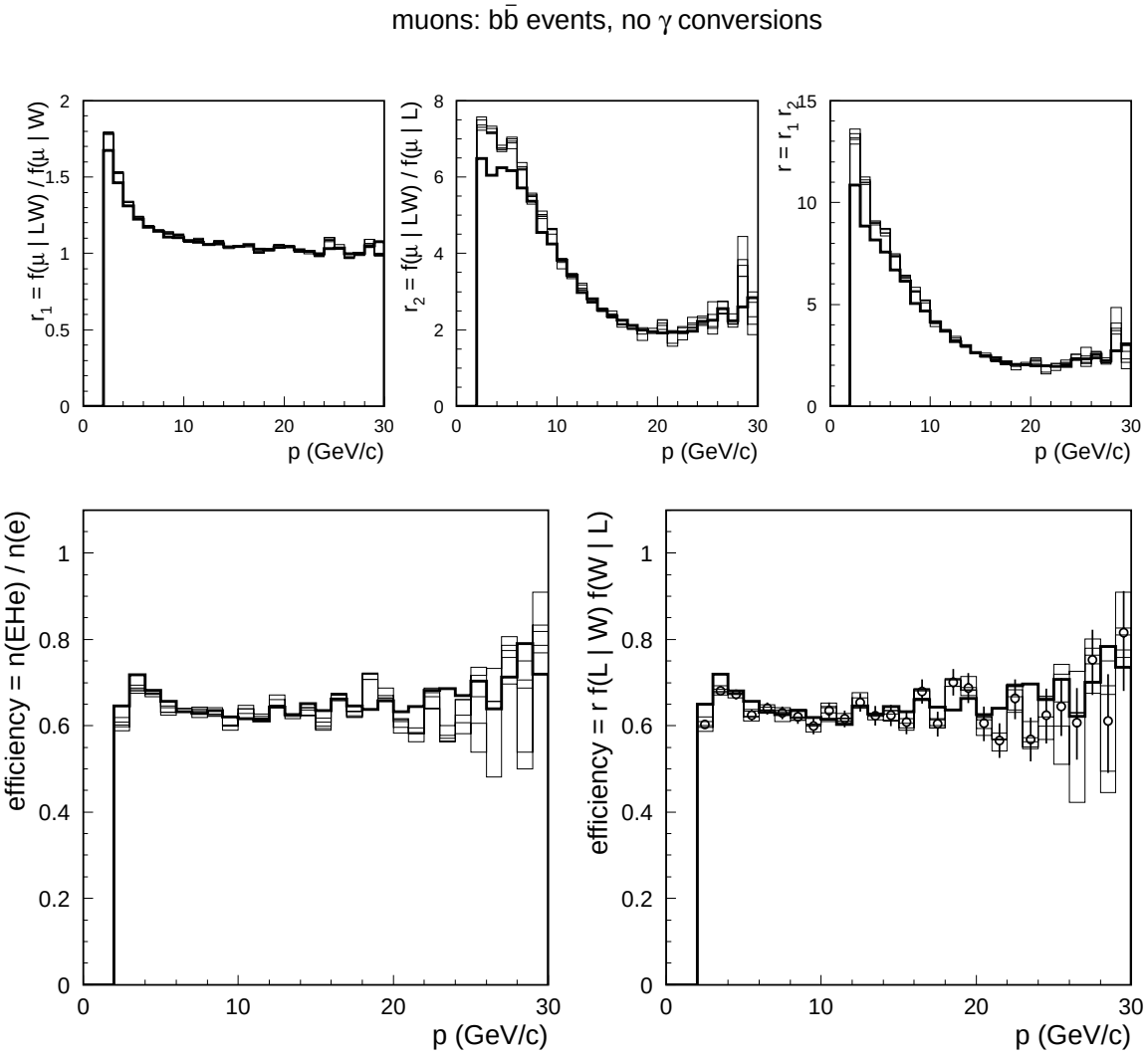


Figure 4.36: Top 3 plots show the two "scaling factors" $r_1 = f(\mu \mid LW)/f(\mu \mid W)$, $r_2 = f(\mu \mid LW)/f(\mu \mid L)$, and their product $r = r_1\, r_2$. The bottom left plot shows true MC efficiency (thick line) and calibrated efficiencies (collection of thin lines) obtained using Eq. (4.20) with the $f(LW)$ calibrated to the data. The bottom right plot shows true MC efficiency (thick line) and calibrated efficiencies (collection of thin lines) obtained using Eq. (4.19) instead, with the $f(L \mid W)\, f(W \mid L)$ calibrated to the data.

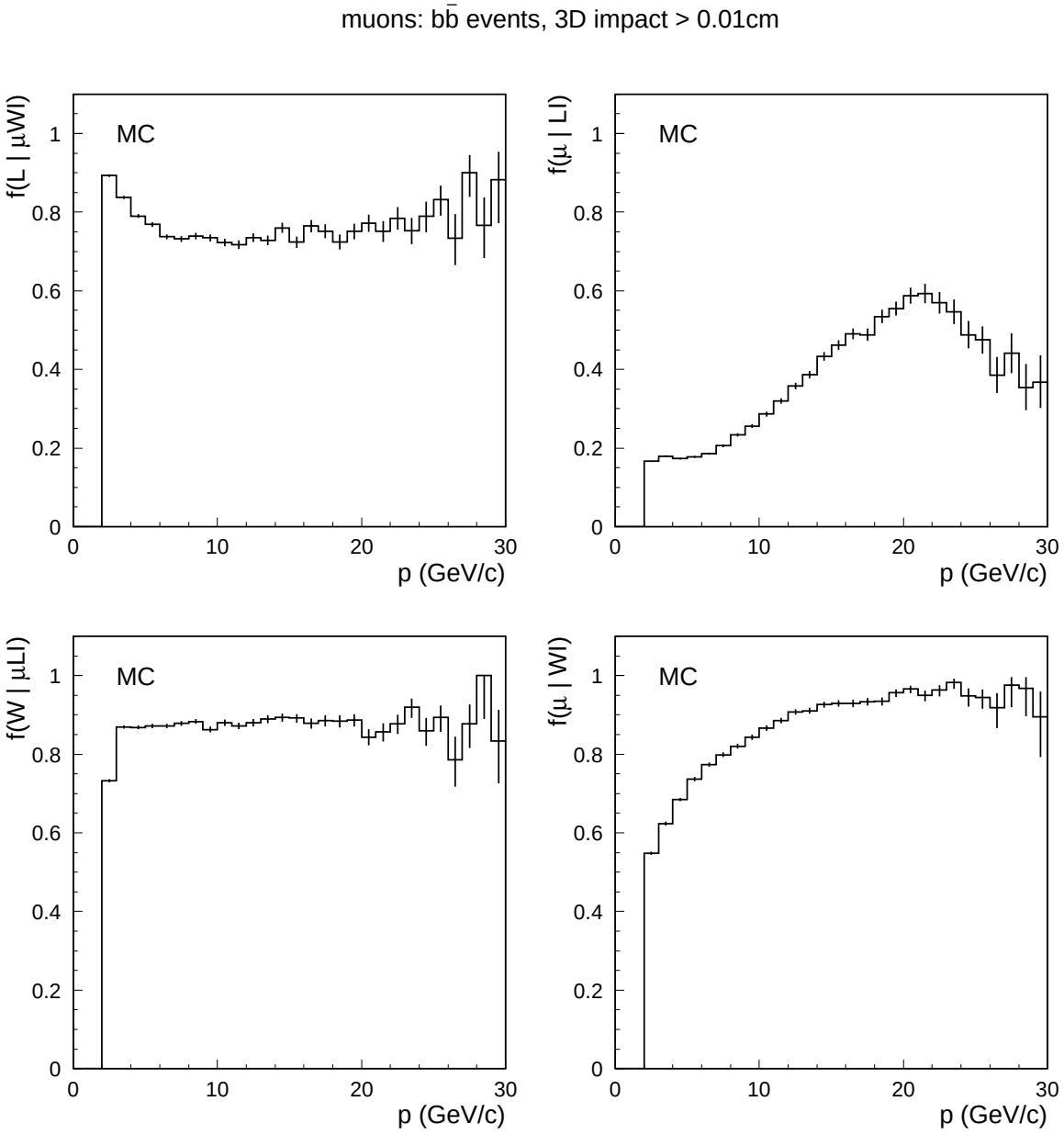


Figure 4.37: True MC efficiencies of the muon L and W tests (left column), and their purities when they are combined with the impact parameter test I (right column). The efficiencies of the L and W remain practically unchanged when the W and L reference tests are replaced with the WI and LI tests.

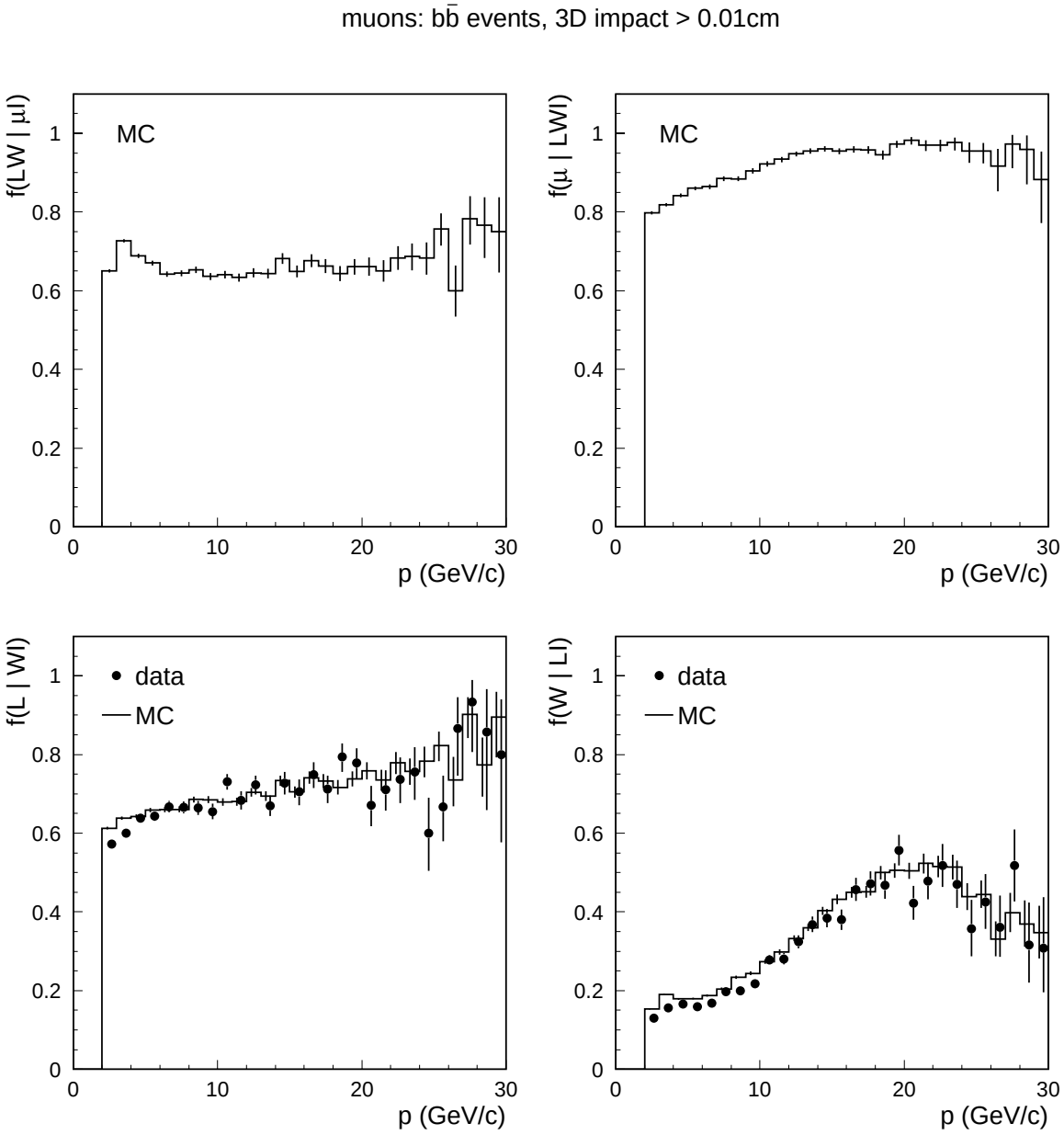


Figure 4.38: True MC efficiency of the combined muon LW test (top left plot), estimated using reference sample of muons whose purity is enhanced by the impact parameter I test. Purity of the combined LWI test (top right plot). Fraction of muons labeled by the L test within the reference sample of muons selected by the WI test (bottom left plot), and when the L and the W are switched (bottom right plot).

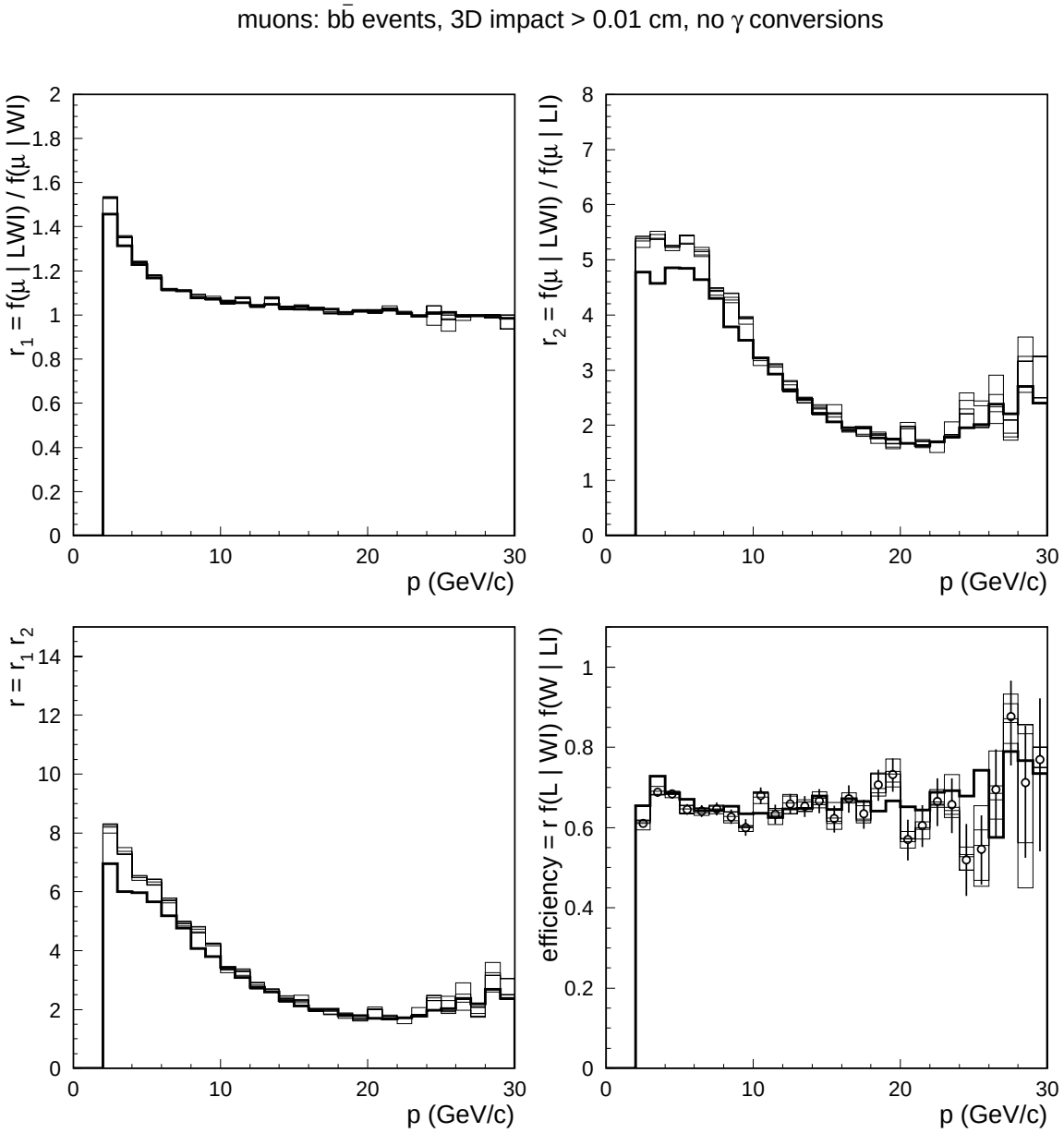


Figure 4.39: The calibration plots for muons when the purity of the reference sample is increased using the impact parameter I test.

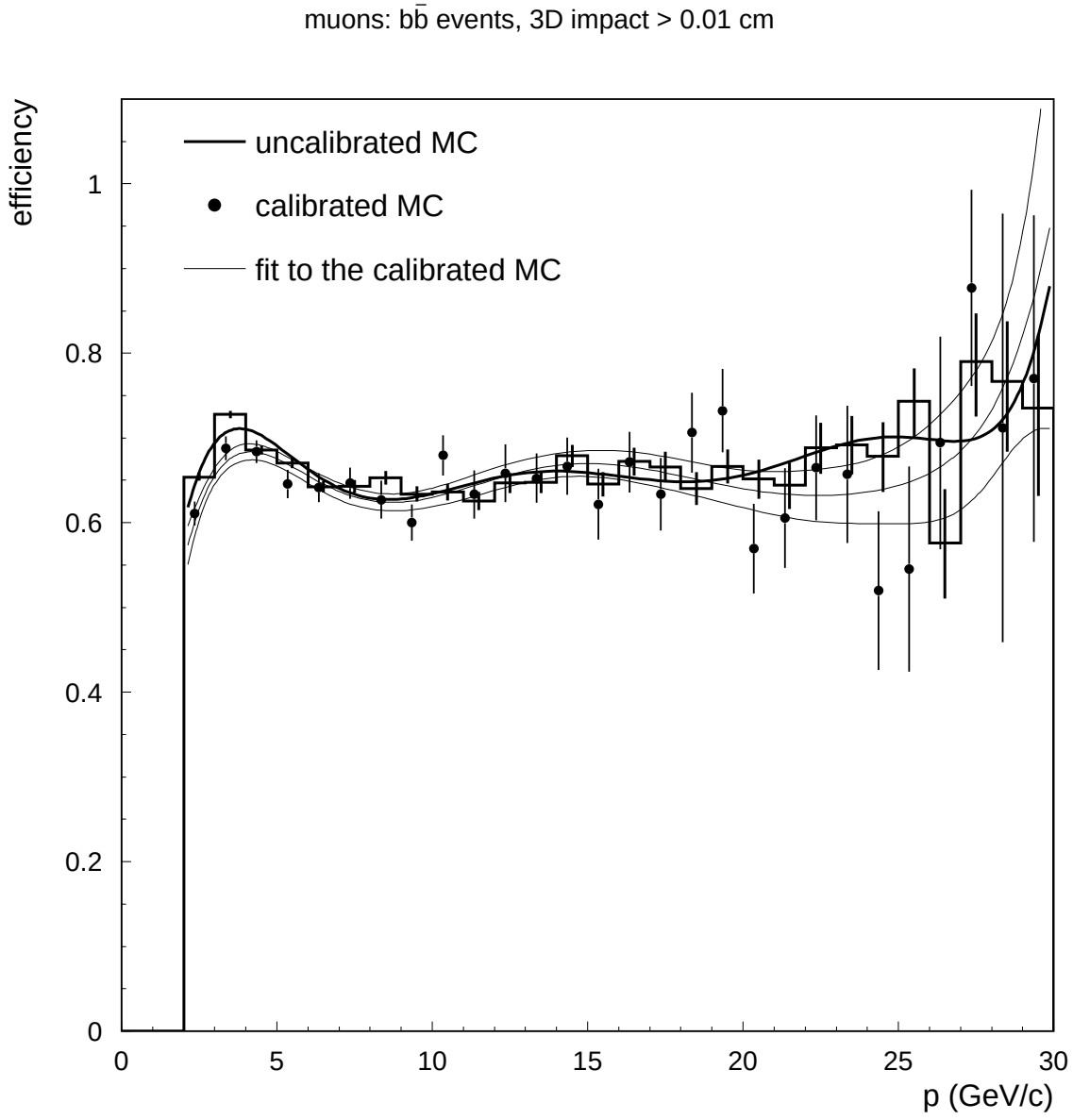


Figure 4.40: Calibrated muon efficiency by using $b\bar{b}$ events with 3D impact parameter cut, Fig. 4.39.

the number of counts falling within $p_0 \pm \epsilon$. For small ϵ such estimator is unbiased but has a large error due to the smallness of the number of counts. For large ϵ the error is small but the bias is large due to the variations of $\varepsilon(p)$ within $p_0 \pm \epsilon$. The role of the expansion of the (unknown) $\varepsilon(p)$ around p_0,

$$\varepsilon(p) = a_0 + a_1(p - p_0) + \cdots + a_n(p - p_0)^n + \text{remainder}\,,$$

where the $a_0, \cdots, a_n$ are obtained by minimizing the χ^2, is to simultaneously ensure both the low bias and the low variance of the estimator [73].

5
Initial state tag

Probability relations from Ch. 3 contain terms that have been identified as the initial state probabilities $f(b \mid \mathrm{B})$ and $f(\bar{b} \mid \mathrm{B})$. They are expressed as conditional probabilities that the initial state flavor in a hemisphere is b or $\bar{b}$ given the collection of variables

$$\mathrm{B} = (\mathrm{mvtx}, Q, P_e, \cos\theta)\,,$$

where mvtx is invariant mass in the opposite hemisphere, Q jet charge in the opposite hemisphere, P_e polarization of incident electrons, and $\cos\theta$ cosine of the thrust axis polar angle. The invariant mass and the jet charge have to be calculated in the *opposite hemisphere* to avoid large correlations to the lepton yield. In other words,

$$f(b\bar{b}, b, b{\to}l \mid \mathrm{mvtx}, Q, P_e, \cos\theta) \neq f(b{\to}l \mid b)\, f(b \mid Q, P_e, \cos\theta)\, f(b\bar{b} \mid \mathrm{mvtx})$$

when mvtx and Q are in the same hemisphere in which the leptons are counted. Variables in the polarized $e^+e^-{\to}b\overline{b}$ scattering, P_e and $\cos\theta$, are hemisphere unrelated, and are largely uncorrelated to the lepton yield.

Regarding the initial state tag, SLD has some advantages over other similar detectors: precision SLD vertex detector allows for more accurate determination of the decay vertices of the short-lived particles, and therefore for more efficient invariant mass tag, while the polarized incident electrons enhance the asymmetry in the Z coupling to $b\bar{b}$, and thus additionally improve the separation between the b and the $\bar{b}$.

5.1 $b\bar{b}$ event selection

This is taken from other SLD analyses *as is* (routines `BZVTMAS` and `BZMASS` written by Eric Weiss). The probability that an event is $b\bar{b}$ is determined from the p_T corrected

invariant mass of displaced vertices of the b-hadron decay candidates, as described in detail in a number of other SLD theses (see for example [74]). The vertices are reconstructed from the 3D probability distribution functions of the joint spatial track positions $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_n$ for $n \geq 2$ [75].

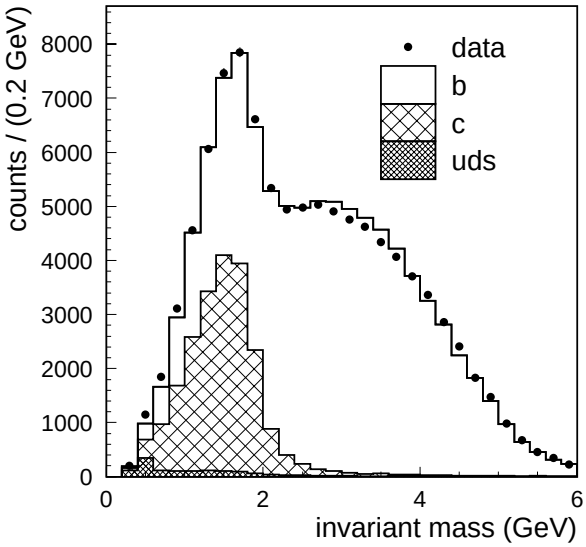


Figure 5.1: Invariant mass distributions for Monte Carlo b, c, and uds quarks and for the data.

Fig. 5.1 compares the distribution of the invariant mass for the data and the MC. From the MC composition of the b, c, and uds events we obtain $f(b\bar{b}) = 0.964$ when the invariant mass cut > 2 GeV is applied. This is an *uncalibrated* value of the $f(b\bar{b})$.

5.2 Flavor tag from polarized asymmetry

The goal is to derive the probability $f(b \mid \cos\theta, P_e) \equiv f(b \mid A)$ that a quark is b given the polarization of the incident electrons P_e and $\cos\theta$, where θ is the opening angle between the directions of the incident e^- and the outgoing b quark. In exactly the same way $f(\bar{b} \mid \cos\theta, P_e) \equiv f(\bar{b} \mid A)$, where θ is again opening angle between directions of the incident e^- and the outgoing b, not $\bar{b}$.

