# A MEASUREMENT OF THE ANOMALOUS MAGNETIC MOMENT OF THE POSITIVE MUON WITH A PRECISION OF 0.7 PARTS PER MILLION

BY

FREDERICK EARL GRAY, JR.

B.A., Rice University, 1997
M.S., University of Illinois at Urbana-Champaign, 1999

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Physics
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2003

Urbana, Illinois

A MEASUREMENT OF THE ANOMALOUS MAGNETIC MOMENT OF THE POSITIVE MUON WITH A PRECISION OF 0.7 PARTS PER MILLION

Frederick Earl Gray, Jr., Ph.D.
Department of Physics
University of Illinois at Urbana-Champaign, 2003
Professor David W. Hertzog, Advisor

The anomalous magnetic moment of the positive muon has been measured to a precision of 0.7 parts per million in an experiment at Brookhaven National Laboratory. The standard model prediction for this quantity can be calculated with a similar uncertainty, although there are currently inconsistent results for one crucial part of this evaluation. Comparison of the theoretical and experimental values tests the standard model and probes for physics beyond it, including supersymmetry, for which a large signal is expected from many models. This thesis describes the principle of the experiment and the status of the theoretical evaluations. It concentrates on the details of an energy-binned method of determining the anomalous precession frequency from the recorded data. The result that is obtained, after combination with previous measurements, is

$$a_\mu^{exp;avg} = 11\,659\,203(8) \times 10^{-10} \ ,$$

to be compared with the present standard model predictions

$$\begin{aligned} a_\mu^{SM;e^+e^-} &= 11\,659\,168.8(7.7) \times 10^{-10} \text{ and} \\ a_\mu^{SM;\tau} &= 11\,659\,193.2(6.8) \times 10^{-10} \ . \end{aligned}$$

which correspond to differences of 3.1 or 1.0 standard deviations, respectively. The implications of this result therefore await the resolution of the theoretical debate.

*In memory of my father*
*Frederick Earl Gray*
*and my grandmother*
*Alice Wilson Griffin.*

# Acknowledgments

The measurement described in this thesis is the work of a "cast of thousands," the Brookhaven National Laboratory (BNL) Muon $(g-2)$ Collaboration. For the data set described in this thesis, the collaboration consisted of

- Boston University: R.M. Carey, E. Efstathiadis, M.F. Hare, X. Huang, F. Krienen, A. Lam, I. Logashenko, J.P. Miller, J. Paley, Q. Peng, O. Rind, B.L. Roberts, L.R. Sulak, and A. Trofimov;
- Brookhaven National Laboratory: G.W. Bennett, H.N. Brown, G. Bunce, G.T. Danby, R. Larsen, Y.Y. Lee, W. Meng, J. Mi, W.M. Morse, D. Nikas, C. Özben, R. Prigl, Y.K. Semertzidis, and D. Warburton;
- Cornell University: Y. Orlov;
- *Universität Heidelberg* (Germany): A. Grossmann, G. zu Putlitz, and P. von Walter;
- University of Illinois at Urbana-Champaign: P.T. Debevec, W. Deninger, D.W. Hertzog, C.J.G. Onderwater, C. Polly, M. Sossong, and D. Urner;
- KEK (Japan): A. Yamamoto;
- *Kernfysisch Versneller Instituut* (The Netherlands): K. Jungmann;
- University of Minnesota: B. Bousquet, P. Cushman, L. Duong, S. Giron, J. Kindem, I. Kronkvist, R. McNabb, T. Qian, and P. Shagin;
- Budker Institute of Nuclear Physics, Novosibirsk (Russia): V.P. Druzhinin, G.V. Fedotovich, D. Grigoriev, B.I. Khazin, N.M. Ryskulov, Yu.M. Shatunov, and E. Solodov;
- Tokyo Institute of Technology (Japan): M. Iwasaki;
- Yale University: M. Deile, H. Deng, S.K. Dhawan, F.J.M. Farley, V.W. Hughes, D. Kawall, M. Grosse-Perdekamp, J. Pretz, S.I. Redin, E. Sichtermann, and A. Steinmetz.