One can start with the formula for the unpolarized cross section for the process $e^+e^- \to Z \to b\bar{b}$

$$\begin{aligned} \frac{d\sigma}{d\cos\theta} &= \frac{\pi\alpha^2}{8\sin^4\theta_W \cos^4\theta_W \Gamma_Z^2} \left(C_L^{b\,2} + C_R^{b\,2}\right) \\ &\quad \times \left[C_L^{e\,2}(1+\cos^2\theta + 2A_b\cos\theta) + C_R^{e\,2}(1+\cos^2\theta - 2A_b\cos\theta)\right] , \end{aligned}$$

where the C_L^e (C_R^e) term corresponds to purely left (right) handed electron [14] p. 508, or calculate each of the two terms separately by using fermion spin projector $(1 + \gamma^5 \not{s})/2$ in trace formulas with $s^0 = \gamma\beta$ and $\mathbf{s} = \pm\gamma\mathbf{p}/|\mathbf{p}|$ for left and right handed incident electrons. One way or the other, $d\sigma(L) \propto C_L^{e\,2}(1 + \cos^2\theta + 2A_b \cos\theta)$ and $d\sigma(R) \propto C_R^{e\,2}(1 + \cos^2\theta - 2A_b \cos\theta)$ so that for the electron beam with polarization P_e

$$\begin{aligned}
\frac{d\sigma(b)}{d\cos\theta} &\equiv f(b,A) = \left.\frac{d\sigma(b)}{d\cos\theta}\right|_L f_L + \left.\frac{d\sigma(b)}{d\cos\theta}\right|_R f_R \\
&\propto C_L^{e\,2}(1 + \cos^2\theta + 2A_b\cos\theta)\frac{1 - P_e}{2} + C_R^{e\,2}(1 + \cos^2\theta - 2A_b\cos\theta)\frac{1 + P_e}{2} \\
&= (1 - A_e P_e)(1 + \cos^2\theta) + 2A_b(A_e - P_e)\cos\theta\,, \qquad (5.1)
\end{aligned}$$

where

$$f_L = \frac{1 - P_e}{2} \quad \text{and} \quad f_R = \frac{1 + P_e}{2}$$

are fractions in the admixture of left and right handed electrons in the incident e^- beam (polarization is actually defined from these as $P_e = f_R - f_L$). The asymmetry for a fermion f is defined as

$$A_f = \frac{C_L^{f\,2} - C_R^{f\,2}}{C_L^{f\,2} + C_R^{f\,2}}\,,$$

and $C_L^f = (g_V^f + g_A^f)/2$ and $C_R^f = (g_V^f - g_A^f)/2$ through the vector and axial couplings as usual.

Similar equation can be obtained for $d\sigma(\bar{b})/d\cos\theta \equiv f(\bar{b}, A)$ by just reversing the sign of $\cos\theta$ in (5.1). The probability of a b given the $A = (\cos\theta, P_e)$ is then calculated as

$$f(b \mid A) = \frac{f(b,A)}{f(b,A) + f(\bar{b},A) + f(x,A)} = f(b\bar{b})\left[\frac{1}{2} + \frac{A_b(A_e - P_e)}{1 - A_e P_e}\,\frac{\cos\theta}{1 + \cos^2\theta}\right]\,, \quad (5.2)$$

where x denotes any quark other than a b or a $\bar{b}$. [The (5.2) has been obtained by taking into account that $f(b, A) = f(b \mid A)f(A)$, $f(x, A) = f(x \mid A)f(A) = f(x)f(A)$ since x and A are considered mutually independent, and $f(b\bar{b}) = 1 - f(x)$.] Radiative corrections are small and are not taken into account in our calculations of the branching fractions. They are studied in detail in [48, 74].

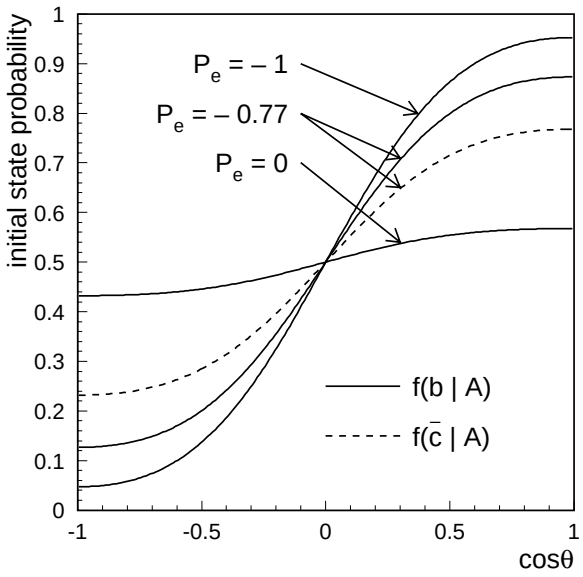


Figure 5.2: Probability $f(b \mid A)$ as a function of $\cos\theta$, Eq. (5.2), plotted for 3 different values of electron polarization P_e. Probability $f(\bar{c} \mid A)$ (dashed line) is only plotted for $P_e = -0.77$, which is approximately the electron polarization at the SLD. We used the PDG value $A_c = 0.65$ [7].

Strictly, $f(x \mid A) = f(x)$ is not correct, because the non $b\bar{b}$ events are mostly $c\bar{c}$, Fig. 5.1, and the c and $\bar{c}$ are correlated to the $\cos\theta$ and P_e. Since the $\bar{c}$ is coupled to W^- in the same way as the b (with the $A_c \approx 0.65$), and decays into leptons of the same sign as that in the prompt b decays, then the probability $f(\bar{c} \mid A)$ has the same dependence on $\cos\theta$ and P_e when A_b is replaced with $-A_c$, and $\cos\theta$ with $-\cos\theta$, Fig. 5.2. Fortunately, the fraction of leptons from c or $\bar{c}$ decays in the entire population of the background leptons is small, and the correlation effects can be neglected.

5.3 Flavor tag from jet charge

An inclusive quantity highly correlated to the flavor of the b hadrons is known as the *jet charge* and is defined as

$$Q = \sum_i q_i |\ \mathbf{p_i} \cdot \mathbf{t}\ |^{\kappa} \operatorname{sign}(\mathbf{p_i} \cdot \mathbf{t}) \tag{5.3}$$

where q_i and $\mathbf{p_i}$ are charges and 3-momenta of the tracks, $\mathbf{t}$ the thrust axis, and κ the phenomenological parameter designed to enhance the discrimination between b and $\bar{b}$. The sum in (5.3) can be taken over all tracks in the event, with the results shown in the top left plot of Fig. 5.3. In that case, however, large correlation between the final state lepton charge and the jet charge Q is observed. The net effect of the correlation is a positive bias toward $\mathcal{B}(b\to l)$ of about 15%, and a negative bias toward $\mathcal{B}(b\to c\to l)$ of about the same size. We therefore calculate the jet charge using only tracks from a single hemisphere, opposite to the hemisphere in which a lepton is tagged, top right plot in Fig. 5.3.

We assume Gaussian distributions of the jet charge for b and $\bar{b}$ hypotheses:

$$f(b,Q) = \frac{1}{\sqrt{2\pi}\sigma}\,\exp\left(-\frac{(Q-Q_b)^2}{2\sigma^2}\right), \qquad f(\bar{b},Q) = \frac{1}{\sqrt{2\pi}\sigma}\,\exp\left(-\frac{(Q-Q_{\bar{b}})^2}{2\sigma^2}\right),$$

where Q_b and $Q_{\bar{b}}$ are the corresponding central values, and the variance σ^2 is assumed equal for both b and $\bar{b}$. Similarly as in Eq. (5.2) one has (assuming $Q_{\bar{b}} = -Q_b$)

$$f(b \mid Q) = \frac{f(b,Q)}{f(b,Q) + f(\bar{b},Q) + f(x,Q)} = \frac{f(b\bar{b})}{1+\exp(-\alpha Q)}, \tag{5.4}$$

where $\alpha = 2Q_b/\sigma^2$ (< 0). From the top right plot in Fig. 5.3, $\langle Q_b\rangle = 1.09$, $\sigma = 2.6$, $\Rightarrow \alpha = 0.323$. Note that the value of the Q_b in a single hemisphere is 1/2 of that of that for both hemispheres, while the value of σ reduces by $1/\sqrt{2}$. The value of α remains unchanged. The probability frequencies for double and single hemisphere jet charge are shown in the lower row of Fig. 5.3.

5.4 Combined initial state flavor tag

One can easily show that for arbitrary random variables $X_1, \ldots, X_n$ and Y satisfying

$$f(x_1, \ldots, x_n \mid y) = \prod_{i=1}^{n} f(x_i \mid y), \tag{5.5}$$

one has

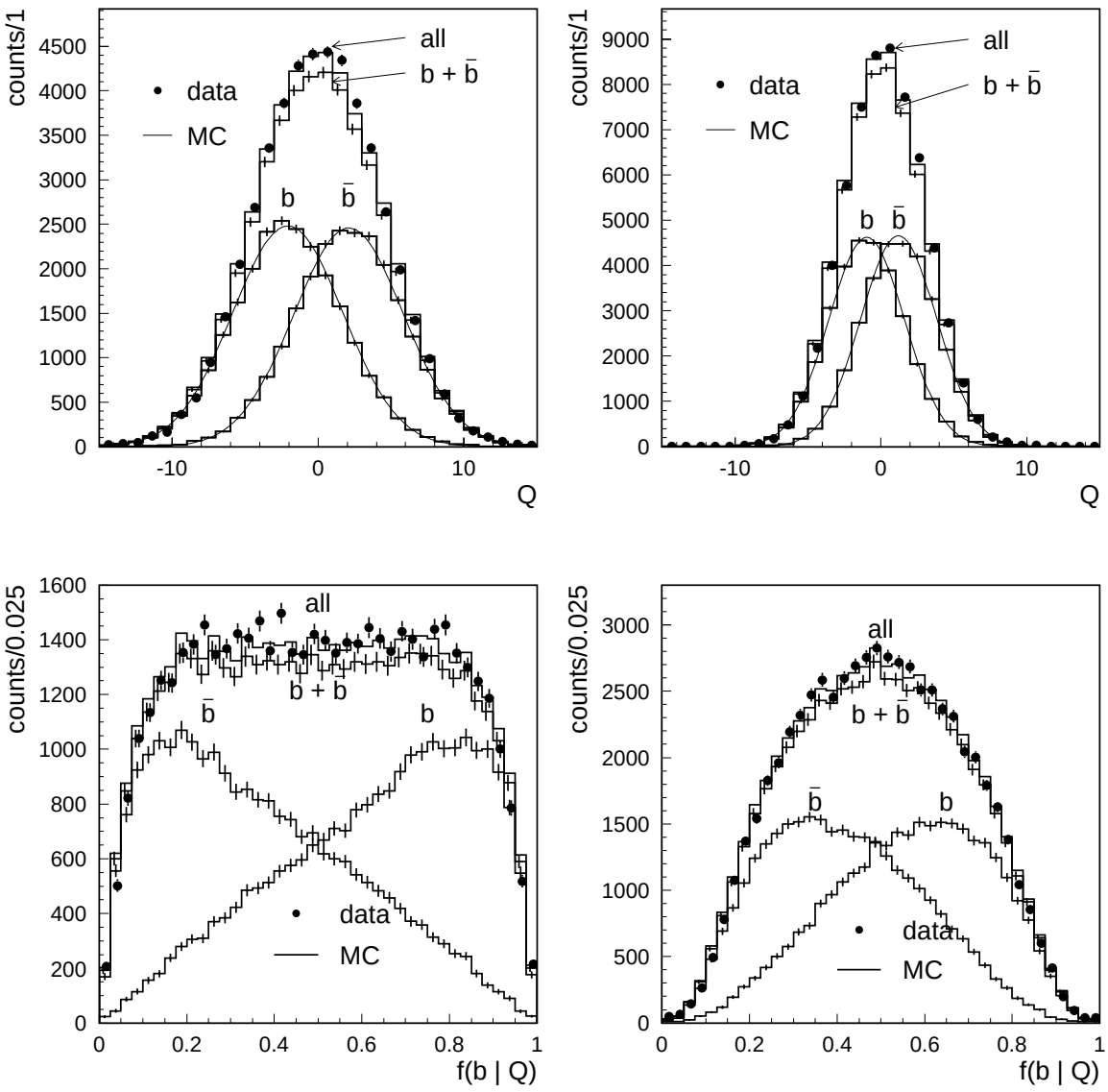


Figure 5.3: Top row: jet charge calculated using tracks from both hemispheres (left), and tracks from single hemispheres (right). Bottom row: the corresponding initial state probabilities.

$$f(y \mid x_1, \ldots, x_n) = \frac{\prod_{i=1}^{n} f(y \mid x_i)}{\int dy' \left[\frac{f(y)}{f(y')}\right]^{n-1} \prod_{i=1}^{n} f(y' \mid x_i)} \,. \tag{5.6}$$

Here Y is a random variable indicating whether a quark is b or $\bar{b}$, in a sample of 100% pure $b\bar{b}$ events, and X_1 and X_2 correspond to the outcomes of the jet charge and the asymmetry tests respectively: $X_1 \equiv Q$, and $X_2 \equiv A$. Assuming Q and A independent, as in Eq. (5.5), one has

$$f(b \mid QA) = \frac{f(b \mid Q) f(b \mid A)}{f(b \mid Q) f(b \mid A) + f(\overline{b} \mid Q) f(\overline{b} \mid A)} \,, \tag{5.7}$$

directly from Eq. (5.6), by taking into account $f(b) = f(\bar{b})$ (otherwise factor $f(\bar{b})/f(b)$ multiplies the second term in the denominator).

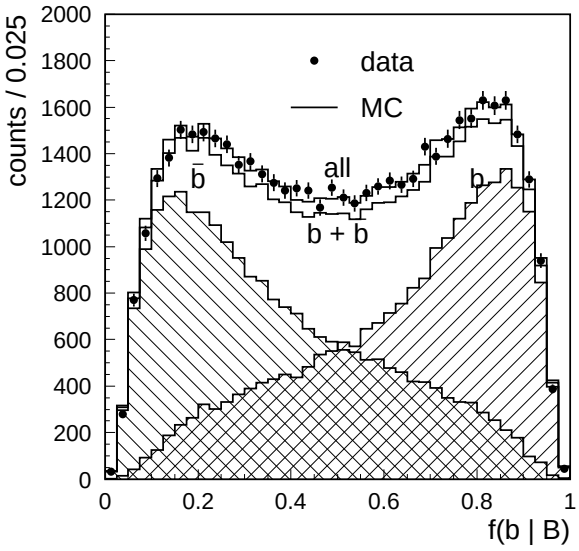


Figure 5.4: Frequency distribution as a function of $f(b \mid B)$, for all the data combined (1993-98). The average polarization is < 0.6 for the 1993 data, and around 0.77 for all other data sets.

If the sample is not composed of 100% pure $b\bar{b}$, then the Y could be b, $\bar{b}$, or something else, which we denote x. In that case, according Eq. (5.6),

$$f(b \mid QA) = \frac{f(b \mid Q) f(b \mid A)}{f(b \mid Q) f(b \mid A) + f(\overline{b} \mid Q) f(\overline{b} \mid A) + \frac{f(b)}{f(x)} f(x \mid Q) f(x \mid A)} \,, \tag{5.8}$$

again assuming that $f(b) = f(\bar{b})$. There is an additional simplification of this expression arising from the property that x is independent with respect to Q and A, which implies $f(x \mid Q) = f(x)$ and $f(x \mid A) = f(x)$. By taking into account $f(b) = f(\bar{b}) = f(b\bar{b})/2$, $f(b\bar{b}) = 0.964$ from the MC, and $f(x) = 1 - f(b\bar{b}) = 0.036$, one gets, Fig. 5.4,

$$f(b \mid QA) = \frac{f(b \mid Q)f(b \mid A)}{f(b \mid Q)f(b \mid A) + f(\overline{b} \mid Q)f(\overline{b} \mid A) + 0.0174} \,.$$

6

Calculation of the branching fractions

Results from the previous three chapters are put together here to estimate the values of $\mathcal{B}(b\to l)$ and $\mathcal{B}(b\to c\to l)$. Joint confidence regions for $\mathcal{B}_L$ and $\mathcal{B}_U$ are first determined from the likelihood function (3.19) bin-by-bin in momentum. Inputs to the likelihood function are efficiency η, background + mis-ID rate $\tilde{\beta}+\delta$, and initial state probabilities $f(b \mid \mathrm{B})$ and $f(\bar{b} \mid \mathrm{B})$. All other inputs are either from other experiments or from the MC. The value of $\tilde{\beta}+\delta$, which is independently calibrated using results from Ch. 4, is cross-checked by using probability relations (3.11) and (3.15) for pairs of leptons in the same hemisphere. The $\mathcal{B}(b\to l)$ is at the end obtained by subtracting branching fractions of the like sign cascades $\mathcal{B}(b\to\bar{c}\to l)$ and $\mathcal{B}(b\to\tau\to l)$ from $\mathcal{B}_L$.

6.1 Preliminaries: efficiency corrections

The subset of tracks used in efficiency studies of Ch. 4 was restricted to a subset of all *reconstructed*, not all *charged* tracks, and was restricted in polar angle to $|\cos\theta| < 0.7$. The acceptance region for reliably reconstructed tracks at the SLD is $|\cos\theta| < 0.8$ (for the VXD2 and the VXD3 data summed), which is also the range of the reconstructed thrust directions. For thrust directions near the boundaries of the acceptance region, the tracking efficiency for tracks in the b tagged hemisphere is additionally reduced: lower momentum tracks tend to curl-out of the acceptance region and escape undetected.

One correction that is needed, therefore, is for the tracking efficiency. Another has to take into account a small fraction of tracks in $0.7 < |\cos\theta| < 0.8$ for which the efficiency drops to about one fifth of that in the $|\cos\theta| < 0.7$ region, and yet another a small fraction of tracks in $|\cos\theta| > 0.8$ which all escape undetected. This is all illustrated in the subsequent figures.

Fig. 6.1 shows correlations between thrust direction and direction of tracks tagged as leptons in b tagged hemisphere.

Fig. 6.2 shows various stages in track selection and lepton identification. The top row refers to electrons and the bottom row to muons. In the top left plot, 4π distribution in $\cos\theta$ for all true electrons that have a B or a D parent (top line) is

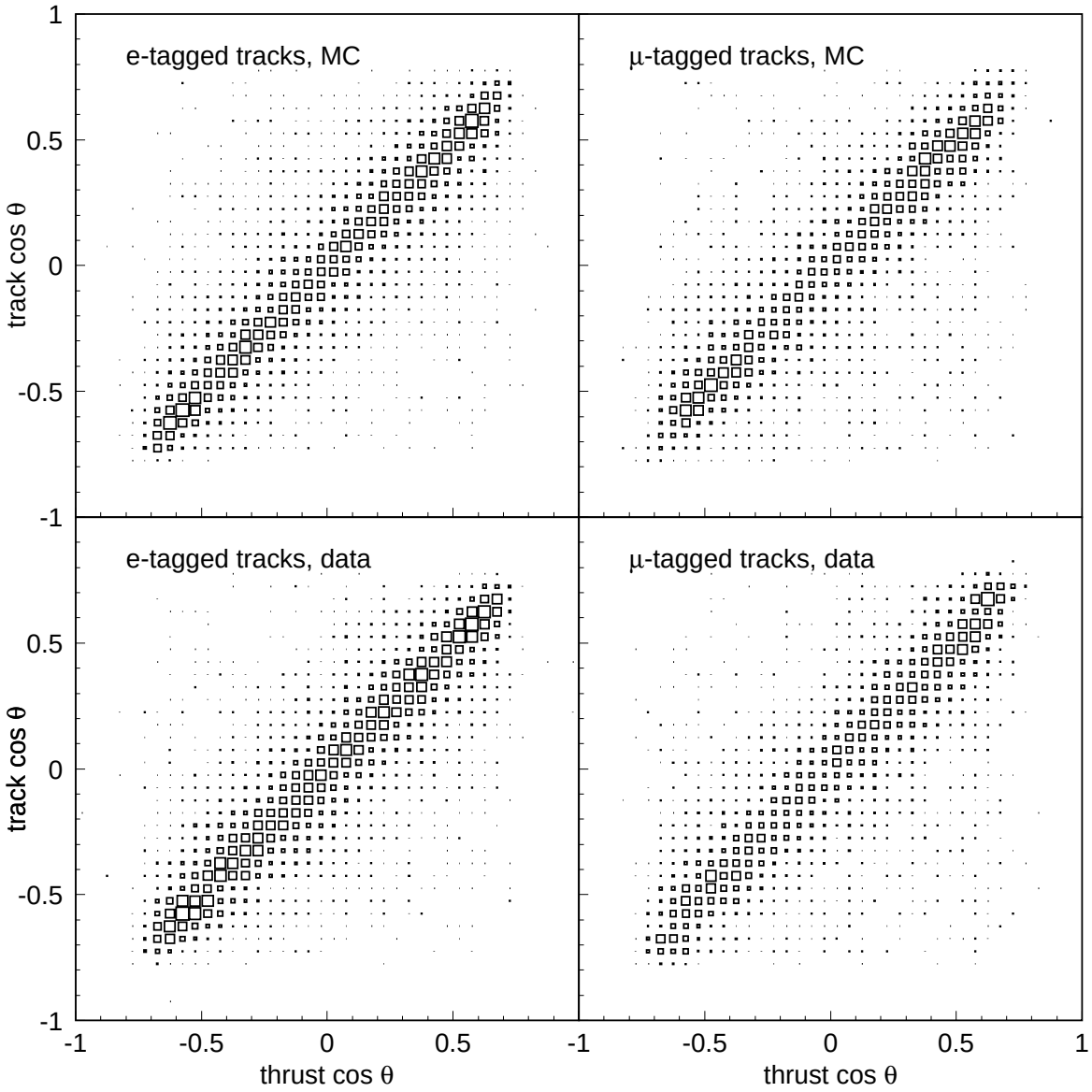


Figure 6.1: Correlations between $\cos\theta$ of reconstructed tracks in b tagged hemispheres and $\cos\theta$ of thrust directions.

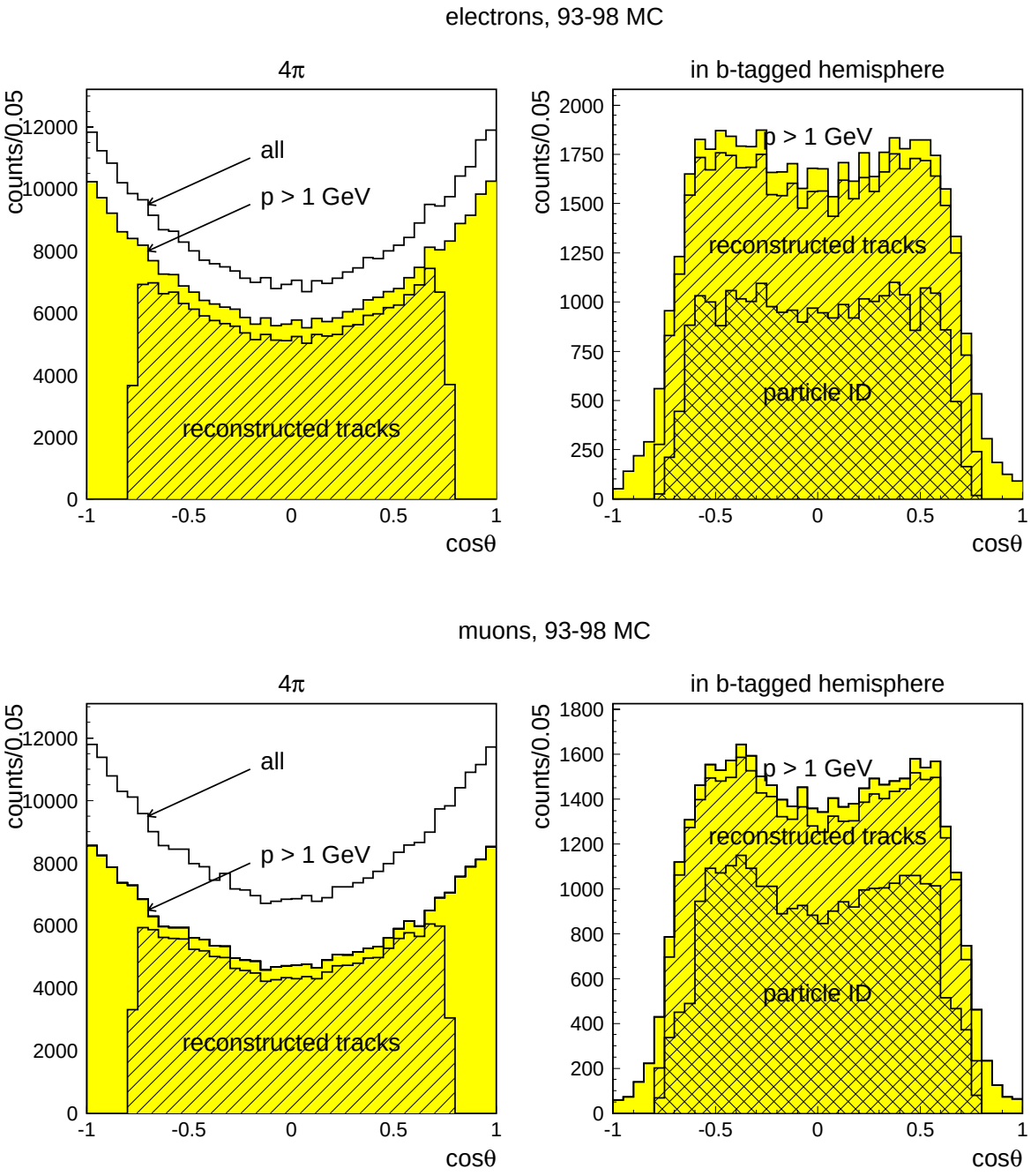


Figure 6.2: Distributions for various subsets of tracks. Distributions in the left column are for the 4π tracks, while in the right column tracks are required to belong to a b-tagged hemisphere.

compared to the distribution of electrons restricted to $p > 1$ GeV (shaded areas), and to the distribution of electrons that are tagged (reconstructed) by the track reconstruction routines (hatched area). The track reconstruction routines introduce an effective cutoff $|\cos\theta| < 0.8$. The ratio between the hatched and the shaded areas in $|\cos\theta| < 0.8$ region is the average tracking efficiency, around 90% for a subset of well reconstructed tracks. Everything repeats for muons (bottom left plot) except that the momentum cut is $p > 2$ GeV.