While interactions with all of these people have been both productive and enjoyable, the Illinois group deserves exceptional mention. I have benefitted greatly from the guidance of my research advisor, David Hertzog, and from working with the other current and past members of the precision muon physics group:

Dan Chitwood, Steven Clayton, Paul Debevec, Bill Deninger, Peter Kammel, Brendan Kiburg, Gerco Onderwater, Cenap Özben, Chris Polly, Andrea Sharp, David Urner, and Mike Sossong.

Funding for construction and operation of the experiment came from the U.S. Department of Energy, the National Science Foundation, the German *Bundesminister für Bildung und Forschung*, the Russian Ministry of Science, and the U.S.-Japan Agreement in High Energy Physics. The computing facilities of the National Computational Science Alliance, specifically the Platinum IA-32 Linux cluster and the UniTree storage system, were employed pursuant to grant PHY-010004N. My tuition, stipend, and travel expenses were supported by the National Science Foundation under contracts NSF-PHY-93-10873 and NSF-PHY-00-72044, by a GE Fellowship, and by a Mavis Memorial Fund Scholarship award.

Finally, a word about the dedication: I am very grateful to my father and my grandmother (who served as a mother figure for most of my childhood) for all of their love. Unfortunately, they both passed away while I was in graduate school, and they are indeed missed. Many thanks are also due to David and Lorraine Griffin, who were invaluable in helping me to pick up the pieces; I really don't know what I would have done without them.

# Table of Contents

# Chapter 1

# Introduction

This thesis describes a precise measurement of the anomalous magnetic moment of the positive muon. The experiment, performed at Brookhaven National Laboratory (BNL) and known as E821, has a long history: a collaboration began to form in 1984, construction began in 1989, and the first data were obtained in 1997. The data set described in this thesis was collected between December 1999 and March 2000; it represents E821's final running period with positive muons.

This chapter defines what is meant by the "anomalous magnetic moment," sets out the motivation for measuring it, and explains the general principle of the experiment. It also provides a brief history of previous measurements of this quantity and a sense of the evolution of the experimental precision with time.

## 1.1 Magnetic moments

The magnetic dipole moment $\vec{\mu}$ of an object is a measure of how much torque it experiences when placed in a magnetic field:

$$\vec{\tau} = \vec{\mu} \times \vec{B} \,. \tag{1.1}$$

The associated potential energy is

$$U = -\vec{\mu} \cdot \vec{B} \,. \tag{1.2}$$

Subatomic particles have a magnetic moment that is generated by their intrinsic spin; these two quantities are related by the gyromagnetic ratio $g$:

$$\vec{\mu} = g \left( \frac{e}{2m} \right) \vec{s} \,. \tag{1.3}$$

Table 1.1 shows the gyromagnetic ratios for various particles. It is apparent that $g$ is approximately 2 for the charged leptons–that is, the electron, muon, and tau. Indeed, the Dirac equation predicts that $g$ is exactly 2 for a pointlike particle. The small discrepancy is caused by corrections from higher-order interactions described by quantum field theories. Composite particles such as the proton and neutron show large differences from 2, which are indications of their rich internal structure.

Table 1.1 also demonstrates that the gyromagnetic ratios of stable and nearly-stable particles can be measured experimentally to extremely high precision. Indeed, the electron $g$ factor, known to a precision of

| Particle | Experimental value | Relative precision | Ref. | Theoretical prediction | Ref. |
|---|---|---|---|---|---|
| Electron | 2.0023193043738(82) | $4\times10^{-12}$ | [1] | 2.00231930492(29) | [2] |
| Muon | 2.0023318406(16) | $8\times10^{-10}$ | [3] | 2.0023318338(14) | [4] |
| Tau | 2.008(71) | $4\times10^{-2}$ | [5] | 2.0023546(6) | [6] |
| Proton | 5.585694674(58) | $1\times10^{-8}$ | [1] | 5.58 | [7] |
| Neutron | -3.8260854(10) | $3\times10^{-7}$ | [1] | -3.72 | [7] |

Table 1.1: Gyromagnetic ratios ($g$) for various subatomic particles.