Plots in the right column of Fig. 6.2 compare similar distributions, but for the tracks that are required to belong to a b-tagged hemisphere. The shaded area in the top right plot is distribution of all true electrons restricted to $p > 1$ GeV. The hatched area is the subset of all true electrons that are also reconstructed by the track reconstruction routines. The doubly hatched area is a subset of that subset that consists of electrons tagged by the electron test. Everything repeats for muons except that the momentum cut is $p > 2$ GeV. What is counted by the detector is therefore represented by the doubly hatched areas, while the shaded areas represent true yields that have to be deduced by knowing the particle identification efficiencies (ratios of the doubly hatched to the singly hatched areas), and the tracking efficiencies (ratios of hatched to shaded areas).

The fraction of tracks that escape undetected, in $0.8 \leq |\cos\theta| \leq 1$ range, is determined from the MC. In principle the problem can be minimized by using a narrower thrust direction cut, say $|\cos\theta| < 0.6$, see Fig. 6.1. The drawback would be the reduced statistics, as well as a lower resolution between the b and the $\bar{b}$ from the polarized asymmetry, which increases with $|\cos\theta|$.

Plots in the top row of Fig. 6.3 show true MC efficiencies for electrons and muons as functions of $\cos\theta$. [Notches seen at $|\cos\theta| \approx 0.5$ for electrons are from the LAC joint washers, Fig. 2.8. Small notches at $|\cos\theta| \approx 0.6$ for muons are from the similar WIC components.]

Different stages of the correction of the lepton efficiency in total momentum are shown in the bottom two plots of Fig. 6.3. Line 1 is the true MC efficiency when the reference set of tracks is taken to be the subset of all reconstructed tracks with the restriction $|\cos\theta| \leq 0.7$. Line 2 is the lepton identification efficiency corrected for the tracking efficiency. The shaded histogram is obtained by taking into account tracks in $|\cos\theta| > 0.7$. Finally *the efficiency relative to all tracks in the events*, represented by the thick line, is obtained when the shaded histogram is scaled by the calibration factor taken from Fig. 4.33 for electrons and Fig. 4.40 for muons. Corrections for the tracking efficiency are rather uniform over the entire momentum range, as expected. Correction for the missing tracks at the endpoints of the acceptance region is visible at lower momenta (low momentum tracks tend to curl-out of the detector acceptance region), and nearly disappears at higher momenta.

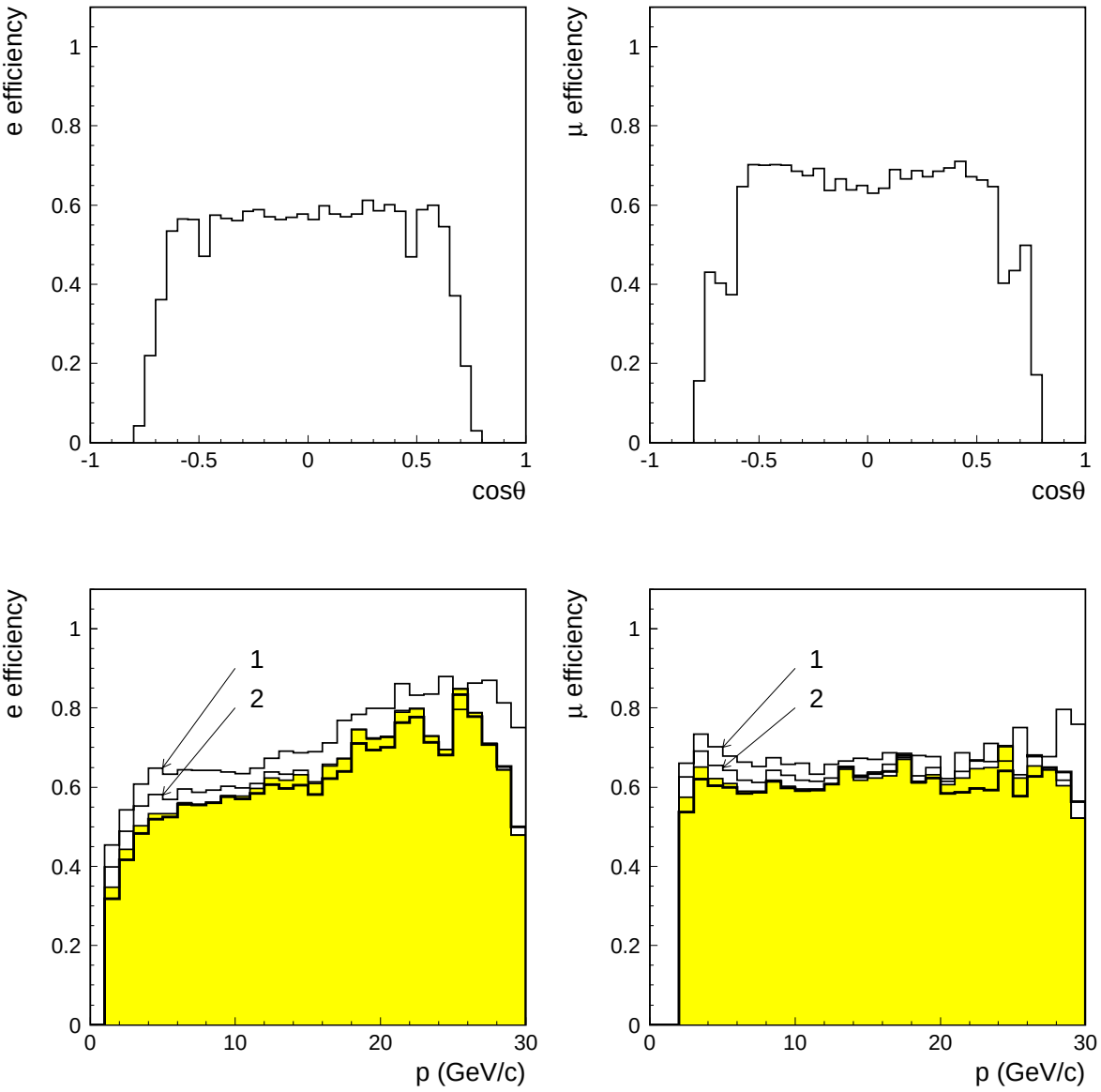


Figure 6.3: Top row: efficiencies as functions of $\cos\theta$. Bottom row: efficiencies as functions of p. Line 1 is the efficiency when the reconstructed tracks are used as a reference sample. Line 2 involves corrections for the tracking efficiency. The shaded area involves corrections for the tracks at the endpoints of the acceptance region in $\cos\theta$. The thick line is calibrated efficiency.

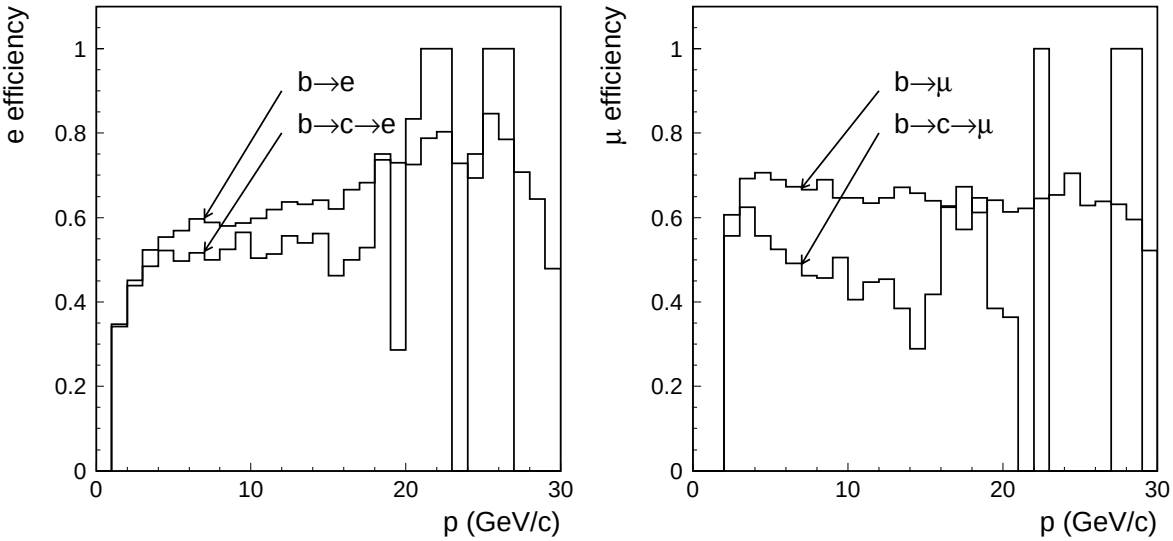


Figure 6.4: True MC efficiencies for $b \to l$ and $b \to c \to l$ class of leptons for both electrons and muons. The discrepancies are explained in the text.

[**An important note:** The SLD Monte Carlo data sets tested by the author, apparently contain an error, which is illustrated in Fig. 6.4. It shows true MC efficiencies for the $b \to l$ and $b \to c \to l$ classes of tracks, for both the electrons (left) and muons (right). The problem is that the SLD Monte Carlo apparently incorrectly links the LAC cluster variables to the particular lepton classes ($b \to l$ and $b \to c \to l$) favoring $b \to l$ over $b \to c \to l$. This is particularly true for some HAD module variables for which the factor is 2. That factor is lower in Fig. 6.4 because the effect is "diluted" somewhat by the presence of other variables. The error is somewhat less visible in electrons, but it is present there as well. The error is purely computational, and has been extensively tested by the author. The bulk lepton yield however ($b \to l$, $b \to c \to l$, ..., summed) is quite correctly tuned to the data, as one can see by examining the conditional probabilities for the outcomes of the various lepton tests of Ch. 4: $f(E \mid H)$ and $f(H \mid E)$ in Fig. 4.24 for electrons, and $f(L \mid W)$ and $f(W \mid L)$ in Fig. 4.35 for muons, all of them showing the discrepancies at the level of $5 - 10\%$ and not a 100%.]

6.2 More preliminaries: the background + mis-ID rate cross-check

Given that the discrepancy in $\tilde{\beta} + \delta$ between the real data and the Monte Carlo is large, Fig. 4.16 and Fig. 4.20, we first want to make a number of calculations involving involving probability relations for pairs of leptons in a hemisphere.

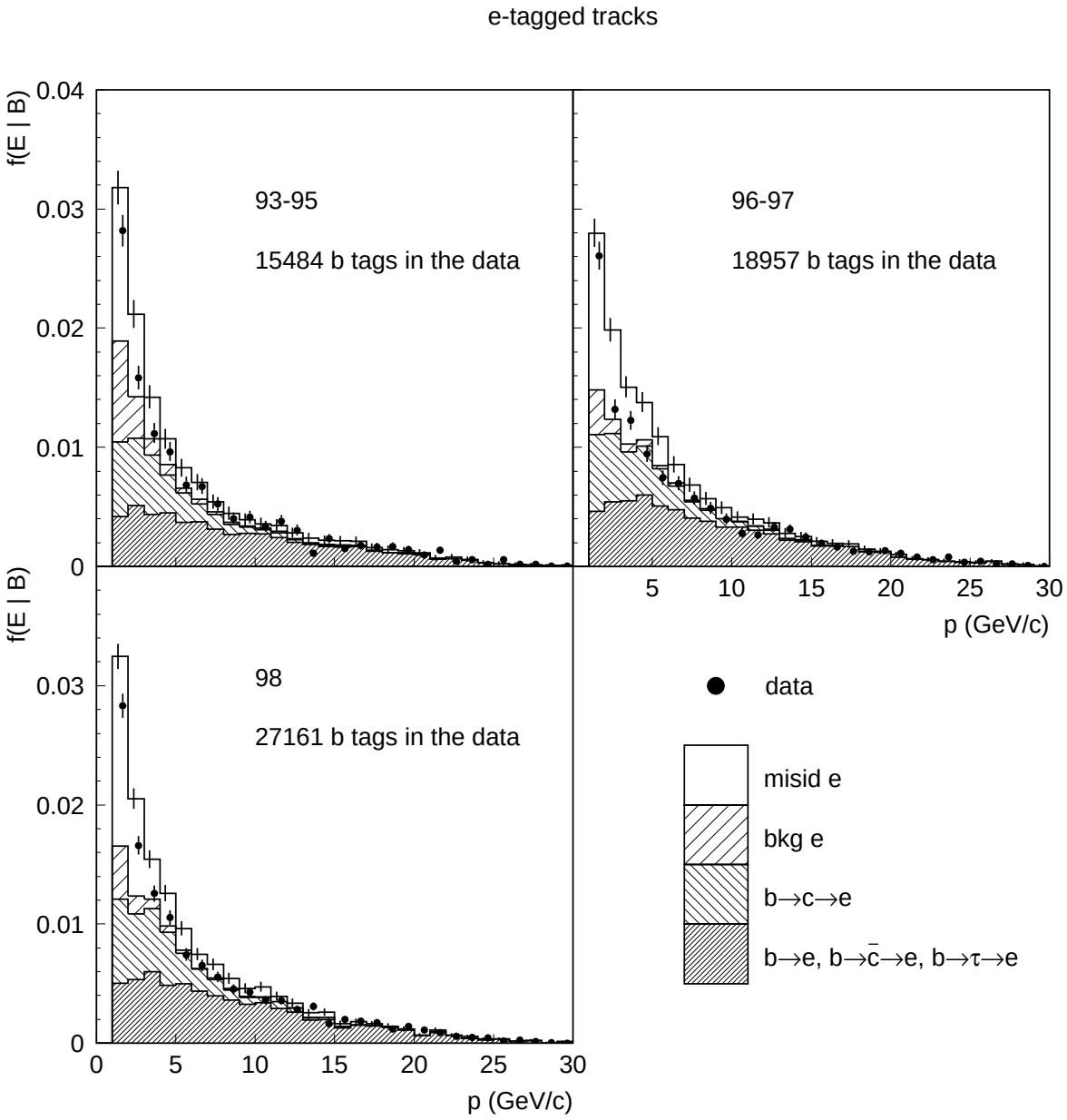


Figure 6.5: Yield of e-tagged tracks in b-tagged hemisphere and their compositions for three periods of the data taking. See Table 6.1 for the related numbers.

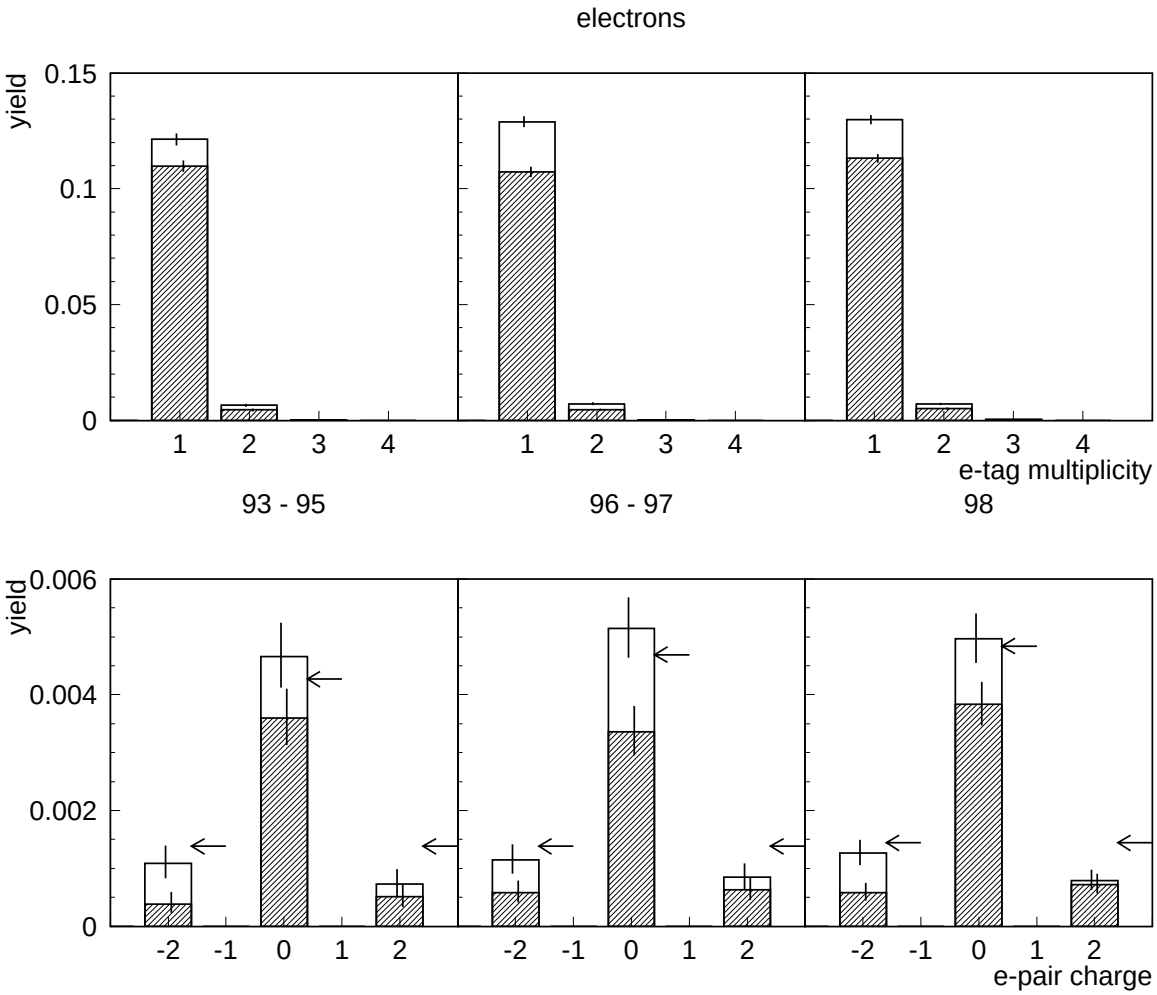
```
n(B)    b tag count
n(L,B)  count of electron tags in b tagged hemisphere
n(bkg)  true background + mis-ID count
n(BL~)  true like sign electron count
n(BU~)  true unlike sign electron count

f(L) = n(L,B)/n(B)    electron tag multiplicity
f(bkg) = n(bkg)/n(B)  true background + mis-ID multiplicity
BL~ = n(BL~)/n(B)     true like sign yield
BU~ = n(BU~)/n(B)     true unlike sign yield
```

Monte Carlo	93-95	96-97	98
n(B)	15283.9	19293.2	27500.4
n(L,B)	2052.18	2771.44	3959.49
n(bkg)	773.103	892.804	1310
n(BL~)	847.961	1299.56	1782.57
n(BU~)	431.119	579.076	866.935
f(L)	0.134271	0.143648	0.14398
f(bkg)	0.050583	0.0462756	0.0476358
BL~	0.0554808	0.0673585	0.0648198
BU~	0.0282075	0.0300145	0.0315245

data			
n(B)	15484	18957	27161
n(L,B)	1826	2204	3353
f(L)	0.117928	0.116263	0.123449

Table 6.1: Summary of the electron counts and fractions for the three data taking periods separately.



f(--)	0.00108905	0.00118391	0.00131708
f(+-)	0.00467672	0.00526096	0.00502906
f(++)	0.000715914	0.000983145	0.000761032

Figure 6.6: Top row: multiplicities of e-tagged tracks. Bottom row: pair charge of e-tagged tracks. Blank histograms are for the MC, hatched histograms are for the data. Arrows indicate the corresponding values in the MC calculated using Eq. (3.11) for unlike sign pairs and Eq. (3.15) for like sign pairs.

6.2.1 Electrons

Yields of e-tagged tracks are shown in Fig. 6.5, for the three data taking periods separately. Somewhat noisier conditions during $96-97$ data taking apparently did not affect the calorimetry readout on which the lepton identification is mostly based. The lower lepton count in the data is systematic in both electrons and muons, as we are going to see, regardless of the data taking period. The numbers related to Fig. 6.5 are summarized in Table 6.1. From these numbers one can calculate the values of $f(L)_{\pm}$ and $f(L)_{\pm\pm}$, using Eq. (3.11) and Eq. (3.15) respectively, and compare them to the directly observed values of $f(L)_{\pm}$ and $f(L)_{\pm\pm}$. This is here done only for the MC by using the measured $f(L)$, true $\tilde{\beta}+\delta$, and true $\tilde{\mathcal{B}}_L$ and $\tilde{\mathcal{B}}_U$. For electrons one has, Eq. (3.11),

$$f(L)_{\pm} = f(b\bar{b})\,\tilde{\mathcal{B}}_L\tilde{\mathcal{B}}_U \;+\; f(\mathrm{L})\,(\tilde{\beta}+\delta) \;-\; (\tilde{\beta}+\delta)^2 \;=$$

$$0.964\times 0.0555\times 0.0282 + 0.134\times 0.0253 - 0.0253^2 \;=\; 0.00426 \quad \text{(93-95 MC)}$$

$$0.964\times 0.0674\times 0.0300 + 0.144\times 0.0231 - 0.0231^2 \;=\; 0.00466 \quad \text{(96-97 MC)}$$

$$0.964\times 0.0648\times 0.0315 + 0.144\times 0.0238 - 0.0238^2 \;=\; 0.00483 \quad \text{(98 MC)}$$

and $f(L)_{\pm\pm} = f(\mathrm{L})\,(\tilde{\beta}+\delta) - (\tilde{\beta}+\delta)^2 = 0.00137$, 0.00140, and 0.00143 for 93-95, 96-97, and 98 MC respectively, in fairly good agreement with the directly observed values for $f(L)_{\pm}$ and $f(L)_{\pm\pm}$, Fig. 6.6.

6.2.2 Muons

Everything repeats for muons. Their yields are shown in Fig. 6.5 also for the three data taking periods separately. The related numbers are summarized in Table 6.2. A systematically lower count of μ-tagged tracks in the data compared to the MC count is also observed. As with the electrons, one can calculate $f(L)_{\pm}$ and $f(L)_{\pm\pm}$ using Eq. (3.11) and Eq. (3.15) respectively to get, Eq. (3.11),

$$f(L)_{\pm} = f(b\bar{b})\,\tilde{\mathcal{B}}_L\tilde{\mathcal{B}}_U \;+\; f(\mathrm{L})\,(\tilde{\beta}+\delta) \;-\; (\tilde{\beta}+\delta)^2 \;=$$

$$0.964\times 0.0569\times 0.0259 + 0.103\times 0.0103 - 0.0103^2 \;=\; 0.00238 \quad \text{(93-95 MC)}$$

$$0.964\times 0.0682\times 0.0269 + 0.122\times 0.0134 - 0.0134^2 \;=\; 0.00322 \quad \text{(96-97 MC)}$$

$$0.964\times 0.0703\times 0.0291 + 0.125\times 0.0129 - 0.0129^2 \;=\; 0.00342 \quad \text{(98 MC)}$$

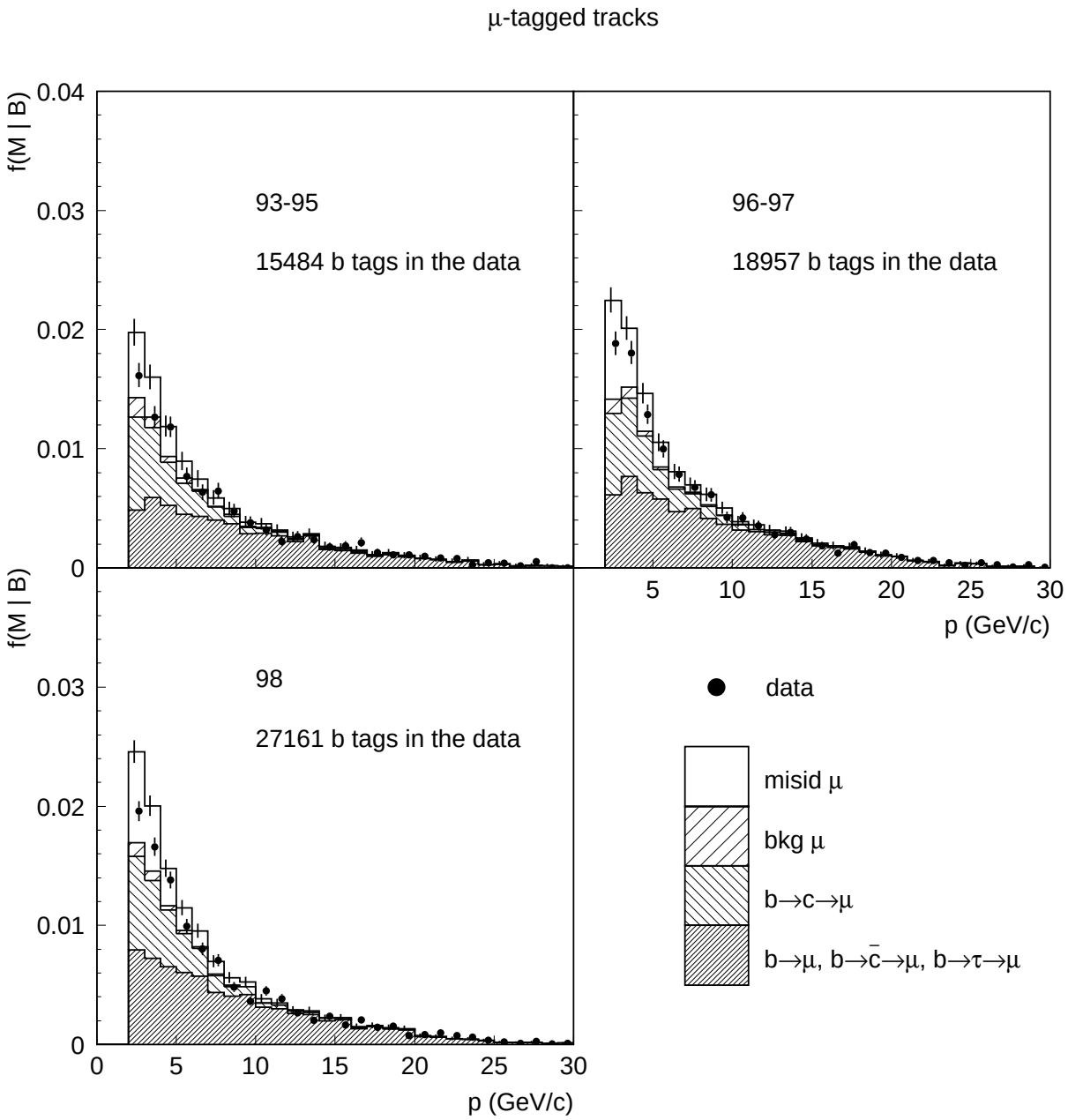
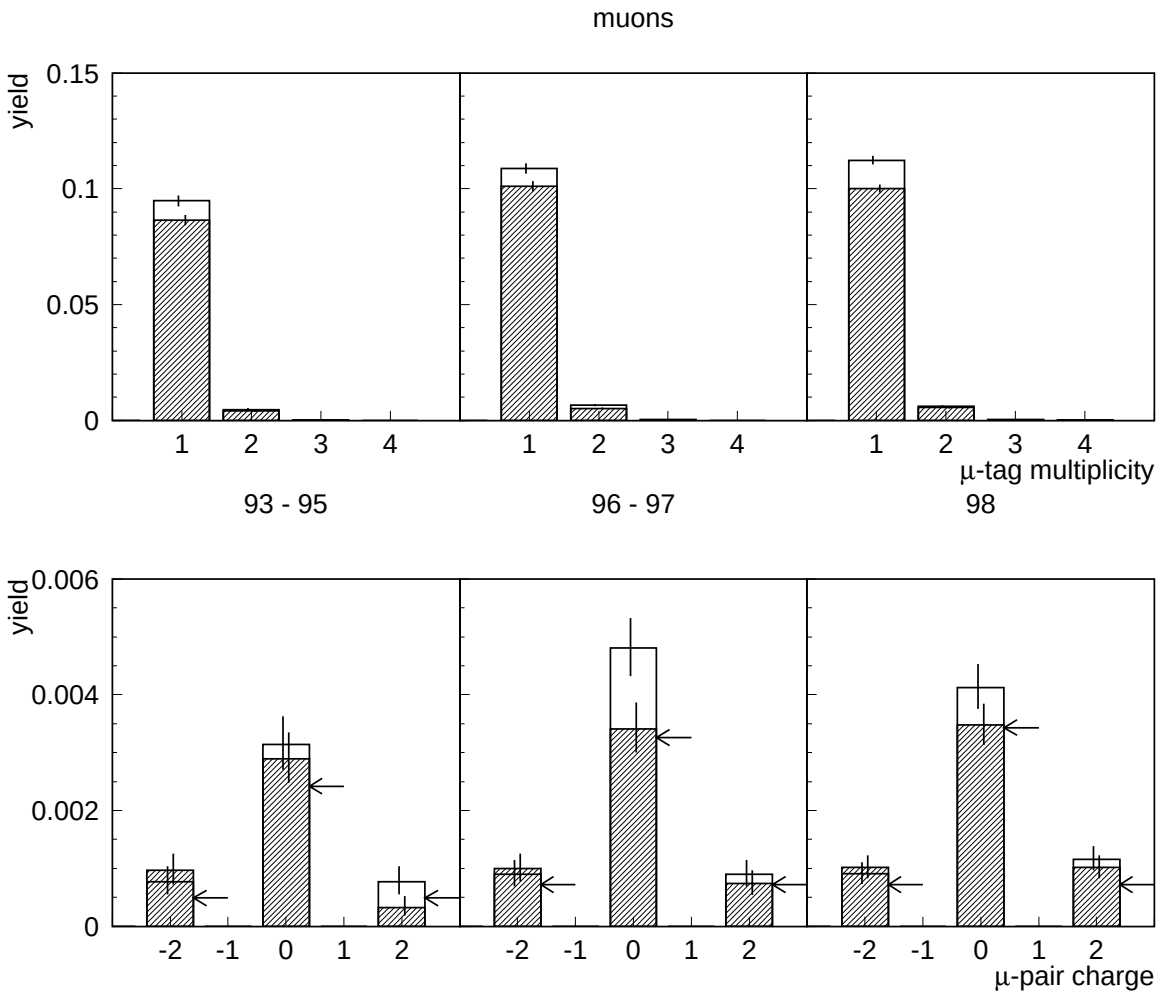


Figure 6.7: Yield of μ-tagged tracks in b-tagged hemisphere and their compositions for three periods of the data taking.



f(--)	0.000727068	0.000866422	0.000906487
f(+-)	0.0031104	0.0046325	0.00406797
f(++)	0.000753804	0.000915034	0.00115207

Figure 6.8: Top row: multiplicities of μ-tagged tracks. Bottom row: pair charge of μ-tagged tracks. Blank histograms are for the MC, hatched histograms are for the data. Arrows indicate the corresponding values in the MC calculated using Eq. (3.11) for unlike sign pairs and Eq. (3.15) for like sign pairs.

```
n(B)    b tag count
n(L,B)  count of muon tags in b tagged hemisphere
n(bkg)  true background + mis-ID count
n(BL~)  true like sign muon count
n(BU~)  true unlike sign muon count

f(L) = n(L,B)/n(B)    muon tag multiplicity
f(bkg) = n(bkg)/n(B)  true background + mis-ID multiplicity
BL~ = n(BL~)/n(B)     true like sign yield
BU~ = n(BU~)/n(B)     true unlike sign yield
```

Monte Carlo	93-95	96-97	98
n(B)	15283.9	19293.2	27500.4
n(L,B)	1581.55	2352.17	3445.05
n(bkg)	315.921	517.709	709.687
n(BL~)	869.312	1315.84	1933.77
n(BU~)	396.326	518.619	801.607
f(L)	0.103479	0.121917	0.125273
f(bkg)	0.0206702	0.0268338	0.0258064
BL~	0.0568778	0.0682023	0.070318
BU~	0.025931	0.0268809	0.0291489
data			
n(B)	15484	18957	27161
n(L,B)	1451	2123	3006
f(L)	0.0937096	0.11199	0.110673

Table 6.2: Summary of the muon counts and fractions for the three data taking periods separately.

and $f(L)_{\pm\pm} = f(\mathrm{L})\,(\tilde{\beta}+\delta) - (\tilde{\beta}+\delta)^2 = 0.00048$, 0.00073, and 0.00072 for 93-95, 96-97, and 98 MC respectively. These numbers turn out to be systematically lower than the ones obtained by direct measurement, Fig. 6.8. The discrepancy is much larger than for the electrons. We have not quantitatively investigated the source of the discrepancy. Apparently, assumptions leading to the derivations of Eq. (3.11) and Eq. (3.15) were too naive, and have not properly taken into account event topologies like the one in Fig. 3.2 which contribute to the crossing of tracks from the "right" to the "wrong" hemisphere, therefore altering the compositions of the pairs.

6.3 Calculation of $\mathcal{B}_L$ and $\mathcal{B}_U$ for electrons

Fig. 6.9 shows the yield of e-tagged tracks for the entire 1993 - 98 data set. Open circles represent $f_i(L \mid \mathrm{B})$ in the real data. Filled circles represent

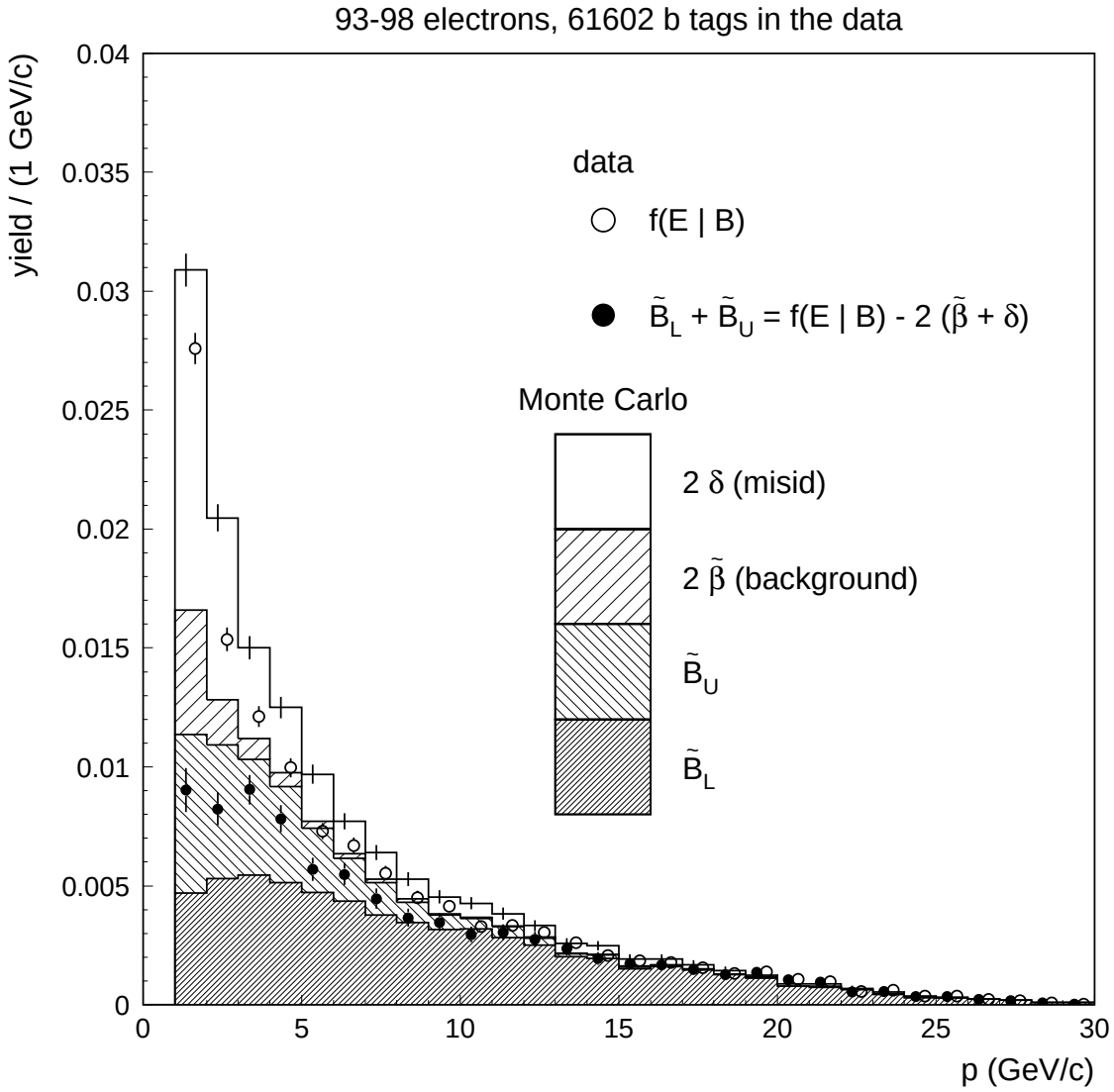
$$\tilde{\mathcal{B}}_{L,i} + \tilde{\mathcal{B}}_{U,i} = f_i(L \mid \mathrm{B}) - 2(\tilde{\beta}_i + \delta_i)$$

also in the real data, where $\tilde{\beta}_i + \delta_i$ is *calibrated* background + mis-ID rate per charge (factor 2 for both signs).

Solutions for $\tilde{\mathcal{B}}_{L,i}$ and $\tilde{\mathcal{B}}_{U,i}$ in the Monte Carlo and the real data are shown in Fig. 6.10. The ellipses represents 68% confidence region for $\mathcal{B}_{L,i}$ and $\mathcal{B}_{U,i}$ jointly, bin-by-bin in momentum (index i), and are obtained as the highest probability density regions of the likelihood function (3.19). The numerical work is done using MINUIT [76]. The plots on the right hand side are just a more traditional way of representing the results, where the error bars correspond to the projections of the confidence regions, and the central values correspond to the outcomes of the point estimators (peaks of the likelihood functions). The MC solutions are compared to the true like and unlike sign distributions (note: this is in the subset of all reconstructed tracks). Polarization of the incident electrons is $P_e = 1$ for the MC and around 0.77 for the data.

The $\tilde{\mathcal{B}}_{L,i}$ and $\tilde{\mathcal{B}}_{U,i}$ are the *visible* branching fractions. The $\mathcal{B}_{L,i}$ and $\mathcal{B}_{U,i}$ are obtained from these as $\mathcal{B}_{L,i} = \tilde{\mathcal{B}}_{L,i}/\eta_i$ and $\mathcal{B}_{U,i} = \tilde{\mathcal{B}}_{U,i}/\eta_i$, for the true MC efficiencies η_i taken from Fig. 6.4. The solutions are practically bias free when compared to the 4π generator level distributions, Fig. 6.11 and Table 6.3. Note that the solutions use no Monte Carlo information at all.

Solutions $\mathcal{B}_{L,i}$ and $\mathcal{B}_{U,i}$ for the real data, Fig. 6.12, are obtained from $\tilde{\mathcal{B}}_{L,i}$ and $\tilde{\mathcal{B}}_{U,i}$, Fig. 6.10, using calibrated efficiency η, Fig. 6.3, and calibrated background + mis-ID rate $\tilde{\beta} + \delta$, Fig. 6.9.



	93-98 Monte Carlo	93-98 data
n(B)	62077.8	61602
n(L,B)	8783.15	7383
n(bkg)	2975.93	2351.2
f(L)	0.141486	0.11985
1/2 f(bkg)	0.0239694	0.0190838

Figure 6.9: Yield of e-tagged tracks for the entire 1993 - 98 data set. Open circles are the total yield in the data, filled circles the combined $\tilde{\mathcal{B}}_L + \tilde{\mathcal{B}}_U$, obtained by subtracting the total background + mis-ID $2(\tilde{\beta} + \delta)$ from $f(L)$.

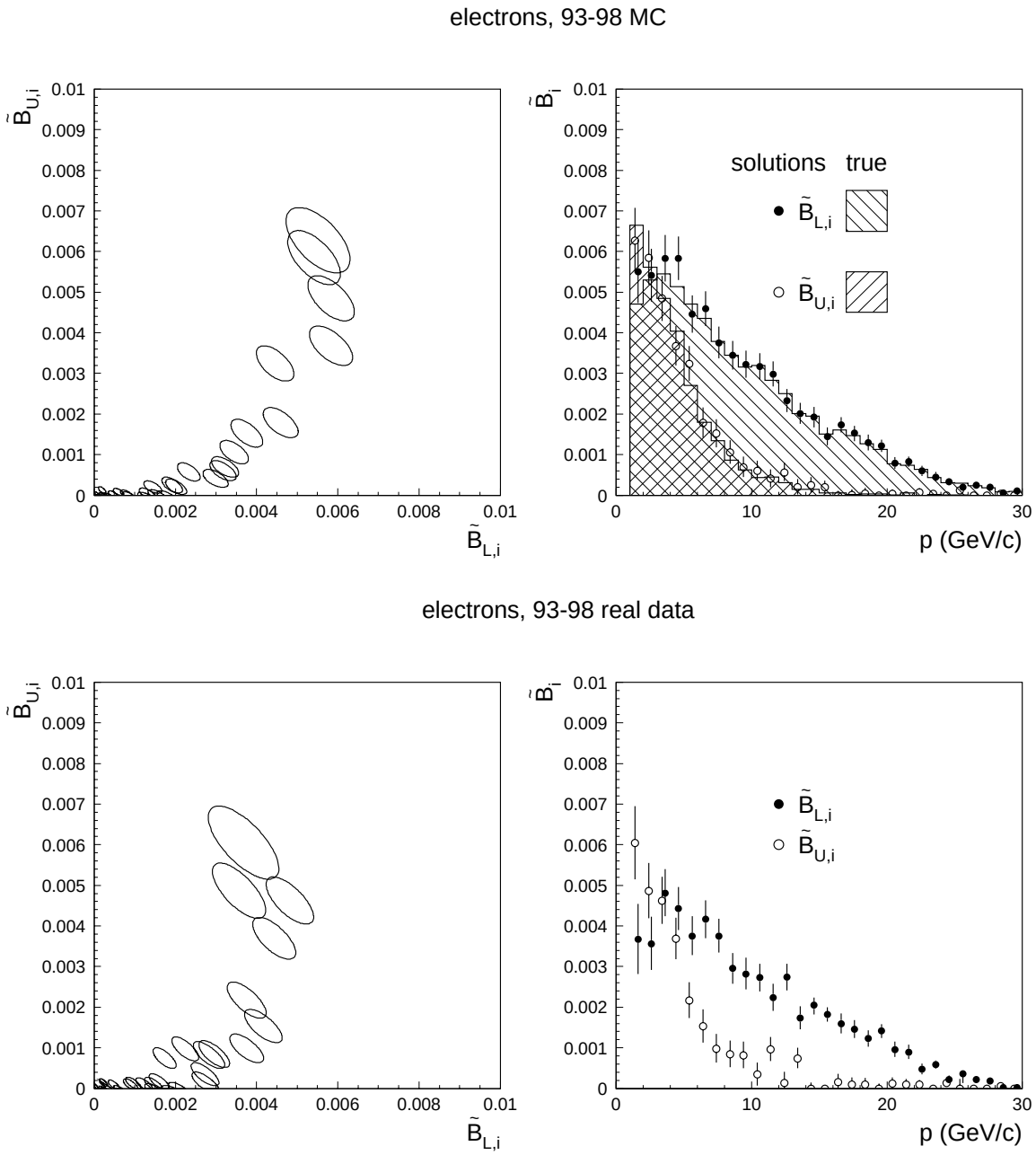


Figure 6.10: Solutions for $\tilde{\mathcal{B}}_{L,i}$ and $\tilde{\mathcal{B}}_{U,i}$ bin-by-bin in total momentum for electrons, MC top row, and data bottom row. Each ellipse in the left hand plots represents 68% joint confidence region for $\tilde{\mathcal{B}}_{L,i}$ and $\tilde{\mathcal{B}}_{U,i}$ in i-th momentum bin. Projections of the confidence regions in terms of error bars are shown in the second plot with the circles being their central values. The MC histograms show true values of $\tilde{\mathcal{B}}_{L,i}$ and $\tilde{\mathcal{B}}_{U,i}$.

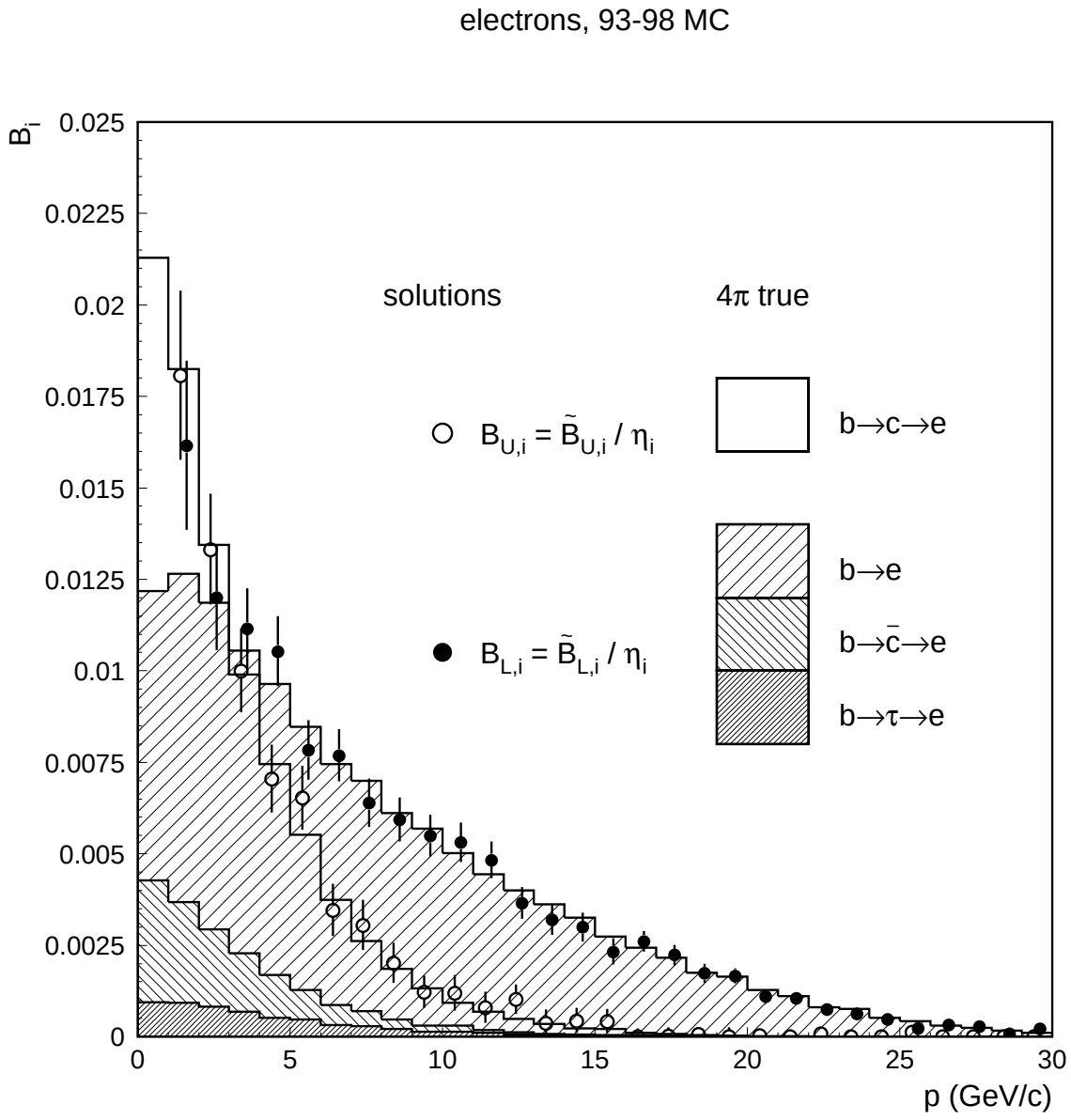


Figure 6.11: MC solutions of Fig. 6.10 scaled by the MC efficiencies of Fig. 6.4. The solutions are compared to the 4π generator level distributions.