$4\times10^{-12}$, is the physical quantity with the smallest quoted uncertainty in the current CODATA table [8]. The lepton $g$ factors can also be calculated very precisely in the context of the standard model, and a comparison of the experimental and theoretical values provides a sensitive test of the theory.

## 1.2 Theoretical calculation

In an extremely naive classical model, the muon's spin can be imagined as a rotation about its own axis, like the spin of a toy top. To compute the associated magnetic moment, it is assumed that the muon moves with speed $v$ along a circular path of radius $r$; eventually, the limit $r\to0$ will be taken. The muon then effectively constitutes a loop that encloses an area $A=\pi r^2$ and carries a current $I=ev/2\pi r$. The magnetic moment is

$$\mu = IA = \frac{evr}{2}\ . \tag{1.4}$$

The angular momentum associated with the particle's motion is $L=mvr$, so

$$\mu = \left(\frac{e}{2m}\right)L\ . \tag{1.5}$$

Thus, $\mu$ is now independent of $r$. Comparison with Equation 1.3 shows that $g=1$. This classical prediction clearly disagrees with the experimental results.

Quantum mechanics leads to a very different conclusion. In the nonrelativistic limit, the Dirac equation may be transformed into the Schrödinger-Pauli equation [9]:

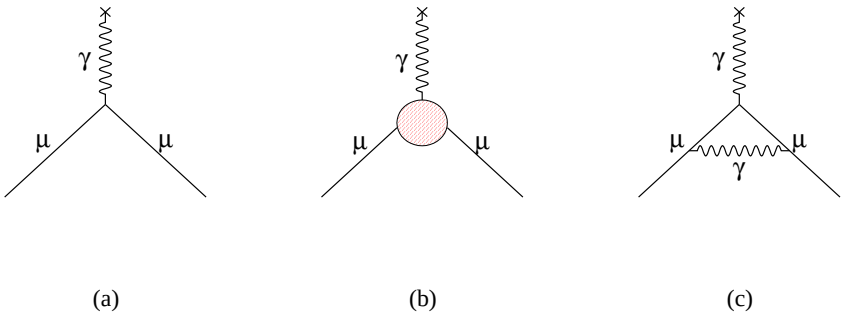
$$\left(\frac{1}{2m}(\vec{P}+e\vec{A})^2+\frac{e}{2m}\vec{\sigma}\cdot\vec{B}-eA^0\right)\psi_A=(E-m)\psi_A\ . \tag{1.6}$$

The magnetic moment term is identified in the Hamiltonian by comparison with Equation 1.2. For a spin-$\frac{1}{2}$ particle, $\vec{S}=\frac{1}{2}\vec{\sigma}$, so

$$\mu=-\frac{e}{2m}\vec{\sigma}=-2\frac{e}{2m}\vec{S}\,. \tag{1.7}$$

The Dirac equation therefore implies that $g=2$. Foldy has shown that this result holds even in a fully relativistic treatment [10].

The transition from quantum mechanics to quantum field theory adds another level of insight into the gyromagnetic ratio. The muon's magnetic moment is illustrated as a Feynman diagram in Figure 1.1(a).




(a) (b) (c)

Figure 1.1: Feynman diagrams for (a) the magnetic moment, corresponding to $g = 2$, (b) the general form of diagrams that contribute to the anomalous magnetic moment $a_\mu$, and (c) the Schwinger term.