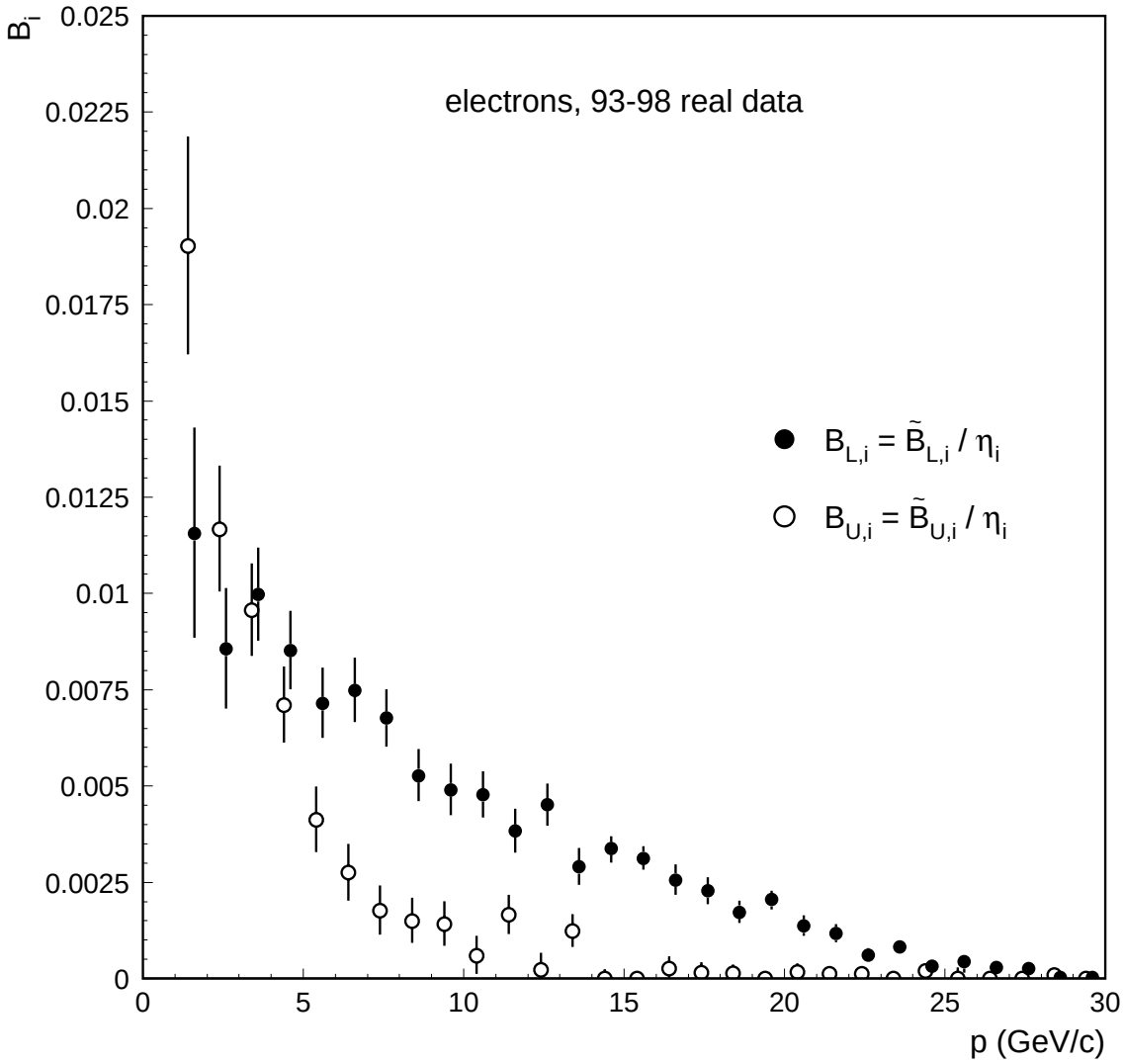


Figure 6.12: Solutions of Fig. 6.10 scaled by the calibrated efficiency of Fig. 6.3.

```
                    ELECTRONS
                    ---------

Solutions for the Monte Carlo in p > 1 GeV
------------------------------------------
BL~ = 0.0654428 + 0.00178975 - 0.00174124
BU~ = 0.0314657 + 0.0016331  - 0.00155158

BL  = 0.118735  + 0.0037092  - 0.00361976
BU  = 0.0691856 + 0.0036944  - 0.00355518

true  all 4 pi          4 pi in p > 1 GeV     ratio
----  --------          -----------------     -----
BL    0.132296          0.119091              1.11088
BU    0.0886017         0.0673175             1.31618

Solutions for the data in p > 1 GeV
-----------------------------------
BL~ = 0.0568537 + 0.00189733 - 0.00186229
BU~ = 0.0285873 + 0.00180435 - 0.00168215

BL  = 0.106649  + 0.00422903 - 0.00414225
BU  = 0.0638867 + 0.00416202 - 0.00399461
```

Table 6.3: Summary of the main results for electrons.

Since $\mathcal{B}_{L,i} = \tilde{\mathcal{B}}_{L,i}/\eta_i$ and $\mathcal{B}_{U,i} = \tilde{\mathcal{B}}_{U,i}/\eta_i$ in individual bins are statistically independent, one has

$$\mathcal{B}_L = \sum_i \mathcal{B}_{L,i}\,, \qquad \sigma_{\mathcal{B}_L} = \sqrt{\sum_i \sigma^2_{\mathcal{B}_L}}\,,$$

and similarly for $\mathcal{B}_U$, Table 6.3.

Results for the generator level branching fractions in Table 6.3 are based on the B decay models and the fragmentation function implemented in the SLD MC. The results will be corrected in Sec. 6.5 for the measured values of the fragmentation function and for the central values in the momentum distributions of the B decay models. This is needed to ensure the correct values of the ratios between the entire and

the visible portions of the spectra. For electrons the visible portion of the spectrum is $p > 1$ GeV, and for muons $p > 2$ GeV. For the real data entries from Table 6.3 one has

$$\begin{aligned} \mathcal{B}_L &= 1.111 \times (0.1066 \pm 0.0042) = 0.1183 \pm 0.0047\,, \\ \mathcal{B}_U \equiv \mathcal{B}(b{\to}c{\to}e) &= 1.316 \times (0.0639 \pm 0.0041) = 0.0841 \pm 0.0054\,. \end{aligned} \tag{6.1}$$

The above calculations can be cross checked using Eq. (3.11) for pairs of like sign tracks and Eq. (3.15) for pairs of unlike sign tracks. The track multiplicity and the pair charge for electrons are given in Fig. 6.13 with the content of each individual bin printed out. We cross-check of the values of $f(L)_{\pm}$ and $f(L)_{\pm\pm}$ in the Monte Carlo and in the data by inserting $f(L)$ and $\tilde{\beta} + \delta$ from Fig. 6.9, and estimated $\tilde{\mathcal{B}}_L$ and $\tilde{\mathcal{B}}_U$ from Table 6.3 into Eq.(3.11):

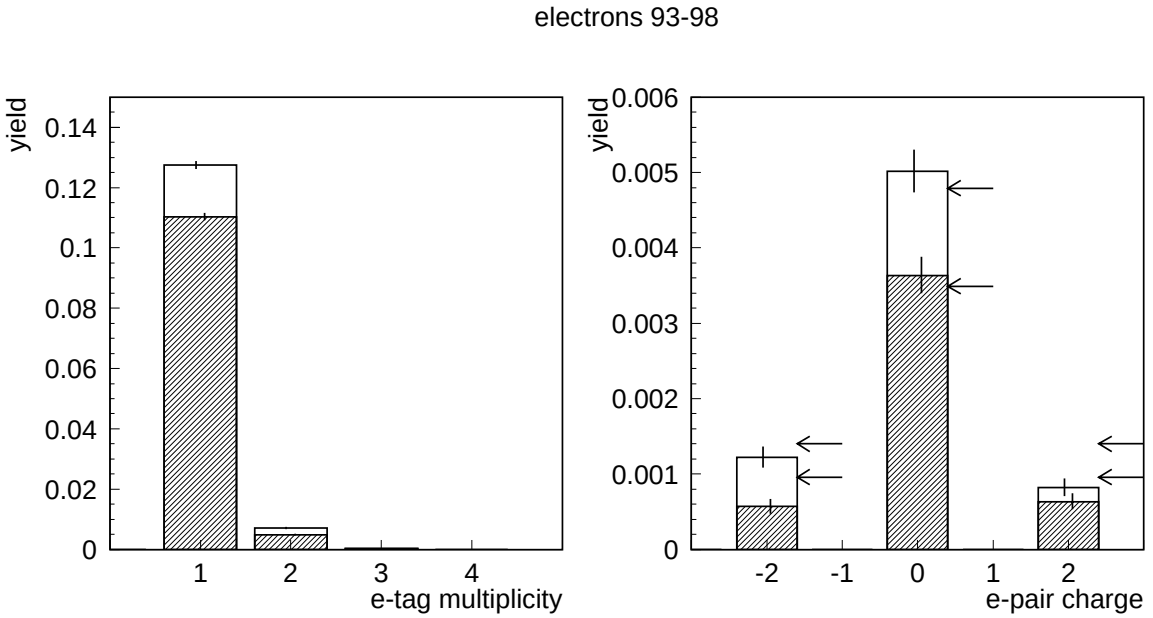
$$f(L)_{\pm} = f(b\bar{b})\,\tilde{\mathcal{B}}_L\tilde{\mathcal{B}}_U \;+\; f(\mathrm{L})\,(\tilde{\beta}+\delta) \;-\; (\tilde{\beta}+\delta)^2 \;=$$

$$\begin{aligned} 0.964 \times 0.0654 \times 0.0315 + 0.141 \times 0.0240 - 0.0240^2 &= 0.00479 \quad \text{(93-98 MC)} \\ 0.964 \times 0.0569 \times 0.0286 + 0.120 \times 0.0190 - 0.0190^2 &= 0.00349 \quad \text{(93-98 data)} \end{aligned} \tag{6.2}$$

Note that the estimated values of $\tilde{\mathcal{B}}_L$ and $\tilde{\mathcal{B}}_U$ have been used above, and that the value of $\tilde{\beta} + \delta$ in the data is the calibrated value. For the values of $f(L)_{\pm\pm}$ one similarly gets 0.00281 and 0.00192 for the MC and the data respectively (0.00140 and 0.00096 are half of these when counted per sign). These numbers are shown with arrows in Fig. 6.13.

While there is a substantial discrepancy between the calculated and the directly measured values of $f(L)_{\pm}$ and $f(L)_{\pm\pm}$, the ratios between the data and the MC values of these quantities turned out to be the same in the directly measured and the calculated quantities:

$$\frac{f(L)_{\pm}(\text{data})}{f(L)_{\pm}(\text{MC})} = \begin{cases} \dfrac{0.00364}{0.00501} = 0.727 & \text{directly measured} \\ \dfrac{0.00341}{0.00473} = 0.729 & \text{calculated} \end{cases}$$



	Monte Carlo	data		Monte Carlo	data
1	0.127399	0.11037	f(--)	0.00121955	0.000568163
2	0.00705289	0.00483751	f(+-)	0.00501439	0.00363625
3	0.000225598	0.000227265	f(++)	0.000818955	0.000633096
4	1.69674E-06				

Figure 6.13: Multiplicity and pair charge for e-tagged tracks. Blank histograms are for the MC; hatched histograms are for the data. The arrows indicate the values of the corresponding quantities calculated using Eq. (3.11) for unlike sign pairs and Eq. (3.15) for like sign pairs. Upper arrows are for the MC; lower arrows are for the data.

and

$$\frac{f(L)_{\pm\pm}(\text{data})}{f(L)_{\pm\pm}(\text{MC})} = \begin{cases} \dfrac{0.00120}{0.00204} = 0.588 & \text{directly measured} \\ \dfrac{0.00192}{0.00281} = 0.683 & \text{calculated} \end{cases}$$

in good agreement given the large errors associated with these ratios (which were not calculated).

6.4 Calculation of $\mathcal{B}_L$ and $\mathcal{B}_U$ for muons

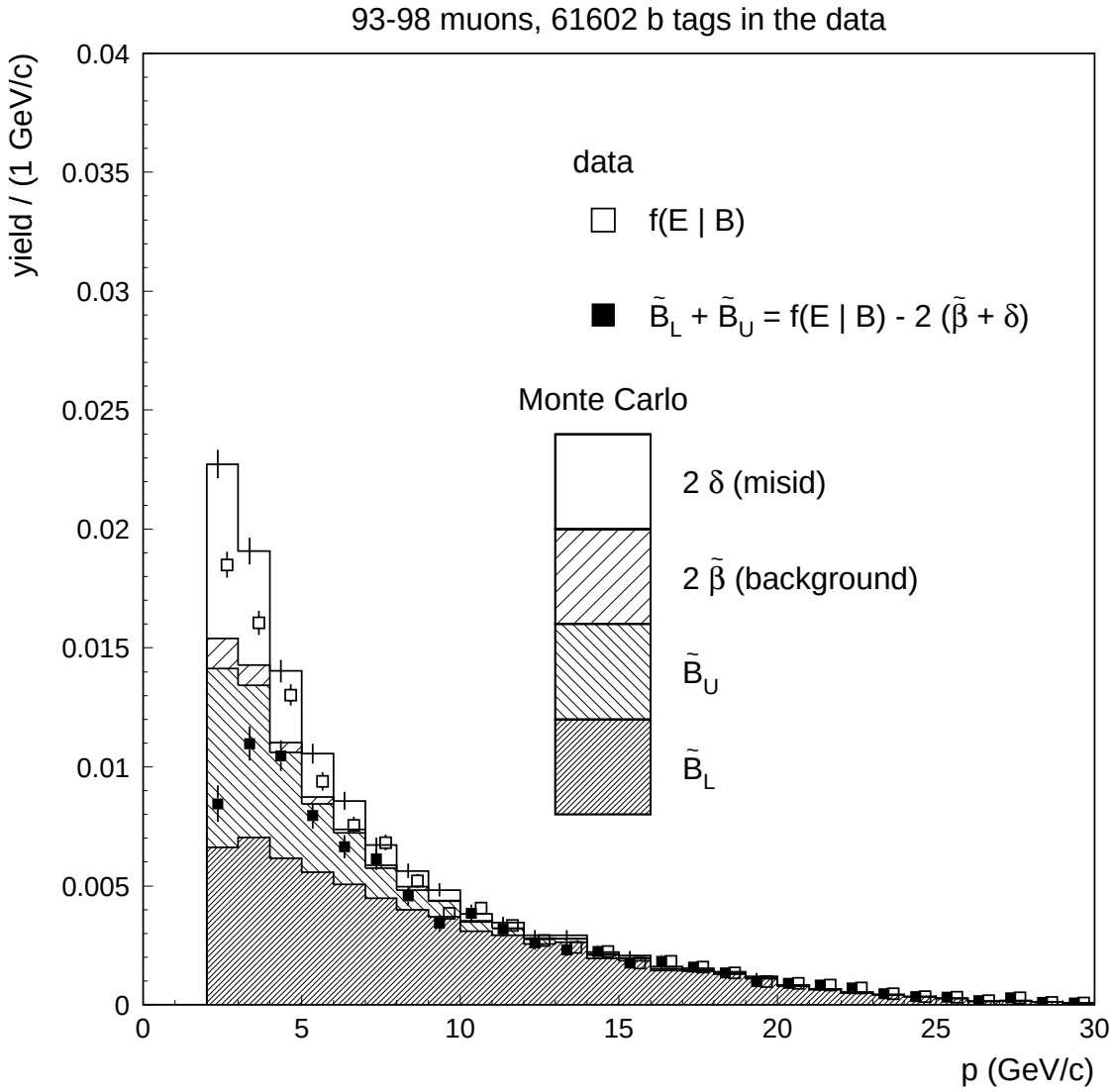
Everything goes the same for muons: Fig. 6.9 shows yield of μ-tagged tracks for the entire data-set used in this analysis (93-98), open circles are $f_i(L \mid \text{B})$ for the data, and filled circles are the differences between $f_i(L \mid \text{B})$ and the calibrated $\tilde{\beta}_i + \delta_i$, i.e. the sum of $\tilde{\mathcal{B}}_{L,i}$ and $\tilde{\mathcal{B}}_{U,i}$.

Everything also repeats for the calculations of $\tilde{\mathcal{B}}_{L,i}$ and $\tilde{\mathcal{B}}_{U,i}$, Fig. 6.15, and for the calculations of $\mathcal{B}_{L,i}$ and $\mathcal{B}_{U,i}$, Fig. 6.16 for the MC, and Fig. 6.17 for the real data. The summary of the results involve the generator level values of the $\mathcal{B}_L$ and $\mathcal{B}_U$ implemented in the SLD Monte Carlo, Table 6.4, which similarly as for the electrons have to be corrected for the measured fragmentation function and the B decay models, to get the correct values of the scaling factors between the entire and the visible portions of the spectra.

For the real data entries from Table 6.4 one has

$$\begin{aligned} \mathcal{B}_L &= 1.249 \times (0.1010 \pm 0.0030) = 0.1261 \pm 0.0037\,, \\ \mathcal{B}_U \equiv \mathcal{B}(b{\to}c{\to}\mu) &= 1.732 \times (0.0468 \pm 0.0027) = 0.0811 \pm 0.0047\,. \end{aligned} \tag{6.3}$$

As with the electrons, we cross checked the above muon results using Eq. (3.11) and Eq. (3.15). The track multiplicity and the pair charge for muons are given in Fig. 6.13 with the content of each individual bin printed out. By taking $f(L)$ and $\tilde{\beta} + \delta$ from Fig. 6.14, and estimated $\tilde{\mathcal{B}}_L$ and $\tilde{\mathcal{B}}_U$ from Table 6.4 one has



	93-98 Monte Carlo	93-98 data
n(B)	62077.8	61602
n(L,B)	7378.81	6580
n(bkg)	1543.31	1376.29
f(L)	0.118864	0.106815
1/2 f(bkg)	0.0124305	0.0111708

Figure 6.14: Yield of μ-tagged tracks for the entire 1993 - 98 data set. Open circles are the total yield, filled circles the combined $\tilde{\mathcal{B}}_L$ and $\tilde{\mathcal{B}}_U$, obtained by subtracting the total background + mis-ID $2(\tilde{\beta}+\delta)$ from $f(L)$.

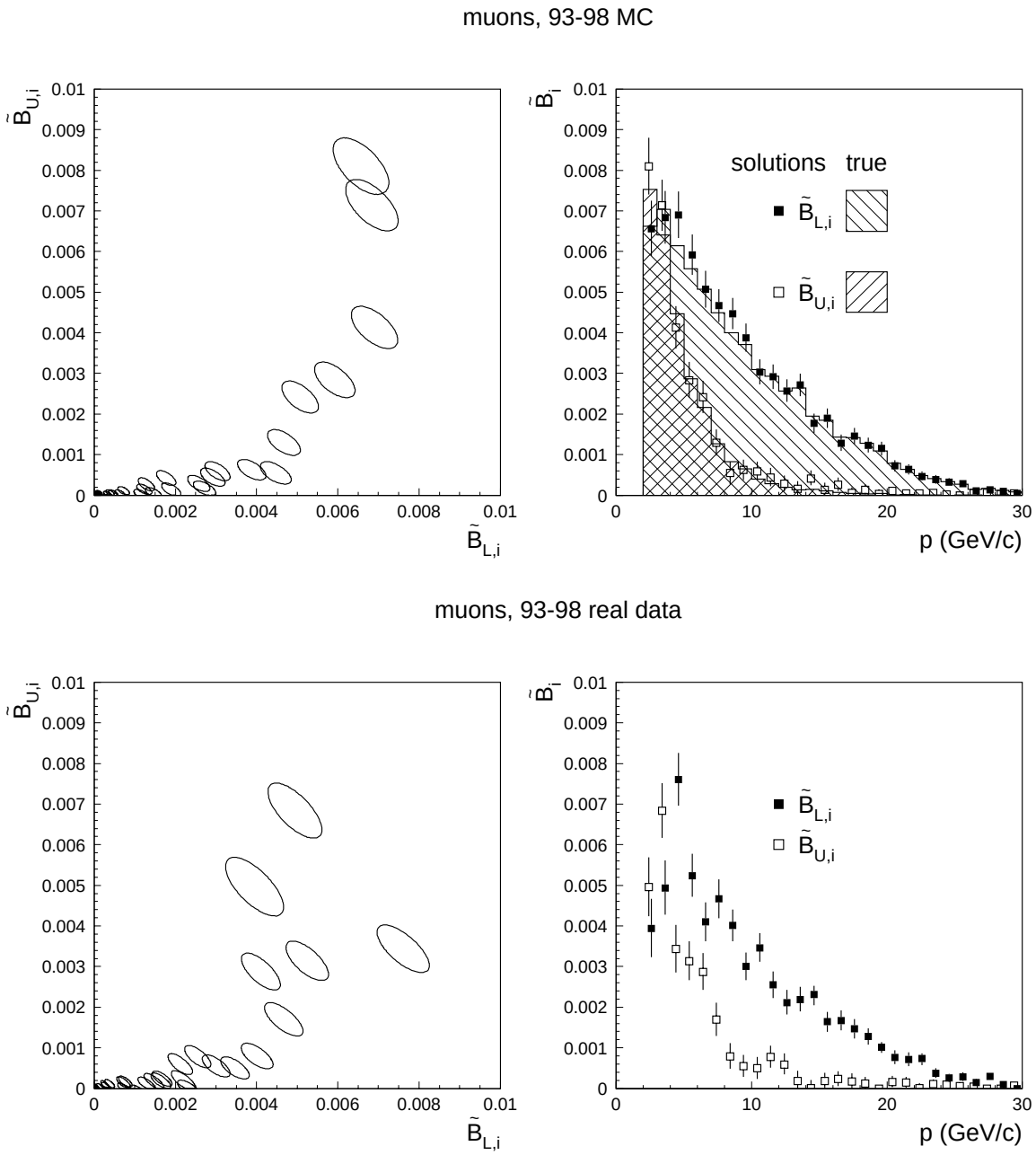


Figure 6.15: Solutions for $\tilde{\mathcal{B}}_{L,i}$ and $\tilde{\mathcal{B}}_{U,i}$ bin-by-bin in total momentum for muons, MC top row, and data bottom row. Each ellipse in the left hand plots represents 68% joint confidence region for $\tilde{\mathcal{B}}_{L,i}$ and $\tilde{\mathcal{B}}_{U,i}$ in i-th momentum bin. Projections of the confidence regions in terms of error bars are shown in the second plot with the circles being their central values. The MC histograms show true values of $\tilde{\mathcal{B}}_{L,i}$ and $\tilde{\mathcal{B}}_{U,i}$.

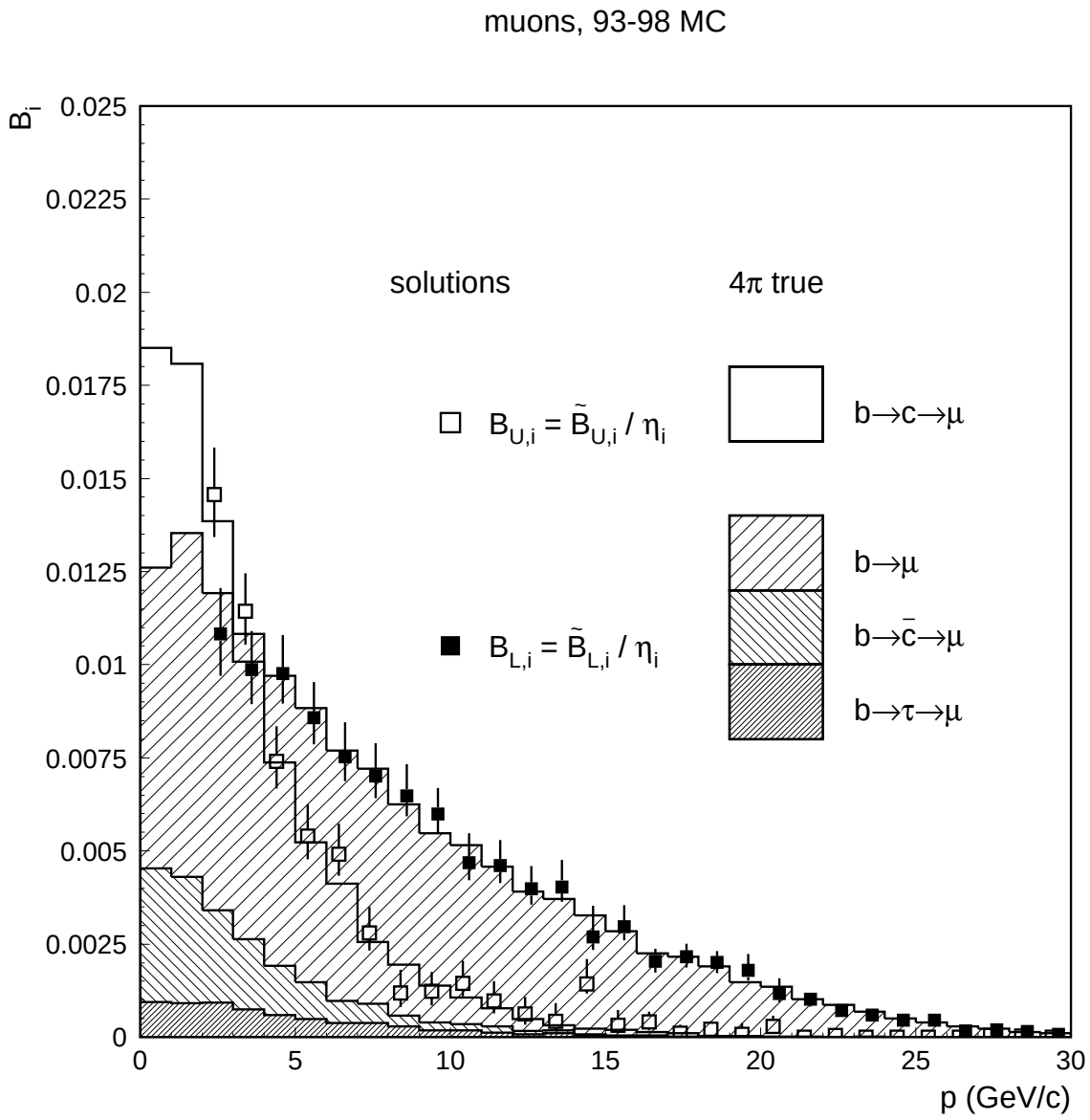


Figure 6.16: MC solutions of Fig. 6.15 scaled by the MC efficiencies of Fig. 6.4. The solutions are compared to the 4π generator level distributions.

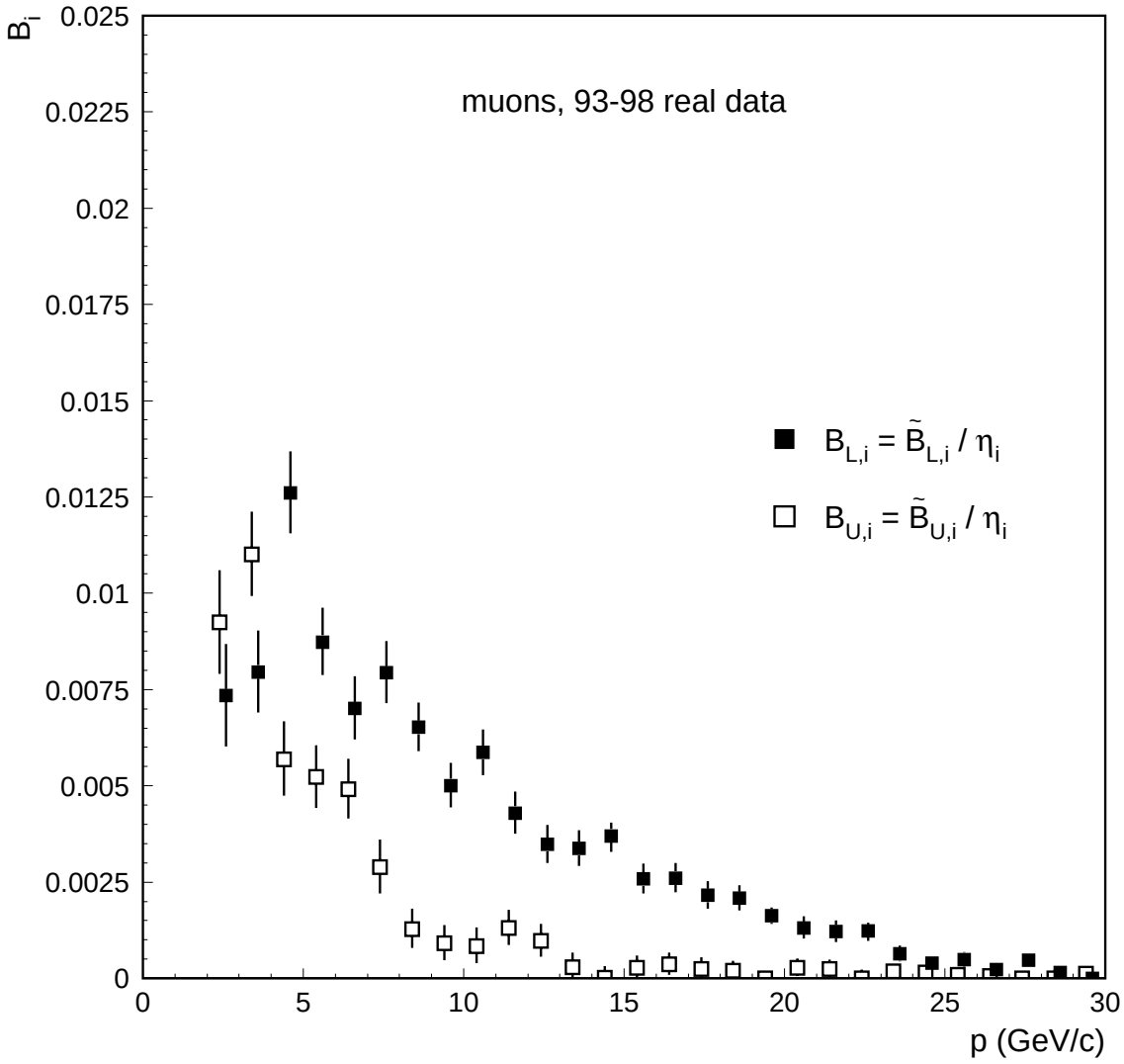
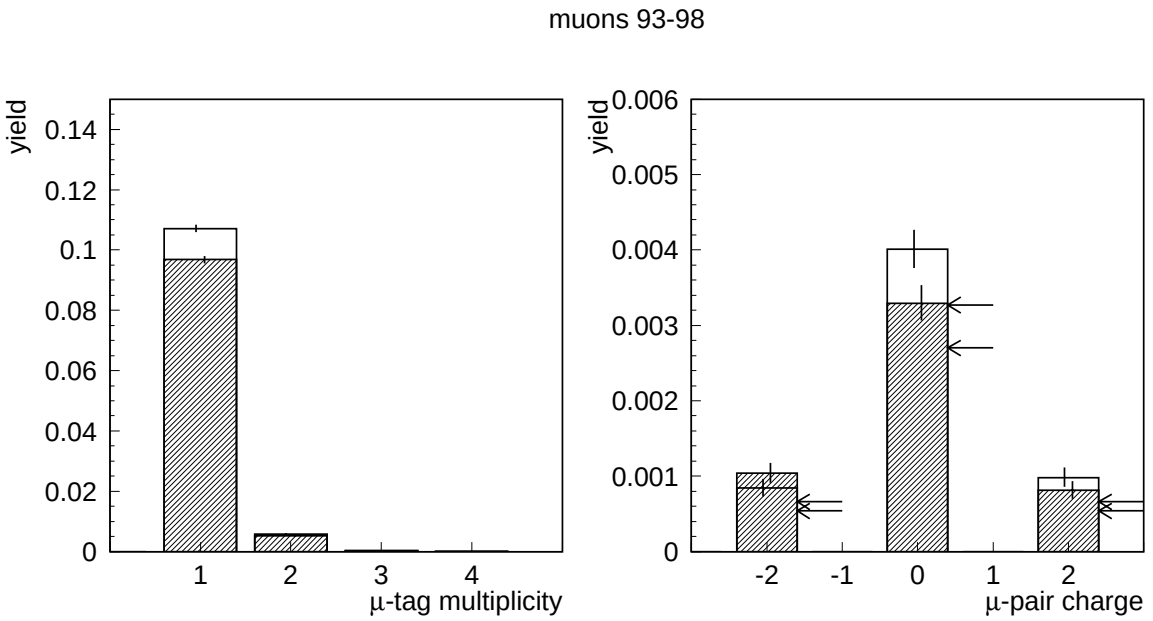


Figure 6.17: Solutions for the data of Fig. 6.15 scaled by the calibrated efficiency of Fig. 6.3.



	Monte Carlo	data		Monte Carlo	data
1	0.107095	0.0967988	f(--)	0.000843133	0.00103893
2	0.00583114	0.00514594	f(+-)	0.00400766	0.00329535
3	0.00022601	0.000211032	f(++)	0.000980347	0.000811662
4	4.30053E-05				

Figure 6.18: Multiplicity and pair charge for μ-tagged tracks. Blank histograms are for the MC; hatched histograms are for the data. The arrows indicate the values of the corresponding quantities calculated using Eq. (3.11) for unlike sign pairs and Eq. (3.15) for like sign pairs. Upper arrows are for the MC; lower arrows are for the data.

```
                    MUONS
                    -----

Solutions for the Monte Carlo in p > 2 GeV
------------------------------------------
BL~ = 0.0674815 + 0.0016921  - 0.00163728
BU~ = 0.0300416 + 0.00147476 - 0.00139752

BL  = 0.102023  + 0.0028455  - 0.00248109
BU  = 0.0553956 + 0.00288555 - 0.00212438

true  all 4 pi          4 pi in p > 2 GeV     ratio
----  --------          -----------------     -----
BL    0.130944          0.104801              1.24945
BU    0.0865618         0.0499881             1.73165

Solutions for the data in p > 2 GeV
-----------------------------------
BL~ = 0.0608721 + 0.00181    - 0.00176852
BU~ = 0.0277083 + 0.00166281 - 0.00156039

BL  = 0.101032  + 0.00304874 - 0.00297809
BU  = 0.0468316 + 0.00281297 - 0.00264976
```

Table 6.4: Summary of the main results for muons.

$$f(L)_{\pm} = f(b\bar{b})\,\tilde{\mathcal{B}}_L\tilde{\mathcal{B}}_U \;+\; f(\mathrm{L})\,(\tilde{\beta}+\delta) \;-\; (\tilde{\beta}+\delta)^2 \;=$$

$$0.964 \times 0.0675 \times 0.0300 + 0.119 \times 0.0124 - 0.0124^2 \;=\; 0.00327 \quad \text{(93-98 MC)}$$

$$0.964 \times 0.0609 \times 0.0277 + 0.107 \times 0.0112 - 0.0112^2 \;=\; 0.00270 \quad \text{(93-98 data)}$$

Note that the estimated values of $\tilde{\mathcal{B}}_L$ and $\tilde{\mathcal{B}}_U$ have been used above, and that the value of $\tilde{\beta}+\delta$ in the data is the calibrated value. For the values of $f(L)_{\pm\pm}$ one similarly gets 0.00132 and 0.00107 for the MC and the data respectively (0.00066 and 0.00054 are half of these when counted per sign). These numbers are shown with arrows in Fig. 6.18.

Absolute values of $f(L)_{\pm}$ and $f(L)_{\pm\pm}$ for muons obtained using Eq. (3.11) and Eq. (3.15) disagree with the directly observed values even more than electrons. The *ratios* between the data and the MC for the two methods of calculating the $f(L)_{\pm}$ and $f(L)_{\pm\pm}$ however,

$$\frac{f(L)_{\pm}(\text{data})}{f(L)_{\pm}(\text{MC})} = \begin{cases} \dfrac{0.00363}{0.00400} = 0.907 & \text{directly measured} \\ \dfrac{0.00270}{0.00327} = 0.826 & \text{calculated} \end{cases}$$

and

$$\frac{f(L)_{\pm\pm}(\text{data})}{f(L)_{\pm\pm}(\text{MC})} = \begin{cases} \dfrac{0.00185}{0.00182} = 1.016 & \text{directly measured} \\ \dfrac{0.00107}{0.00132} = 0.811 & \text{calculated} \end{cases}$$

are good agreement given the large errors associated with these ratios in the directly observed yields (not shown).

One source of the discrepancy between these two results is that the Eq.(3.11) and similar have been derived for an idealized topology of a back-to-back $b\bar{b}$ pair with all their decaying products contained in one hemisphere. In reality, the situation is more complicated: for lower values of x_B the topologies are more isotropic and the effects of tracks crossing into the "wrong" hemisphere becomes significant. Moreover, gluon splitting effects and their contribution to the pairs combinatorics has to be taken into account.

6.5 Inputs and systematic uncertainties

Calculation of the dominant systematic uncertainties in this analysis is quite simple. Uncertainties in $\tilde{\mathcal{B}}_L$ and $\tilde{\mathcal{B}}_U$ from the background + mis-ID probability calibration, and uncertainties in $\mathcal{B}_L = \tilde{\mathcal{B}}_L/\eta$ and $\mathcal{B}_U = \tilde{\mathcal{B}}_U/\eta$ from the efficiency calibration (which also absorbs uncertainties from the tracking efficiency and very small uncertainties from the momentum resolution), are first calculated.

This is followed by the calculations of the uncertainties from $f(b\bar{b})$ and $f(b \mid \mathrm{B})$. Corrections for the invisible portions of the spectra introduce some model related uncertainties as well as uncertainties from the fragmentation function.

At the end, we estimate the value of $\mathcal{B}(b{\to}\bar{c}{\to}l)$, which is evaluated from the $\mathcal{B}(b{\to}c{\to}l)$ (measured in our analysis as a single quantity), and from the ratios of the "wrong" to "right" sign yields of the intermediate c-hadrons in b decays.

6.5.1 Uncertainties from the background + mis-ID probability and efficiency calibrations

These are quite straightforward to calculate. Recall Eq. (3.18):

$$\tilde{\mathcal{B}}_{L,i} = \frac{g^-_{ij}\, f_{L,j}(b) - g^+_{ij}\, f_{L,j}(\bar{b})}{\det C_j}\,, \qquad \tilde{\mathcal{B}}_{U,i} = \frac{g^-_{ij}\, f_{U,j}(\bar{b}) - g^+_{ij}\, f_{U,j}(b)}{\det C_j}\,,$$

where $g^-_{ij} = k^-_{ij}/n_j - (\tilde{\beta}+\delta)_{ij}$ and $g^+_{ij} = k^+_{ij}/n_j - (\tilde{\beta}+\delta)_{ij}$. Clearly, $(\tilde{\beta}+\delta)_{ij}$ can be factored out in both expressions. Since it is uncorrelated to the initial state b tag, it does not depend on j. The consequence is that it is factored out in the solution that take the entire array of initial state b flavor probabilities. Therefore, the uncertainties from the combined mis-ID probability and background calibrations σ_i is simply obtained by inserting

$$(\tilde{\beta}+\delta)_{ij}\,(1 \pm \sigma_i)$$

into the likelihood function (3.19) and by repeating the calculations for $-\sigma_i$ (to get the upper uncertainties) and for $+\sigma_i$ (to get the lower uncertainties). Given the statistical independence of the individual bins the systematic errors from the $\tilde{\beta}+\delta$ are summed in quadrature as $\sigma = \sqrt{\sum_i \sigma_i^2}$.

Along the same line of reasoning, uncertainties in $\mathcal{B}_L = \tilde{\mathcal{B}}_L/\eta$ and $\mathcal{B}_U = \tilde{\mathcal{B}}_U/\eta$ are estimated just like uncertainty in X/Y for two mutually independent random variables X and Y, for which

$$\begin{aligned}
\frac{1}{Y} &= \frac{1}{\mu_y} - \frac{1}{\mu_y^2}(Y-\mu_y) + \frac{1}{\mu_y^3}(Y-\mu_y)^2 + \cdots \\
\frac{1}{Y^2} &= \frac{1}{\mu_y^2} - \frac{2}{\mu_y^3}(Y-\mu_y) + \frac{3}{\mu_y^4}(Y-\mu_y)^2 + \cdots \\
\sigma^2_{X/Y} \equiv E\left(\frac{X}{Y} - E\frac{X}{Y}\right)^2 &= E\frac{X^2}{Y^2} - \left(E\frac{X}{Y}\right)^2 = (\mu_x^2+\sigma_x^2)E\frac{1}{Y^2} - \mu_x^2\left(E\frac{1}{Y}\right)^2 \\
&\approx \frac{\mu_x^2+\sigma_x^2}{\mu_y^2}\left(1+3\frac{\sigma_y^2}{\mu_y^2}\right) - \frac{\mu_x^2}{\mu_y^2}\left(1+\frac{\sigma_y^2}{\mu_y^2}\right)^2 \\
&\approx \frac{1}{\mu_y^2}\left(\sigma_x^2 + \frac{\mu_x^2}{\mu_y^2}\sigma_y^2\right)
\end{aligned}$$

when higher order terms in $\sigma_x \ll \mu_x$ and $\sigma_y \ll \mu_y$ are neglected. This translates into $\tilde{\mathcal{B}}_{L,i}$, $\tilde{\beta}_i + \delta_i$, and η_i as

$$\sigma^2_{\mathcal{B}_{L,i}} = \frac{\sigma^2_{\tilde{\mathcal{B}}_{L,i}}(\text{stat}) + \sigma^2_{\tilde{\mathcal{B}}_{L,i}}(\text{from } \tilde{\beta}_i + \delta_i) + \mathcal{B}^2_{L,i}\sigma^2_{\eta_i}}{\eta_i^2} \tag{6.4}$$

and similarly for $\mathcal{B}_{U,i}$ (i is momentum bin index). Outputs of these calculations are shown in Table 6.5.

```
ELECTRONS p > 1 GeV
-------------------
               statistical  sys bkg+mis sys eff     combined
               -----------  ----------- -------     --------
BL = 0.106649 +-0.00418433+-0.00178542+-0.000537744 (+-0.004581)
BU = 0.0638867+-0.00406954+-0.00195378+-0.00052637  (+-0.00454484)


MUONS p > 2 GeV
---------------
               statistical  sys bkg+mis sys eff     combined
               -----------  ----------- -------     --------
BL = 0.101032 +-0.00301176+-0.00163957+-0.000480234 (+-0.00346259)
BU = 0.0468316+-0.00272176+-0.00140003+-0.000333969 (+-0.0030789)
```

Table 6.5: Outputs for the central values, statistical uncertainties, systematic uncertainties from $\tilde{\beta} + \delta$, and systematic uncertainties from η for both electrons and muons. These uncertainties add in quadrature as in Eq. (6.4).

6.5.2 The initial state uncertainties

These are the uncertainties in $f(b\bar{b})$ and $f(b \mid \text{B})$. The latter absorbs uncertainties from the jet charge calibration, A_e, A_b, QCD corrections to the tree level value of the A_b, χ, initial state electron polarization P_e, and a very small uncertainty from the $c\bar{c}$ asymmetry (A_c). These uncertainties are calculated for the bulk values of the $\mathcal{B}_L$ and $\mathcal{B}_U$ (their bin-by-bin values are maximally correlated: if one bin goes up or down all other bins go up or down proportionally). The number are given in Table 6.6 (we thank Su Dong for a plentiful supply of the newest data used in the systematic studies here [64]).

	MC	data	data $\sigma_{\mathcal{B}_L}(e)$	$\sigma_{\mathcal{B}_U}(e)$	$\sigma_{\mathcal{B}_L}(\mu)$	$\sigma_{\mathcal{B}_U}(\mu)$
A_e	0.1382	0.1496 ± 0.0016	± 0.000099	∓ 0.000120	$\pm 0.000070*$	$\pm 0.000068*$
A_b	0.897	0.905 ± 0.030	∓ 0.000607	± 0.000812	∓ 0.000632	± 0.000536
χ	0.1300	0.1186 ± 0.0043	± 0.000158	∓ 0.000158	± 0.000305	∓ 0.000305
P_e	± 1	see Fig. 6.19	∓ 0.000191	± 0.000193	$\pm 0.000225*$	$\mp 0.000278*$
		data combined	± 0.000663	± 0.000858	± 0.000740	± 0.000680

```
el BL      Ae           Ab          chi           Pe
           --           --          ---           --
h       0.106748     0.106359    0.106807      0.106716
c       0.106649     0.106649    0.106649      0.106649
l       0.106655     0.107256    0.106494      0.10684

el BU      Ae           Ab          chi           Pe
           --           --          ---           --
h       0.0637662    0.0643888   0.063729      0.0637859
c       0.0638867    0.0638867   0.0638867     0.0638867
l       0.0639009    0.0630748   0.0640405     0.0636936

mu BL      Ae           Ab          chi           Pe
           --           --          ---           --
h       0.100962*    0.1004       0.101306     0.101012*
c       0.101032     0.101032     0.101032     0.101032
l       0.101035*    0.101373     0.100727     0.100807*

mu BU      Ae           Ab          chi           Pe
           --           --          ---           --
h       0.0468997*   0.0473678   0.0465578     0.0468432*
c       0.0468316    0.0468316   0.0468316     0.0468316
l       0.0468327*   0.0466219   0.0471364     0.0471101*
```

Table 6.6: Uncertainties from the initial state inputs for the *visible* portions of the spectra ($p > 1$ GeV for electrons and $p > 2$ GeV for muons). Suspicious numbers (although quite insignificant) are marked by an $*$.

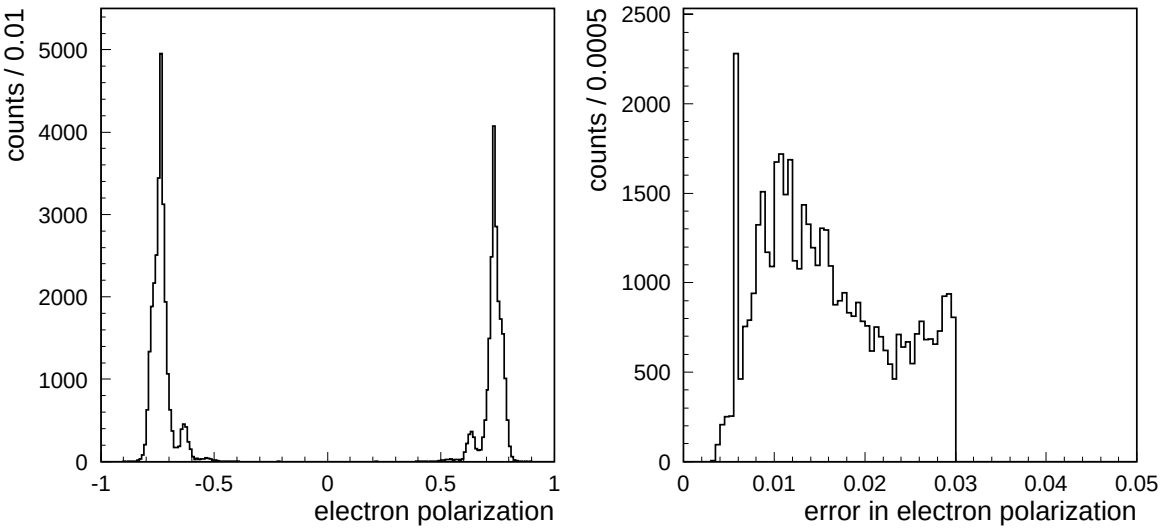


Figure 6.19: Polarization of incident electrons for 1993-98 data. Two lower peaks are from early data taking (1993).

6.5.3 Open charm multiplicities and their flavor-specific ratios

For $b \equiv B^-/\bar{B}^0/\bar{B}^0_s/b$-baryon admixture at the Z, and $D \equiv D^+/D^0/D^+_s/c$-baryon admixture in b hadron decay products, the most convenient relation for our purpose would be

$$r_D \equiv \frac{\mathcal{B}(b\to\bar{D}X)}{\mathcal{B}(b\to DX)} \approx \frac{\mathcal{B}(b\to\bar{c}\to l)}{\mathcal{B}(b\to c\to l)} \tag{6.5}$$

where the second ratio is obtained by multiplying the numerator and the denominator of the first ratio with $\mathcal{B}(c\to l)$, and by assuming $\mathcal{B}(c\to l)$ for the charm composition at upper and lower vertices approximately equal. There are other ways of estimating $\mathcal{B}(b\to\bar{c}\to l)$: for example it could be obtained directly as $\mathcal{B}(b\to\bar{D}X)\cdot\mathcal{B}(c\to l)$, which requires the knowledge of two quantities as opposed to a single ratio r_D.

The r_D has been directly measured at $\Upsilon(4S)$:

$$r_D \equiv \frac{\mathcal{B}(b\to\bar{D}X)}{\mathcal{B}(b\to DX)} = 0.100 \pm 0.026 \pm 0.016 \ \ (\pm 0.031 \ \text{ combined}), \tag{6.6}$$

for $b \equiv B^-/\bar{B}^0$ and $D \equiv D^+/D^0$ admixture, [22] (D instead of c to suggest that only open charm is counted). Open charm multiplicity is given by

$$n_D \;=\; 1\cdot\mathcal{B}_{\bar D}(1-\mathcal{B}_D)+1\cdot(1-\mathcal{B}_{\bar D})\mathcal{B}_D+2\cdot\mathcal{B}_{\bar D}\mathcal{B}_D=\mathcal{B}_{\bar D}+\mathcal{B}_D$$

where $\mathcal{B}_{\bar D}\equiv\mathcal{B}(b{\to}\bar DX)$ and $\mathcal{B}_D\equiv\mathcal{B}(b{\to}DX)$. For a D^+/D^0 admixture [24]

$$\begin{aligned} n_D &= \underbrace{53.4\pm2.7\pm3.1(\pm1.8)\%}_{n_{D^0,\bar D^0}}+\underbrace{18.8\pm1.5\pm1.3(\pm1.2)\%}_{n_{D^\pm}} \\ &= 72.2\pm5.1\% \end{aligned}$$

For the fractions of $(D^0,\bar D^0)$ and $D^\pm$ are at the Z (not at $\Upsilon(4S)$) one calculates

$$\mathcal{B}(b{\to}\bar DX)=n_D\,\frac{r_D}{1+r_D}=0.066\pm0.023\,. \tag{6.7}$$

[The second moment of $Y=X/(1+X)$ is given by $\sigma_y^2=\sigma_x^2/(1+\mu_x)^4$, when higher order terms in σ_x are neglected, and is obtained from the Taylor expansion of Y in $X-\mu_x$. Also $\sigma_{zy}^2=\sigma_z^2\,y^2+z^2\,\sigma_y^2$ as usual. This translates into a second moment of (6.7) as $\sigma^2=n_D^2\sigma_{r_D}^2/(1+r_D)^4+\sigma_{n_D}^2r_D^2/(1+r_D)^2$.]