This diagram shows the muon coupling directly to a photon from the external magnetic field; it corresponds to the Dirac equation's prediction that $g = 2$. However, there is a large set of radiative corrections; these couplings to virtual fields lead to an anomalous part $a_\mu$ of the magnetic moment, defined by

$$a_\mu = \frac{1}{2}(g-2) \; . \tag{1.8}$$

These corrections are represented symbolically by the "blob" in Figure 1.1(b); any allowed intermediate state may be inserted in its place. The dominant correction arises from a coupling to a single virtual photon, as shown in Figure 1.1(c). This leading-order contribution was first evaluated by Schwinger [11]:

$$a_\mu^{Schwinger} = \frac{\alpha}{2\pi} \approx 0.0012 \; . \tag{1.9}$$

Many higher-order processes lead to additional perturbations; they will be discussed in detail in Chapter 2. Contributions are included from all known particles and interactions. Any discrepancy remaining between the experimental and theoretical values must then be the result of new physics: particles and interactions that are not included in the standard model. The search for such a difference, if it exists, provides the motivation for the experiment.

## 1.3 Experimental principles

In the experiment, a collection of longitudinally polarized muons is injected into a storage ring where they follow a circular central orbit in a magnetic dipole field $\vec{B}$ and a vertically focusing electric quadrupole field $\vec{E}$. As shown in Figure 1.2, the field rotates the direction of their polarization with respect to their momentum.

In its own rest frame, the muon spin rotates at a rate that is proportional to $g$ according to the Larmor

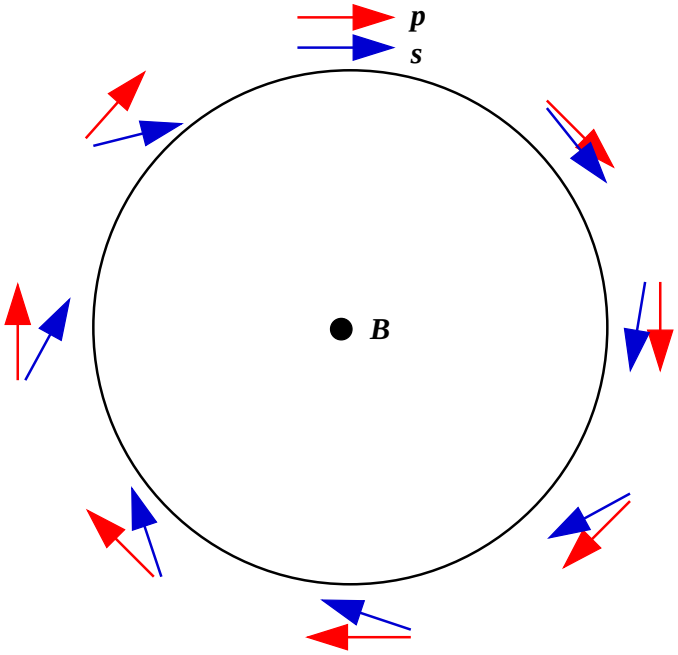



Figure 1.2: Illustration of the motion of the muon spin vector relative to its momentum vector.

formula

$$\vec{\omega}_s^* = g\frac{e}{2m}\vec{B}^* \,. \tag{1.10}$$

The frequency is written as a 3-vector; each component describes the motion of the spin about one of the three spatial axes so that

$$\frac{\partial \vec{s}}{\partial t} = \vec{\omega}_s \times \vec{s} \,. \tag{1.11}$$

The starred quantities are measured in the muon's rest frame.

The muon moves relativistically along a circular path in the lab frame, and its cyclotron frequency–the frequency at which it completes circular orbits–is given by

$$\vec{\omega}_c = \frac{e\vec{B}}{m\gamma} \,. \tag{1.12}$$

The muon experiences a transverse centripetal acceleration, which leads to a Thomas precession [12] of all observables in the lab frame. This effect may be thought of as a Lorentz contraction of the rest frame axes [13], and it causes the spin to appear to turn at a frequency

$$\vec{\omega}_T = (\gamma - 1)\omega_c = (\gamma - 1)\frac{e\vec{B}}{m\gamma} \,. \tag{1.13}$$

Because of mutually canceling factors of $\gamma$, the Larmor spin precession occurs at the same rate in both frames. Time moves faster in the muon rest frame, but the magnetic field is stronger there, and the product

$\vec{B}t$ is an invariant. Altogether, the spin in the lab frame rotates at a frequency

$$\vec{\omega}_s = g\frac{e}{2m}\vec{B} + (\gamma - 1)\frac{e\vec{B}}{m\gamma} \,. \tag{1.14}$$