We first want to check whether (6.7) agrees with similar measurements at the Z. In a recent paper by ALEPH [23],

$$p_D\equiv\frac{\mathcal{B}(b{\to}D\bar DX)}{\mathcal{B}(b{\to}DX)}=(7.8^{+2.0}_{-1.8}\;(\text{stat})^{+1.7}_{-1.5}\;(\text{syst})^{+0.5}_{-0.4}\;(\mathcal{B}_D))\% \tag{6.8}$$

has been reported, for $b\equiv B^-/\bar B^0/\bar B_s^0/b$-baryon and $D\equiv D^+/\bar D^0$ admixture. [ALEPH calls ratio in (6.8) just $\mathcal{B}(b{\to}D\bar DX)$ which is somewhat confusing because what is actually measured is $\mathcal{B}(b{\to}D\bar DX)$ relative to the reference subset of all $b{\to}DX$ decays. We will sometimes use $f(b{\to}\bar DX\mid b{\to}DX)$ for the ratio in (6.8) and other similar quantities to stress that they are conditional probabilities, and to also bring many different notations in accord with one accepted in this document.]

At the Z, the $D\equiv D^+/D^0$ admixture has to be augmented with D_s and Λ_c modes:

$$\begin{aligned}\mathcal{B}(b\to\bar{D}X) &= f(b\to\bar{D}X \mid b\to DX)\mathcal{B}(b\to DX) + f(b\to\bar{D}X \mid b\to D_s^+X)\mathcal{B}(b\to D_s^+X) \\ &\quad + f(b\to\bar{D}X \mid b\to\Lambda_c^+X)\mathcal{B}(b\to\Lambda_c^+X)\,. \qquad (6.9)\end{aligned}$$

Assuming $f(b\to\bar{D}X \mid b\to DX)$, $f(b\to\bar{D}X \mid b\to D_s^+X)$, and $f(b\to\bar{D}X \mid b\to\Lambda_c^+X)$ approximately equal, and by taking into account that the three semileptonic branching fractions in (6.9) sum to 1, one has

$$p_D \approx \mathcal{B}(b\to\bar{D}X)\,,$$

for p_D from Eq. (6.8) and $\mathcal{B}(b\to\bar{D}X)$ from Eq. (6.9) respectively. Given the size of errors in (6.7) and (6.8), CLEO and ALEPH results agree quite well. Based on these comparisons, we take for granted the "bulk" value [23]

$$\mathcal{B}(b\to\bar{D}X) \approx p_D \equiv \frac{\mathcal{B}(b\to D\bar{D}X)}{\mathcal{B}(b\to DX)} = (20.9^{+3.2}_{-2.8}\ (\text{stat})^{+2.5}_{-2.2}\ (\text{syst})^{+4.5}_{-2.8}\ (\mathcal{B}_D))\%\,, \qquad (6.10)$$

where $\bar{D} \equiv D^-/\bar{D}^0/D_s^-$ is upper vertex charm admixture and $D \equiv D^+/D^0/D_s^+$ lower vertex charm admixture. Along the same line of reasoning as that which established that p_D in Eq. (6.8) and $\mathcal{B}(b\to\bar{D}X)$ in Eq. (6.9) are approximately equal, and by taking into account that [20]

$$r_{\Lambda_c} = \frac{\mathcal{B}(b\to\bar{\Lambda}_cX)}{\mathcal{B}(b\to\Lambda_cX)} = 0.20 \pm 0.14 \qquad (6.11)$$

is approximately the same as the value of p_D in Eq. (6.10), one can take the p_D in Eq. (6.10) to be the same for the totally inclusive $D \equiv D^+/D^0/D_s^+/\Lambda_c$ lower vertex admixture, and for the same (conjugate) admixture at the upper vertex. From the inclusive open charm multiplicity [24]

$$\begin{aligned}n_D &= n_{D^0,\bar{D}^0} + n_{D^\pm} + n_{D_s^\pm} + n_{\Lambda_c^\pm} \equiv \mathcal{B}(b\to\bar{D}X) + \mathcal{B}(b\to DX) \\ &= (106.1 \pm 4.5 \pm 6.0\,(\pm 1.7))\%\ (\pm 7.7\%\ \text{combined})\,,\end{aligned}$$

and from the (6.10) we calculate

$$r_D \equiv \frac{\mathcal{B}(b\to\bar{D}X)}{\mathcal{B}(b\to DX)} = \frac{p_D}{n_D - p_D} = 0.245 \pm 0.056\,.$$

Since $r_D \approx \mathcal{B}(b\to\bar{c}\to l)/\mathcal{B}(b\to c\to l)$, Eq. (6.5), one has

$$\mathcal{B}(b\to\bar{c}\to l) = r_D\,\mathcal{B}(b\to c\to l) = (0.245 \pm 0.056) \times \mathcal{B}(b\to c\to l)\,, \tag{6.12}$$

with the $\mathcal{B}(b\to c\to l)$ directly measured in our analysis.

The disadvantage in our approach of calculating the $\mathcal{B}(b\to\bar{c}\to l)$ is that a relatively large error from the $\mathcal{B}(b\to c\to l)$ creeps in. The advantage compared to the calculation of the $\mathcal{B}(b\to\bar{c}\to l)$ as a product of p_D and $\mathcal{B}(c\to l)$, on the other hand, is that the PDG value of $\mathcal{B}(c\to l)$ refers to the $Z\to c\bar{c}$ charm composition and does not properly reflect the charm composition in the b decay products, which contain larger fraction of D_s final states, etc.

[**Note:** All of the above $b\to D$ branching fractions already include all known strong intermediate decays via excited D^* states.]

The value of $\mathcal{B}(b\to\tau\to l)$ is much simpler to calculate: to order $1/m_b^2$ one has [26]

$$r_\tau \equiv \frac{\mathcal{B}(b\to c\tau\bar{\nu})}{\mathcal{B}(b\to ce\bar{\nu})} = 0.22 \pm 0.02\,. \tag{6.13}$$

The $\mathcal{B}(\tau\to l)$ is accurately known to within $0.4 - 0.5\%$ [7]:

$$\begin{aligned}\mathcal{B}(\tau\to e^-\bar{\nu}_e\nu_\tau) &= (17.81 \pm 0.07)\%\\ \mathcal{B}(\tau\to\mu^-\bar{\nu}_\mu\nu_\tau) &= (17.37 \pm 0.09)\%\,.\end{aligned}$$

Therefore

$$\mathcal{B}(b\to\tau\to l) = r_\tau\,\mathcal{B}(\tau\to l)\,\mathcal{B}(b\to l) = (0.0385 \pm 0.0039) \times \mathcal{B}(b\to l)\,. \tag{6.14}$$

6.5.4 The $\mathcal{B}(b\to c\to l)$ composition

Inclusive $\mathcal{B}(b\to c\to l)$ can be written as a product of the branching fractions of the contributing exclusive subprocesses

$$\begin{aligned}\mathcal{B}(b\to c\to l) \;=\; & (f_{B^-}\, f_{B^-\to D^0} + f_{\bar{B}^0}\, f_{\bar{B}^0\to D^0} + f_{\bar{B}^0_s}\, f_{\bar{B}^0_s\to D^0} + f_{\Lambda_b}\, f_{\Lambda_b\to D^0})\, \mathcal{B}(D^0\to \bar{l}\nu_l X) \\ & + \text{modes with } D^+,\, D_s, \text{ and } \Lambda_c \text{ intermediate states.}\end{aligned} \tag{6.15}$$

where f_{B^-}, $f_{B^-\to D^0}$, $\cdots$ are just the shorthand notations for the branching fractions. This is a lengthly and redundant expression. The $\mathcal{B}(b\to c\to l)$ is directly measured as a single quantity in our analysis and the only reason for knowing its composition is to estimate the effects of ratio r in Eq. (6.19) that is needed to get the correction of $\mathcal{B}(b\to c\to l)$ for the "invisible" portion of the spectrum.

This can be elegantly done by using a trick which not only simplifies the calculations but also yields much more accurate results. We first introduce

$$\begin{aligned} n_D(\mathrm{sl}) = \mathcal{B}(b\to c\to l) + \mathcal{B}(b\to \bar{c}\to l) = \\ \underbrace{(0.534 \pm 0.027 \pm 0.031(\pm 0.018))\ (\pm 0.045 \text{ combined})}_{n_{D^0,\bar{D}^0}}\ \mathcal{B}(D^0\to e^+\nu_e X) \\ +\ \underbrace{(0.188 \pm 0.015 \pm 0.013(\pm 0.012))\ (\pm 0.023 \text{ combined})}_{n_{D^\pm}}\ \mathcal{B}(D^+\to e^+\nu_e X) \\ +\ \underbrace{(0.214 \pm 0.023 \pm 0.021(\pm 0.025))\ (\pm 0.040 \text{ combined})}_{n_{D_s^\pm}}\ \mathcal{B}(D_s\to e^+\nu_e X) \\ +\ \underbrace{(0.125 \pm 0.024 \pm 0.010(\pm 0.017))\ (\pm 0.031 \text{ combined})}_{n_{\Lambda_c^\pm}}\ \mathcal{B}(\Lambda_c\to e^+\nu_e X) \end{aligned} \tag{6.16}$$

which is just a sum of weighted intermediate D semileptonic branching fractions with exclusive open charm multiplicities as the weighting factors, [24]. Given the accuracy of the exclusive open charm multiplicities, the role of the D decay shapes on the value of r can be simply estimated by weighting the MC D semileptonic decay shapes with the 16 different combinations of the charm multiplicities, each being varied within its error, Eq. (6.16). The results are

	lowest	highest	σ_r
r for $\mathcal{B}(b\to c\to e)$	1.31075	1.31277	0.000202
r for $\mathcal{B}(b\to c\to \mu)$	1.72872	1.73242	0.0037

(6.17)

where we took $\sigma_r = r(\text{highest}) - r(\text{lowest})$, assuming it is large enough to absorb the uncertainties from the cascade semileptonic decay shapes.

An interesting calculation can be done by inserting the PDG values for the exclusive D semileptonic branching fraction into (6.16). For [7]

$$\begin{aligned}
\mathcal{B}(D^0\to e^+\nu_e X) &= 0.0675 \pm 0.0029 \\
\mathcal{B}(D^+\to e^+\nu_e X) &= 0.172 \pm 0.019 \\
\mathcal{B}(D_s\to e^+\nu_e X) &= 0.08^{+0.06}_{-0.05} \\
\mathcal{B}(\Lambda_c\to e^+\nu_e X) &= 0.021 \pm 0.006
\end{aligned}$$

one obtains

$$n_D(\text{sl}) \approx 0.088 \qquad \text{and} \qquad \mathcal{B}(b\to c\to l) = \frac{n_D(\text{sl})}{1.245} \approx 0.071\,,$$

close to our measured value $\mathcal{B}(b\to c\to l) \approx 0.078$. This agreement gave us an additional confidence in both the numbers from already cited [24] and in our calculations.

6.5.5 The role of the fragmentation function

The effects of the fragmentation function are simple to estimate. We thank Danning Dong here for providing us with two "endpoint" distributions, f_1 and f_2, Fig. 6.20, which are extracted from the envelopes of all acceptable fragmentation functions measured at the SLD [77]. Function f in Fig. 6.20 is the standard SLD MC fragmentation function. The f_1 and f_2 roughly correspond to highest and lowest value of $\langle x_B\rangle$. We take the central value of r to be $r = (r_{f_1} + r_{f_2})/2$ and the uncertainty $\sigma_r = r - r_f$. This σ_r is large, and absorbs a number of other small uncertainties like the uncertainty in the center of mass energy of the initial e^+e^- pair, b-decay shapes, and similar.

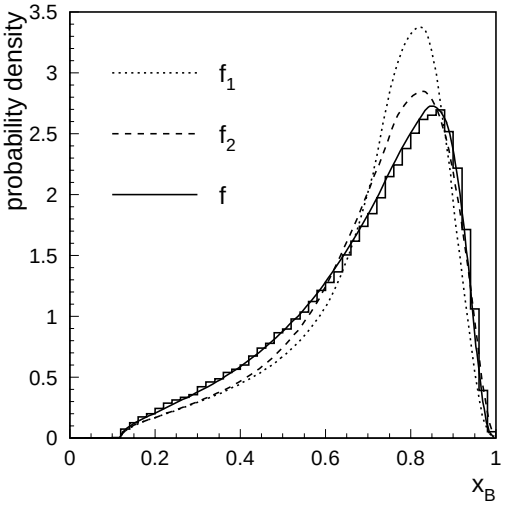


Figure 6.20: Fragmentation functions used in the studies here. The $x_B = 2E_B/E_{CM}$ is the scaled b hadron energy where E_{CM} is the estimated center of mass energy of the initial state $b\bar{b}$ pair. The three normalized distributions are: f the SLD MC distribution, f_1 and f_2 two "endpoint" distribution measured at the SLD [77]. The histogram is the cross-check of the SLD MC distribution using a large sample of true electrons and muons with B, D, or τ parents.

From Fig. 6.21 and Table 6.7 we obtain

$$\boxed{\begin{array}{rcl} r_{\mathcal{B}_L}(\text{electrons}) & = & 1.1072 \pm 0.0037 \\ r_{\mathcal{B}_U}(\text{electrons}) & = & 1.304 \pm 0.012 \\ r_{\mathcal{B}_L}(\text{muons}) & = & 1.2368 \pm 0.013 \\ r_{\mathcal{B}_U}(\text{muons}) & = & 1.699 \pm 0.036 \end{array}} \tag{6.18}$$

* * *

So far our calculations were restricted only to the visible portions of the spectra. Branching fractions for the entire momentum range are then calculated as

$$\mathcal{B}(\text{entire spectrum}) \;=\; r\,\mathcal{B}(\text{visible portion}) \tag{6.19}$$

where r is from (6.18). When the uncertainties in r and the uncertainties in the visible portions of $\mathcal{B}_L$ and $\mathcal{B}_U$ are all taken into account, one has

$$\begin{array}{rcl} \mathcal{B}_L(\text{electrons}) & = & (1.1072 \pm 0.0037) \times (0.1066 \pm 0.0046 \pm 0.00066) \\ & = & 0.1180 \pm 0.0052 \end{array}$$

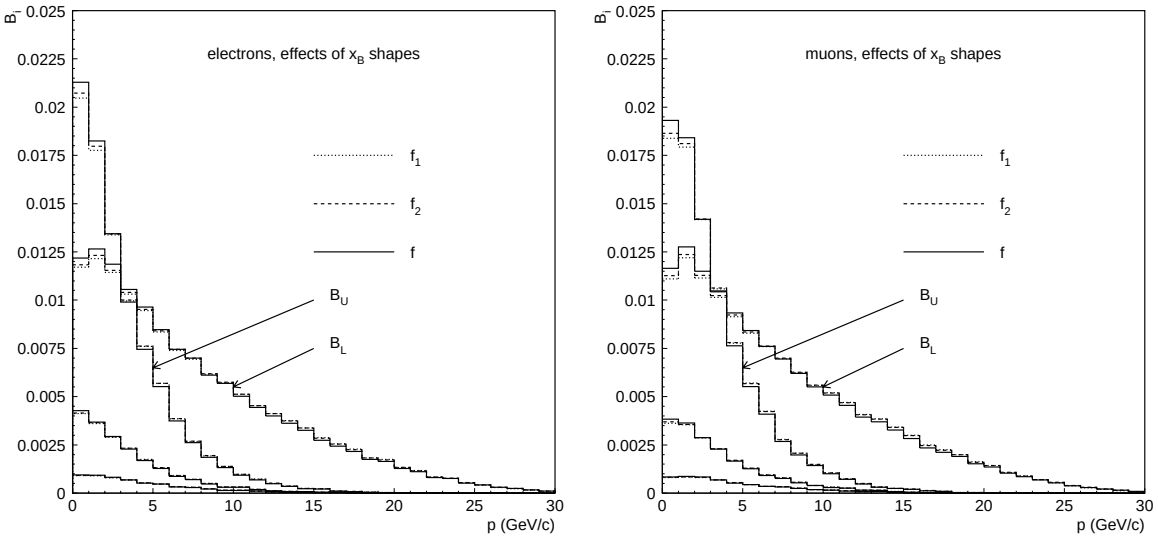


Figure 6.21: Generator level spectra for electrons (left) and muons (right) corresponding to different fragmentation functions of Fig. 6.20. Ratios r between the entire and the visible portions of the spectra are calculated based on the data summarized in Table 6.7. For electrons we obtain $r_{\mathcal{B}_L} = 1.1072 \pm 0.0037$ and $r_{\mathcal{B}_U} = 1.304 \pm 0.012$, and for muons $r_{\mathcal{B}_L} = 1.2368 \pm 0.013$ and $r_{\mathcal{B}_U} = 1.699 \pm 0.036$.

	ELECTRONS	MUONS
	r = BL(1:30)/BL(2:30)	r = BL(1:30)/BL(3:30)
f1	1.10696	1.23542
f2	1.10746	1.2381
f	1.11088	1.24945
	r = BU(1:30)/BU(2:30)	r = BL(1:30)/BL(3:30)
f1	1.30281	1.69571
f2	1.30475	1.70235
f	1.31618	1.73552

Table 6.7: Outputs corresponding to Fig. 6.21. We use $r = (r_{f_1} + r_{f_2})/2$ as an estimate of the central value of r, and $\sigma_r = r - r_f$ as an estimate of its error.

$$\begin{aligned}
\mathcal{B}_U(\text{electrons}) &= (1.304 \pm 0.012 \pm 0.0002) \times (0.0639 \pm 0.0045 \pm 0.00086) \\
&= 0.0833 \pm 0.0060 \\
\\
\mathcal{B}_L(\text{muons}) &= (1.2368 \pm 0.013) \times (0.1010 \pm 0.0035 \pm 0.00074) \\
&= 0.1249 \pm 0.0046 \\
\\
\mathcal{B}_U(\text{muons}) &= (1.699 \pm 0.036 \pm 0.0037) \times (0.0468 \pm 0.0031 \pm 0.00068) \\
&= 0.0795 \pm 0.0057
\end{aligned} \tag{6.20}$$

where the first error in r is from the fragmentation function, and the second from the cascade D semileptonic decay shapes and compositions (applies only to $\mathcal{B}_U$). The first error in the visible portions in $\mathcal{B}_L$ and $\mathcal{B}_U$ is from the lepton background + mis-ID and efficiency combined, Table 6.5. The second error is from the initial state inputs, Table 6.6.

6.5.6 Getting final results for $\mathcal{B}(b\to l)$ and $\mathcal{B}(b\to c\to l)$

Correction factors for the biases in the estimators of the $\mathcal{B}_L$ and $\mathcal{B}_U$ are obtained by comparing the 4π generator level values for $p > 1$ GeV for electrons, Table 6.3, and $p > 2$ GeV for muons, Table 6.4, with the corresponding values of the solutions (in the same tables):

$$\boxed{\begin{aligned}
c_{\mathcal{B}_L}(\text{electrons}) &= 1.003 \\
c_{\mathcal{B}_U}(\text{electrons}) &= 0.973 \\
c_{\mathcal{B}_L}(\text{muons}) &= 1.027 \\
c_{\mathcal{B}_U}(\text{muons}) &= 0.902
\end{aligned}}$$

The corrected values for $\mathcal{B}_L$ and $\mathcal{B}_U \equiv \mathcal{B}(b\to c\to l)$ are

$$\boxed{\begin{aligned}
\mathcal{B}_L(\text{electrons}) &= 0.1184 \pm 0.0055 \\
\mathcal{B}_U(\text{electrons}) &= 0.0811 \pm 0.0061 \\
\mathcal{B}_L(\text{muons}) &= 0.1283 \pm 0.0049 \\
\mathcal{B}_U(\text{muons}) &= 0.0717 \pm 0.0051
\end{aligned}}$$

where the uncertainties have been additionally scaled as $\sigma\to\sigma\sqrt{1+1/n}$ to take into account the error in the bias estimation. The $n = 10$ is the number of MC samples equivalent in size to a single data sample.

Then

$$\begin{aligned}\mathcal{B}_L &= \mathcal{B}(b\to l) + \mathcal{B}(b\to\bar{c}\to l) + \mathcal{B}(b\to\tau\to l)\\ &= \mathcal{B}(b\to l) + r_D\,\mathcal{B}(b\to c\to l) + r_\tau\,\mathcal{B}(\tau\to l)\,\mathcal{B}(b\to l)\,.\end{aligned}$$

For r_D from (6.12) and $\alpha \equiv r_\tau\,\mathcal{B}(\tau\to l)$ from (6.14) one has

$$\begin{aligned}\mathcal{B}(b\to l) &= \frac{\mathcal{B}_L - r_D\,\mathcal{B}_U}{1 + r_\tau\,\mathcal{B}(\tau\to l)} = \frac{\mathcal{B}_L - (0.245 \pm 0.056)\times\mathcal{B}_U}{1 + (0.0385 \pm 0.0039)} =\\ &\quad 0.0948 \pm 0.0065 \quad \text{(for electrons)}\,,\\ &\quad 0.1010 \pm 0.0060 \quad \text{(for muons)}\,,\end{aligned}$$

where the error has been calculated in quadratic approximation:

$$\sigma^2_{\mathcal{B}(b\to l)} = \frac{1}{(1+\alpha)^2}\left\{\left(\sigma^2_{\mathcal{B}_L} + r_D^2\sigma^2_{\mathcal{B}_U} + \sigma^2_{r_D}\mathcal{B}_U^2\right) + \sigma^2_\alpha\left(\frac{\mathcal{B}_L - r_D\mathcal{B}_U}{1+\alpha}\right)^2\right\}\,.$$

The final results are

$$\begin{aligned}\mathcal{B}(b\to e) &= 0.0949 \pm 0.0070\\ \mathcal{B}(b\to\mu) &= 0.1066 \pm 0.0062\\ \text{combined}\quad \mathcal{B}(b\to l) &= 0.1015 \pm 0.0046\\ \\ \mathcal{B}(b\to c\to e) &= 0.0811 \pm 0.0061\\ \mathcal{B}(b\to c\to\mu) &= 0.0717 \pm 0.0051\\ \text{combined}\quad \mathcal{B}(b\to c\to l) &= 0.0756 \pm 0.0039\,.\end{aligned}$$

where the combined results have again been obtained as

source	$\sigma_{\mathcal{B}(b\to l)}$ electrons	$\sigma_{\mathcal{B}(b\to l)}$ muons	$\sigma_{\mathcal{B}(b\to c\to l)}$ electrons	$\sigma_{\mathcal{B}(b\to c\to l)}$ muons
bkg + mis-ID calibr	± 0.00209	± 0.00203	± 0.00216	± 0.00240
eff calibr	± 0.00062	± 0.00059	± 0.00063	± 0.00057
prompt/cascade eff corr	NE	NE	NE	NE
tracking eff calibr	NE	NE	NE	NE
A_e	± 0.000117	$\mp 0.000092*$	∓ 0.000158	$\pm 0.000116*$
A_b	∓ 0.000728	∓ 0.000821	± 0.001069	± 0.00092
χ	± 0.000184	± 0.000402	± 0.000208	∓ 0.000523
P_e	∓ 0.000222	$\pm 0.000304*$	± 0.000254	$\mp 0.000427*$
jet charge calibr	NE	NE	NE	NE
QCD corr to A_b	NE	NE	NE	NE
$f(b\bar{b})$ calibr	NE	NE	NE	NE
$c\bar{c}$ asymmetry	NE	NE	NE	NE
r from fragm func	± 0.00039	± 0.0013	± 0.00070	± 0.0015
r from B decay models	NE	NE	± 0.000011	± 0.00016
τ cascades	0.00037	0.00039	–	–
n_D and p_D combined	0.0039	0.0039	–	–

Table 6.8: Summary of the systematic uncertainties. Uncertainties from combined open charm multiplicity n_D and open charm flavor ratio r_D dominate the total systematic uncertainty in $\mathcal{B}(b\to l)$. Of all the systematic effects that were not estimated (NE) only the tracking efficiency has a potential of changing the central values somewhat.

$$z = \frac{\sigma_y^2}{\sigma_x^2+\sigma_y^2}\,x + \frac{\sigma_x^2}{\sigma_x^2+\sigma_y^2}\,y\,,$$
$$\frac{1}{\sigma_z^2} = \frac{1}{\sigma_x^2} + \frac{1}{\sigma_y^2}\,.$$

6.6 Summary of the results

We conclude this analysis with the summaries of the systematic uncertainties, Table 6.8, and the final results, Fig. 6.22. Following a firmly established tradition, we disentangle the statistical and the systematic errors for the purpose of this presentation. [This separation is rather artificial: we call statistical errors terms that contain

uncertainties from the counting, scaled by the central values of other variables. All other uncertainties are called systematic.]