Jackson [14], citing Thomas [12] and Bargmann, Michel, and Telegdi [15], expands this expression to include electric as well as magnetic fields:

$$\vec{\omega}_s = \frac{e}{m}\left[\left(\frac{g}{2} - 1 + \frac{1}{\gamma}\right)\vec{B} - \left(\frac{g}{2} - 1\right)\frac{\gamma}{\gamma+1}(\vec{\beta}\cdot\vec{B})\vec{\beta} - \left(\frac{g}{2} - \frac{\gamma}{\gamma+1}\right)(\vec{\beta}\times\vec{E})\right] \,. \tag{1.15}$$

In this general treatment, the cyclotron frequency is written as [16]

$$\vec{\omega}_c = \frac{1}{\beta^2}\frac{\partial\vec{\beta}}{\partial t}\times\vec{\beta} = \frac{e}{m}\left[\frac{1}{\gamma}\vec{B} - \frac{1}{\gamma\beta^2}(\vec{\beta}\cdot\vec{B})\vec{\beta} - \frac{\gamma}{\gamma^2-1}(\vec{\beta}\times\vec{E})\right] \,. \tag{1.16}$$

The anomalous precession frequency $\vec{\omega}_a$ (the "$(g-2)$ frequency") is defined as the *difference* of the spin precession and cyclotron frequencies. It is the frequency at which the muon's spin advances relative to its momentum. Assuming that the motion of the muons is purely longitudinal so that no component of the momentum is parallel to the magnetic field, the terms containing $\vec{\beta}\cdot\vec{B}$ drop out, leaving

$$\vec{\omega}_a = \vec{\omega}_s - \vec{\omega}_c = \frac{e}{m}\left[a_\mu\vec{B} - \left(a_\mu - \frac{1}{\gamma^2-1}\right)(\vec{\beta}\times\vec{E})\right] \,. \tag{1.17}$$

This expression may be simplified further by choosing $\gamma = \sqrt{\frac{1}{a_\mu}+1} \approx 29.3$ so that the dependence on $\vec{E}$ disappears. This $\gamma$ corresponds to a "magic" muon momentum of 3.09 GeV/$c$; the experiment runs at this momentum so that a large electric quadrupole field can be used to confine the particles in the storage ring. The relationship between $a_\mu$ and $\vec{B}$ is then

$$\vec{\omega}_a = \frac{e}{m}a_\mu\vec{B} \,. \tag{1.18}$$

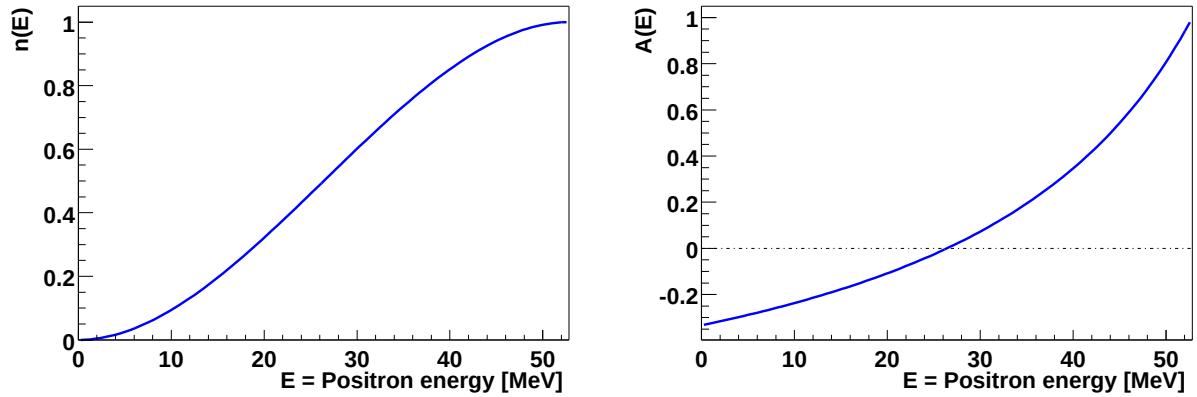
The quantity $a_\mu$ may in principle be determined by measuring the anomalous precession frequency and the magnetic field very precisely and taking their ratio. It must be emphasized that $\vec{\omega}_a$ is proportional to $a_\mu$, not to $g$. The experiment directly measures only the perturbations to the magnetic moment, permitting three orders of magnitude higher precision than if it measured the entire quantity.

Parity violation in the weak interaction allows the spin direction to be measured as a function of time. Eventually, essentially every muon decays by $\mu^+ \to e^+\nu_\mu\bar{\nu}_e$. In the rest frame of the muon, the differential probability for the positron to emerge with a normalized energy $y = E/E_{max}$ at an angle $\theta_s$ with respect to the muon spin is [13]

$$\frac{dP}{dy\,d\Omega} = n(y)[1 + A(y)\cos\theta_s] \quad \text{with} \tag{1.19}$$

$$n(y) = y^2(3-2y) \quad \text{and} \tag{1.20}$$

$$A(y) = \frac{2y-1}{3-2y} \,. \tag{1.21}$$




(a) $n(E)$ (b) $A(E)$

Figure 1.3: Number density and asymmetry distributions for decay positrons in the muon rest frame.

The quantities $n(E)$ and $A(E)$ are plotted in Figure 1.3, normalized to $E_{max} = 52.8$ MeV. Positrons with $y > 0.5$ are most likely to be emitted in the direction of the muon spin, and those with lower energy are most likely to be emitted opposite it. Because more positrons are emitted with $y > 0.5$ and because their asymmetry is higher, the overall effect is that the decay positrons tend to go in the direction of the muon spin. Integrating over all energies,

$$\int_0^1 \frac{dP}{dy\, d\Omega}\, dy = \frac{1}{2}[1 + \frac{1}{3}\cos\theta] \; . \tag{1.22}$$

If a nonzero energy threshold is established, then the asymmetry is even higher than $\frac{1}{3}$. Because the spin appears to rotate at the frequency $\omega_a$, so does the distribution of decay positrons.

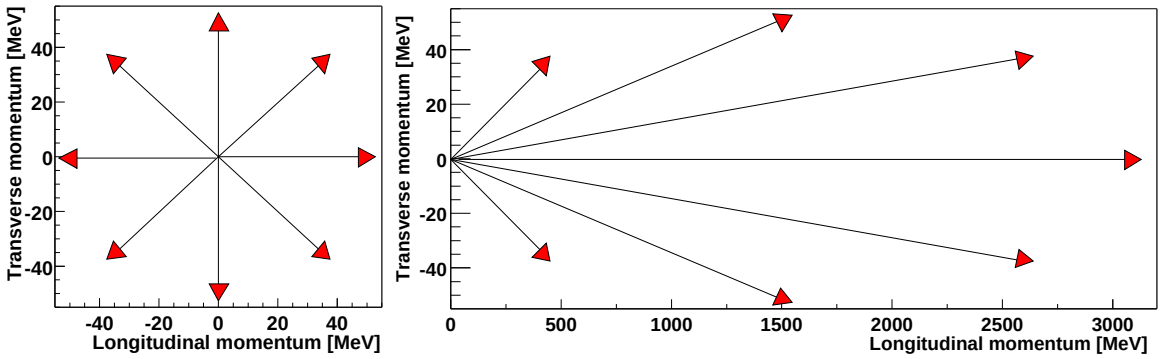
In the laboratory frame, the stored muons are highly relativistic so that the range of observed decay angles is extremely compressed; the effect of the Lorentz boost is illustrated in Figure 1.4. Compared to the angle subtended by a practical detector, all decays are forward. The energy of the positron in the laboratory frame is given by

$$E_{lab} = \gamma(E^* + \beta p^* \cos\theta^*) \; . \tag{1.23}$$

The positron energy $E^*$ is also high enough, in general, to justify a fully relativistic treatment, so

$$E_{lab} = \gamma E^*(1 + \cos\theta^*) \; . \tag{1.24}$$

The laboratory energy clearly depends strongly on the decay angle $\theta^*$. To have a high energy in the laboratory frame, a positron must have a high energy in the CM frame and also be emitted at a forward angle there. Setting a laboratory energy threshold therefore selects a range of angles in the muon rest frame.