6.7 Prospects

The analysis can be improved and extended in may different ways.

6.7.1 Immediate improvements to the existing analysis

Substantial reduction of errors in $\mathcal{B}(b\to l)$ and $\mathcal{B}(b\to c\to l)$ and removal of potential biases can be achieved within the context of the existing method, by increasing the efficiencies and the purities, and by decreasing the mis-ID rates of the ingredients in our probability relations.

- Calibration of the tracking efficiency. This can be done very accurately by using two independent tests for a track hypothesis: one that uses only VXD and another that uses only CDC. The approach has been discussed briefly at the end of Ch. 2. The task *is* requiring, however, which mostly has to do with the fact that a track hypothesis has many attributes, and can rarely be formulated as a simple hypothesis. The lack of the calibrated efficiency is the main shortcoming of our results, which may prove to be uniformly biased over the entire momentum range.

- Increasing particle identification efficiencies. The electron tests use almost all variables available, and there is very little room for the improvements. The muon efficiency, however, can be increased by at least 20% by using more discriminating WIC variables [69], and by considering momentum dependencies. Since the branching fractions are calculated as $\mathcal{B}_L = \tilde{\mathcal{B}}_L/\eta$ and $\mathcal{B}_U = \tilde{\mathcal{B}}_U/\eta$, any correction to the efficiencies η would be simple to implement.

- Efficiency of the $b\bar{b}$ event selection can be substantially increased by using likelihood ratio tests similar to the lepton tests of Ch. 4. Alternative methods based on Neural Networks have been recently experimented with at the SLD (Tom Wright), which improve the efficiency versus purity of the $b\bar{b}$ identification from about $35\% : 96.4\%$ to about $55\% : 97\%$. The results were not implemented into our analysis by the time of writing of this document.

- The average values of $f(b \mid \mathrm{B})$ and $f(\bar{b} \mid \mathrm{B})$ can substantially be increased in two ways. The first is to improve the jet charge, by actually taking into account not all the tracks in the hemisphere, but only tracks from the jet which contains the B decay vertex. This way neutral gluon jets are eliminated. What

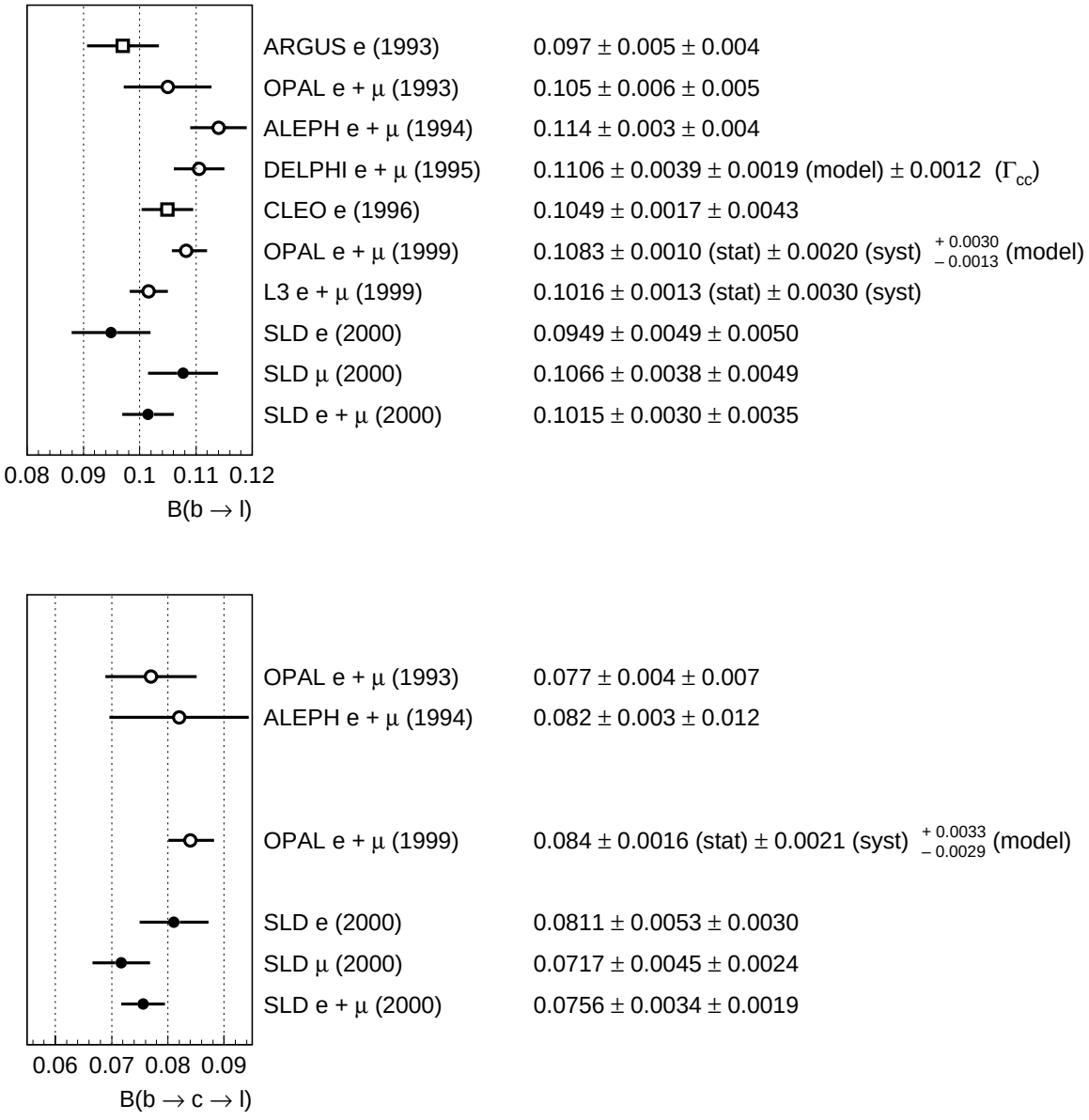


Figure 6.22: Comparison of the SLD results with other measurements. CLEO and ARGUS measurements are at $\Upsilon(4S)$, all other measurements are at the Z. CLEO also reports $\mathcal{B}(b{\to}c{\to}e)$ of $(7.8 \pm 0.2 \pm 1.2)\%$ and $(8.3 \pm 0.2 \pm 1.2)\%$, depending on the B decay model used to extract the number [2].

we have now is actually *hemisphere* and not jet charge. Another improvements to the $f(b \mid \mathrm{B})$ and $f(\bar{b} \mid \mathrm{B})$ can be achieved by using lepton and kaon tags in addition to the jet charge and polarized asymmetry. Substantial improvements in that direction have been recently achieved (Thomas Moore), but are also not implemented into the analysis at the time of writing of this document.

6.7.2 Possible expansions of the analysis

In principle, a number of quantities that are inputs to our analysis, like χ, A_b, or $\mathcal{B}(b{\to}D\bar{D}X)$, can be turned into independent variables and solved for in extended measurements. In practice however, formulating analysis that is self-contained (does not need many inputs) and simple at the same time, is not always straightforward.

The problem is rather generic in nature, and boils down to handling hierarchical statistical models that involve distribution in the sample spaces spanned by a large number of variables. The Neural Networks recently emerged as a tool of choice for handling such problems. For example, if a sample space is spanned by track variables: p, $p_\perp$, 3D impact parameter, etc., then $f(\mathbf{x} \mid b{\to}l)$ and $f(\mathbf{x} \mid b{\to}c{\to}l)$ taken from the existing B decay models, would serve as the likelihood functions for the $b{\to}l$ and $b{\to}c{\to}l$ hypotheses. By combining these distributions with the distributions that use detector variables for the lepton hypotheses, Ch. 4, one can, at least in principle, achieve high lepton identification efficiencies and accurate calibrations at the same time.

Some of the extensions of our analysis are quite natural. For example, the integrated mixing probability in the lepton hemisphere can be determined on event by event basis to increase the probabilities of the parent b or $\bar{b}$ quarks. The only problem in that kind of analysis would be high correlations between the final state lepton yield and any of the quantities used to determine the mixing probabilities.

A very natural extension of our analysis would be to simultaneously measure A_b and the two branching fractions. The A_b is input to our analysis, Eq. (5.2). On the other hand, accurately known final state (with high purity and preferably high efficiency) can be used to tag the b flavor (up to the mixing) and therefore calculate A_b; which can be inserted back into Eq. (5.2), and so on until a fixed point is reached. In practice, an extended probabilistic model can easily be formulated to simultaneously solve for A_b, χ, $\mathcal{B}(b{\to}l)$, and $\mathcal{B}(b{\to}c{\to}l)$.

References

[1] I. Bigi, B. Block, M. Shifman, A. Vainshtein, "The baffling semileptonic branching ratio of B mesons", Phys. Lett. **B323**: 408-416, 1994.

[2] CLEO collaboration, "Measurement of the B Semileptonic Branching Fraction with Lepton Tag", Phys. Rev. Lett., **76** 1570 (1996);
ARGUS Collaboration, "A Model Independent Determination of the Inclusive Semileptonic Decay Fractions of B Mesons", Phys. Lett., **B318** 397-404 (1993).

[3] The ALEPH Collaboration, "Heavy Flavor Production and Decay with Prompt Leptons in the ALEPH Detector", Z. Phys. **C62**:179-198,1994;
DELPHI Collaboration, "Measurement of $\Gamma_{b\bar{b}}/\Gamma_{had}$ using Impact Parameter Measurements and Lepton Identification", Z. Phys. **C66**:323-340,1995;
The OPAL Collaboration, "Measurement of $\mathrm{B^0-\bar{B}^0}$ mixing, $\Gamma(\mathrm{Z^0 \to b\bar{b}})/\Gamma(\mathrm{Z^0 \to}$ hadrons) and semileptonic branching ratios for b-flavoured hadrons in hadronic $\mathrm{Z^0}$ decays", Z. Phys. **C60**:199-216,1993;
The OPAL Collaboration, "Measurement of inclusive semileptonic branching fractions of b hadrons in $\mathrm{Z^0}$ decays", hep-ex/9906041, Submitted to Eur. Phys. J. **C**;
The L3 Collaboration, "Measurement of R_b and Br($b \to l\nu X$ at LEP Using Double-Tag Methods", CERN-EP/99-121 and hep-ex/9909045, Submitted to Eur. Phys. J.

[4] N. Isgur and M. B. Wise, "Weak decays of the heavy mesons in the static quark approximation", Phys. Lett. **B232**, 113 (1989);
N. Isgur and M. B. Wise, "Weak transition form-factors between heavy mesons", Phys. Lett. **B237**, 527 (1990);
M. Neubert, "Heavy Quark Symmetry", Phys. Rept. **245** (1994) 259-395.

[5] E. Bagan, P. Ball, V. M. Braun, and P. Gosdzinsky, "Charm quark masses dependence of QCD corrections to nonleptonic inclusive B decays", Nucl. Phys. **B432**:3-38, 1994;

E. Bagan, P. Ball, V. M. Braun, and P. Gosdzinsky, "Theoretical update of the semileptonic branching ratio of B mesons", Phys. Lett. **B342** (1995) 362-368; Erratum Phys. Lett. **B374** (1996) 363-364;
E. Bagan, P. Ball, B. Fiol, and P. Gosdzinsky, "Next-to-leading order radiative corrections to the decay $b \to ccs$", Phys. Lett. **B351** (1995) 546-554.

[6] M. Neubert, C. T. Sachrajda, "Spectator effects in inclusive decays of beauty hadrons", Nucl. Phys. **B483** (1997) 339-367.

[7] C. Caso et al., "Review of Particle Physics", Eur. Phys. J. **C3**, 1-794 (1998).

[8] The ALEPH Collaboration, "Measurement of the b baryon lifetime and branching fractions in Z decays", Eur. Phys. J. **C2**, 197-211 (1998);
The DELPHI Collaboration, "Measurement of the lifetime of b-baryons", Eur. Phys. J. **C10**, 185-199 (1999).

[9] D. Liko, "B hadron lifetime measurements at LEP", Nucl. Phys. **B65** (1998) 194-198 (Proc. Suppl.).

[10] C. H. Jin, W. F. Palmer, and E. A. Paschos, "A Parton Model for Inclusive Semileptonic B Meson Decays", Phys. Lett. **B329**:364-368 (1994);
C. H. Jin and E. A. Paschos, "Inclusive Semileptonic Decays and the Structure of B Mesons", hep-ph/9504375;
C. Jin, "Probing hadron structure and strong interactions with inclusive semileptonic decays of B mesons", hep-ph/9808313;
C. Jin, E. A. Paschos "Radiatively corrected semileptonic spectra in B meson decays", Eur. Phys. J. **C1**, 523-529 (1998);
C. Jin "Analysis of the hadronic invariant mass spectrum in inclusive charmless semileptonic B decays", Phys. Rev. **D57** 6851 (1998).

[11] C. Jin, "Nonperturbative QCD contributions to the semileptonic decay width of the B meson", Phys. Rev. **D56** 2928 (1997).

[12] ALEPH Collaboration, "Measurement of $|V_{cb}|$, form factors and branching fractions in the decays $\bar{B} \to D^{*+} l^- \bar{\nu}_l$ and $\bar{B} \to D^+ l^- \bar{\nu}_l$", Phys. Lett. **B395**:373-387, 1997;
CLEO Collaboration, "Measurement of the $\bar{B} \to D l \bar{\nu}$ Partial Width and Form Factor Parameters", Phys. Rev. Lett. **79** 2208 (1997).

[13] C. Jin, "Determinations of $|V_{ub}|$ and $|V_{cb}|$ from Measurement of $B \to X_{u,c} l \nu$ Differential Decay Rates".

[14] P. Renton, "Electroweak Interactions", Cambridge 1990.

[15] J. Donoghue, E. Golowich, B. Holstein, "Dynamics of the Standard Model", Cambridge 1994.

[16] B. Guberina, R. D. Peccei, and R. Rückl, "Dimensional regularization techniques and their uses in calculating infrared safe weak decay processes", Nucl. Phys. **B171** (1980) 333-361.

[17] M. Neubert, "Heavy-Quark Effective Theory and Weak Matrix Element", CERN-TH/98-2 and hep-ph/9801269.

[18] I. Dunietz, "On the Necessity of Recalibrating Heavy Flavor Decays and its Impact on Apparent Puzzles in High Energy Physics", FERMILAB-PUB-96/104-T, hep-ph/9606247.

[19] G. Buchalla, I. Dunietz, and H. Yamamoto, "Hadronization of $b \to c\bar{c}s$", Phys. Lett. **B364**, 188 (1995).

[20] I. Dunietz, J. Incandela, F. D. Snider, and H. Yamamoto, "Large charmless yield in B decays and inclusive B decay puzzles", Eur. Phys. J. **C1**, 211-219 (1998).

[21] A. F. Falk, M. B. Wise, and I. Dunietz, "Inconclusive inclusive nonleptonic B decays", Phys. Rev. **D51** 1183-1191 (1995).

[22] CLEO Collaboration, "Flavor-Specific Inclusive B Decays to Charm", Phys. Rev. Lett **80**:1150-1155, 1998.

[23] The ALEPH Collaboration, "Observation of doubly-charmed B decays at LEP", Eur. Phys. J. **C4**, 387-407 (1998).

[24] OPAL Collaboration, "A study of charm hadron production in $Z^0 \to c\bar{c}$ and $Z^0 \to b\bar{b}$ decays at LEP", Z. Phys. **C72** 1-16 (1996).

[25] K. Schubert and R. Waldi, "Semileptonic B meson decays and interfering amplitudes", hep-ph/9409341.

[26] M. Neubert, "Theory of Inclusive B Decays", hep-hp/9702310.

[27] P. Ball, "$|V_{cb}|$ from inclusive semileptonic B decays", in "QCD and Hight Energy Hadronic Interactions", Les Arcs, France, March 1995.

[28] The L3 Collaboration, "Measurement of the Average Lifetime of b-Hadrons in Z Decays", Phys. Lett. **B416**: 220-232 (1998);
DELPHI Collaboration, "Updated Precision Measurement of the Average Lifetime of B Hadrons", Phys. Lett. **B377**: 195-204 (1996).

[29] Internal notes:
Su Dong, "Summary of MC Semi-Leptonic Decay Model Branching Ratios", May 2, 1995;
Su Dong, "Summary of R15 MC Semi-Leptonic Decay Model Branching Ratios", Nov 16, 1998.

[30] J. F. Donoghue, private communication, 21 and 22 Oct 1997.

[31] R. Belusevic, "Neutral Kaons", KEK-PREPRINT-97-264, 1998.

[32] I. Caprini and M. Neubert, "Improved Bounds for the Slope and Curvature of $\bar{B}\to D^{*+}l^-\bar{\nu}_l$ Form Factors", Phys. Lett. **B380** (1996) 376-384.

[33] Downloaded from
`http://slac.stanford.edu/exp/sld/figure/intro.html`.

[34] J. Kent et al., SLAC-PUB-4922 (1989).

[35] P. C. Rowson, R. Frey, S. Hertzbach, R. Kofler, M. Swartz, and M. Woods, "Calibration of the WISRD Energy Spectrometer with a Z Peak Scan", SLD note **264** (1999).

[36] A. W. Weidemann, private communication, 26 Jan 1999;
P. Chen, A. Spitkovsky, and A. W. Weidemann, "Beam-beam disruption and the case for a plasma lens in e^-e^- collisions", Int. J. Mod. Phys. **A11**:1687-1692, 1996.

[37] S. C. Berridge et al., "Beam Test of the SLD Silicon-Tungstren Luminosity Monitor", IEEE Trans. Nucl. Sc. **NS-37**, 1191 (1990).

[38] P. Raimondi et al., "Recent Luminosity Improvements at SLD", SLAC-PUB-7995, 1998.

[39] A. Lath and M. Woods, "Compton Laser Analysis: Determination of Polarization and Systematic Errors", SLD Note **236** (1994);
J. Fernandez, "Determination of the Compton Plarimeter Laser Polariztion for the 96 SLD Run", SLD Note **257** (1997);
J. Fernandez, "Determination of the Compton Plarimeter Laser Polariztion for the 1997/98 SLD Run", SLD Note **265** (1999).

[40] B. R. Holstein, "Topics in Advanced Quantum Mechanics", Addison-Wesley, 1992.

[41] H. Veltman, "Radiative corrections to polarized Compton scattering", Phys. Rev. **D40**:2810 (1989), Erratum Phys. Rev. **D42**:1856 (1990).

[42] M. L. Swartz, "Physics With Polarized Electrons", SLAC-PUB-4656 (1986).

[43] M. L. Swartz, "Complete order-α^3 calculation of the cross section for polarized Compton scattering", Phys. Rev. **D58**:014010, 1998.

[44] M. B. Smy, "Measurement of Z^0 Lepton Coupling Asymmetries", PhD Thesis, SLAC-REPORT-515, 1997.

[45] M. Fero, P. L. Reinertsen, B. A. Schumm, M. Swartz, and E. Torrence, "Compton Polarization Measurement: 1995", SLD Physics Note **50**, 1996.

[46] K. Abe et al., "Design and performance of the SLD vertex detector: a 307 Mpixel tracking system", Nucl. Inst. Meth. **A400** (1997) 287-343.

[47] C. J. S. Damerell, "Vertex Detectors", RAL 86-077 (1986).

[48] T. R. Junk, "Measurement of Polarized Forward-Backward Asymmetry of B Quarks Using Momentum-Weighted Track Charge at SLD", PhD Thesis, SLAC-REPORT-95-476, 1995.

[49] M. O. Dima, "Production of Strange Vector Mesons at the Z^0 Resonance", PhD Thesis, SLAC-REPORT-505, 1997.

[50] T. J. Pavel, "Measurement of Charged Hadron Spectra at the Z^0 with Chrenkov Ring Imaging", PhD Thesis, SLAC-REPORT-491, 1997.

[51] D. Axen et al., "The lead-liquid argon sampling calorimeter of the SLD detector", Nucl. Inst. Meth. **A328** (1993) 472-494.

[52] W. R. Leo, "Techniques for Nuclear and Particle Physics Experiments", Springer-Verlag, 1987.

[53] W. R. Nelson, T. M. Jenkins, R. C. McCall, and J. K. Kobb, Phys. Rev. **149**, 201 (1966);
G. Bathow et. al, Nucl. Phys. **B20** 592, (1970);
E. Longo and I. Sestili, Nucl. Inst. Meth. **128**, 283 (1975).

[54] D. Falciai, "Electron Identification is SLD with a Neural Network", SLD Physics Note **44**, 1995.

[55] A. C. Benevenuti et al., "The Iron Calorimeter and Muon Identifier for SLD", Nucl. Inst. Meth. **A276** (1989) 94-104.

[56] E. Iarocci, "Plastic streamer tubes and their application in high-energy physics", Nucl. Inst. Meth. **217**, 30 (1983);
G. Battistoni et al., "Electrodeless plastic streamer tubes", Nucl. Inst. Meth. **217**, 429 (1983).

[57] The SLD Collaboration, "SLD Design Report", SLAC-REPORT-273 (1984);
The SLD Collaboration, "A Status Report on the SLD Data Acquisition Sytstem", IEEE Trans. Nucl. Sci. **36**: 23-28, 1989.

[58] J. T. Walker, Soo-Ik Chae, S. Shapiro, R. S. Larsen, "Microstore - The Stanford Analog Memory Unit", IEEE Trans. Nucl. Sc. **NS-32**, 616 (1985).

[59] A. Marchioro, W. von Rüden, and G. McPherson, "The Aleph Event Builder", IEEE Trans. Nucl. Sci. **NS-34** 133 (1987).

[60] J. M. Yamartino, "Hadronic Event Selection Using the LAC", SLD Physics Note **14**, 11/04/92;
E. Vella, J. Yamartino, "Tower Thresholds for LAC Total Energy Trigger", SLD Note **213**, February 6, 1992.

[61] J. M. Yamartino, "A Measurement of the e^+e^- Decay Width of the Z^0", PhD thesis, SLAC-REPORT-426 (1994).

[62] T. Sjöstrand, "The Lund Monte Carlo for Jet Fragmentation and e^+e^- Physics: Jetset Version 6.2", Comput. Phys. Commun. **39**: 347 (1986);
T. Sjöstrand and M. Bengston, "The Lund Monte Carlo for Jet Fragmentation and e^+e^- Physics: Jetset Version 6.3: An Update", Comput. Phys. Commun. **43**: 367 (1987).

[63] T. Sjöstrand, "Pythia 5.7 and Jetset 7.4: Physics and Manual", LU-TP 95-20 and hep-ph/9508391.

[64] Su Dong, private communication, November 22, 1999.

[65] CLEO Collaboration, "Measurement of $B \to D_s^+ X$ decays", Phys. Rev. **D53** 4734-4746 (1996).

[66] Neyman, J. and Pearson, E. S., "On the problem of the most efficient tests of statistical hypotheses", Phil. Trans. **A231** 289, 1933.

[67] P. G. Hoel, "Introduction to Mathematical Statistics", 3rd ed., Wiley 1962.

[68] K. Kleinknecht, "Detectors for particle radiation", Cambridge 1986.

[69] G. Mancinelli, "Improved Muon Identification at SLD", SLD Physics Note **65**, 1998.

[70] D. Aston, T. Pavel and D. Muller, "Application of CRID Particle Identification", SLD Physics Note **35**, 1995;
P. Antilogus, M. Kalelkar and D. Muller, "Doing Physics with the CRID", SLD Physics Note **208**, 1995;
J. Va'vra, "Finding Cherenkov Angles and Testing Hypotheses", CRID MEMO **49**, 1993.

[71] Sir Maurice Kendall and Alan Stuart, "The Advanced Theory of Statistics" v.2, 4th ed., Macmillan 1979.

[72] W. H. Press, B. P. Flannery, S. A. Teukolsky, W. T. Vetterling, "Numerical Recipes in C", Cambridge 1989.

[73] G. Casella and R. L. Berger, "Statistical Inference", Duxbury Press 1990.

[74] V. V. Serbo, "Measurement of the Polarized Forward-Backward Asymmetry of B Quarks at SLD", PhD Thesis, SLAC-REPORT-510, 1997.

[75] D. J. Jackson, "A topological vertex reconstruction algorithm for hadronic jets", Nucl. Inst. Meth. **A388** (1997) 247-253.

[76] "MINUIT - Function Minimization and Error Analysis" Version 94.1, CERN Program Library entry **D506**.

[77] Daning Dong, "Measurement of the b-quark Fragmentation Function in Z^0 Decays", SLAC-PUB-8153, and private communication Dec 14, 1999.