(a) CM frame

(b) Laboratory frame

Figure 1.4: Illustration of the Lorentz boost of the decay positron from the muon rest frame to the laboratory frame.

Consequently, the number of particles detected above such a threshold as a function of time is modulated with the frequency $\omega_a$. The positron arrival time histogram for the data set analyzed in this thesis appears in Figure 1.5. The oscillations from the anomalous precession are indeed present as expected, visible on top of an exponential baseline that results from the decay of the muon population. In the absence of any instrumental backgrounds, this spectrum is described by the functional form

$$N(t) = N_o e^{-t/\tau}[1 - A\cos(\omega_a t + \phi_a)] \,, \tag{1.25}$$

and $\omega_a$ is extracted by fitting this functional form to the data.

## 1.4 Evolution of experimental precision

The anomalous magnetic moment $a_\mu$ was measured by a succession of three experiments of increasing precision and complexity at CERN, beginning in 1958 and concluding in 1977.

The first CERN experiment [17] employed a 6 m straight magnet. Particles were injected at one end and followed a spiral path to the other, "walking" slowly down the magnet because of a small gradient in the field. The amount of time spent in the field could be varied by tuning the gradient. The muons were then stopped in a polarimeter outside the magnet in which their front/back decay asymmetry was measured; this asymmetry was plotted as a function of storage time to determine $\omega_a$. The experiment obtained a result with an uncertainty of 4300 ppm that agreed with the prediction of quantum electrodynamics (QED) for a structureless particle. The result was surprising at the time because it had been assumed, in order to account for the mass difference between the muon and the electron, that the muon had significant internal structure.

The second CERN experiment [18] used a storage ring with a muon momentum of 1.3 GeV/$c$. A radial magnetic field gradient was used to focus the beam. The primary proton beam was injected directly onto an internal target on one side of the ring. Pions were produced at the target; a small fraction of the muons

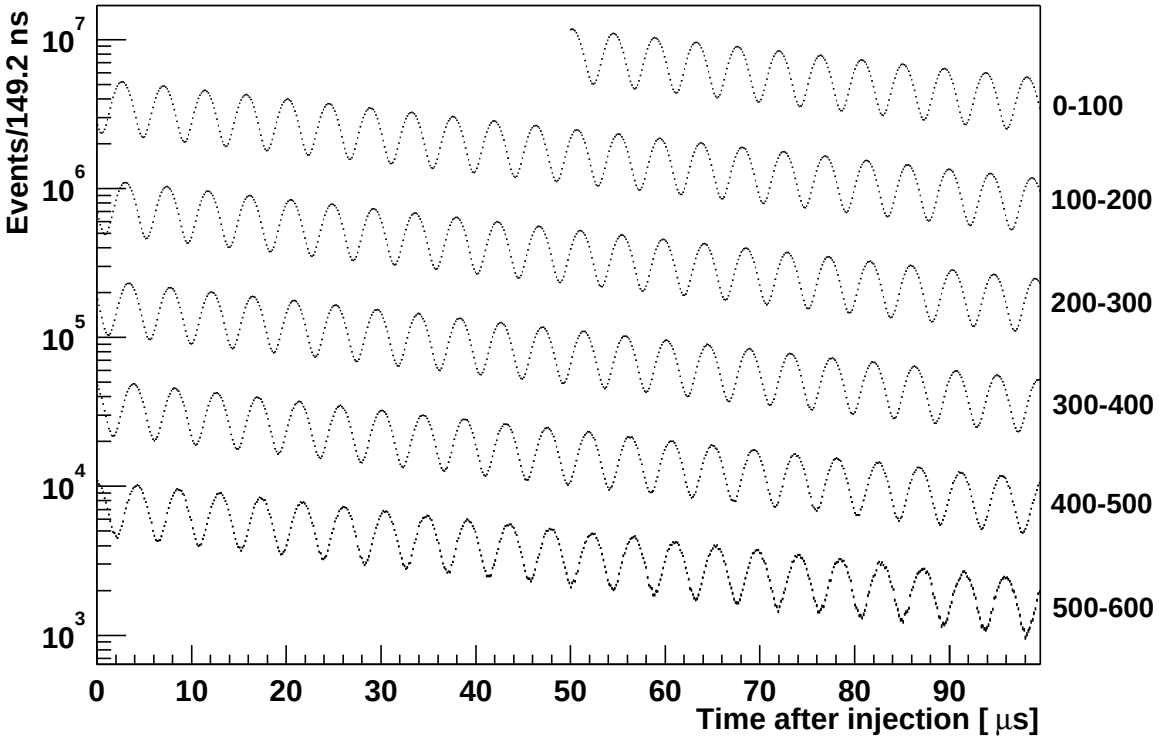



Figure 1.5: Spectrum of arrival times for decay positrons with detected energies greater than 2 GeV ($y \approx 0.6$). The time axis wraps around several times in order to facilitate the presentation of a long time range in one figure.

| Experiment | Years | Polarity | $a_\mu \times 10^{10}$ | Precision [ppm] | Reference |
|---|---|---|---|---|---|
| CERN I | 1961 | $\mu^+$ | 11 450 000(220 000) | 4300 | [17, 13] |
| CERN II | 1962-1968 | $\mu^+$ | 11 661 600(3100) | 270 | [18, 13] |
| CERN III | 1974-1976 | $\mu^+$ | 11 659 100(110) | 10 | [19, 13] |
| CERN III | 1975-1976 | $\mu^-$ | 11 659 360(120) | 10 | [19, 13] |
| BNL | 1997 | $\mu^+$ | 11 659 251(150) | 13 | [20] |
| BNL | 1998 | $\mu^+$ | 11 659 191(59) | 5 | [21] |
| BNL | 1999 | $\mu^+$ | 11 659 202(15) | 1.3 | [22] |
| BNL | 2000 | $\mu^+$ | 11 659 204(9) | 0.7 | [3] |
| BNL | 2001 | $\mu^-$ | Analysis in progress | $\approx 0.7$ | |
| Average | | | 11 659 203(8) | 0.7 | |

Table 1.2: Summary of $a_\mu$ results from various experiments and data sets, showing the evolution of experimental precision over time.

resulting from their decay were produced within the accepted phase space of the ring. This process produced a tremendous flash of prompt radiation when the beam was injected, so the positron detectors were placed only on the side of the ring away from the target, and the space in the middle of the ring was filled with concrete shielding. This experiment obtained a result with a precision of 270 ppm, and the value was again in agreement with QED.

For the third CERN experiment [19], a larger storage ring with a higher field increased the muon momentum to the "magic" value of 3.1 GeV/$c$. An electric quadrupole field was employed to focus the beam. The production target was located well away from the storage ring, so the injected beam consisted of pions, which decayed within the storage ring to produce muons. Pion injection necessitated the introduction of an inflector magnet to cancel out the field of the storage ring and provide a field-free path for the entering beam. This experiment achieved a precision of 10 ppm for each muon polarity. If $CPT$ symmetry is assumed, the results can be combined to give a 7 ppm measurement.

The present BNL experiment continues to refine the experimental technique. As in the final CERN experiment, a 3.1 GeV/$c$ storage ring is used. A continuous superconducting magnet is used instead of a lattice of discrete resistive magnets. Muons, rather than pions, are the injected species. The increased storage efficiency allows higher intensity with lower background. This change requires, in addition to an inflector, a pulsed kicker magnet. The measurement reported in this thesis is of the positive muon and has a precision of 0.7 ppm.

The results of the CERN and BNL experiments are summarized in Table 1.2. To underscore the tremendous improvement over time in experimental precision, the most recent measurements are presented graphically in Figure 1.6.

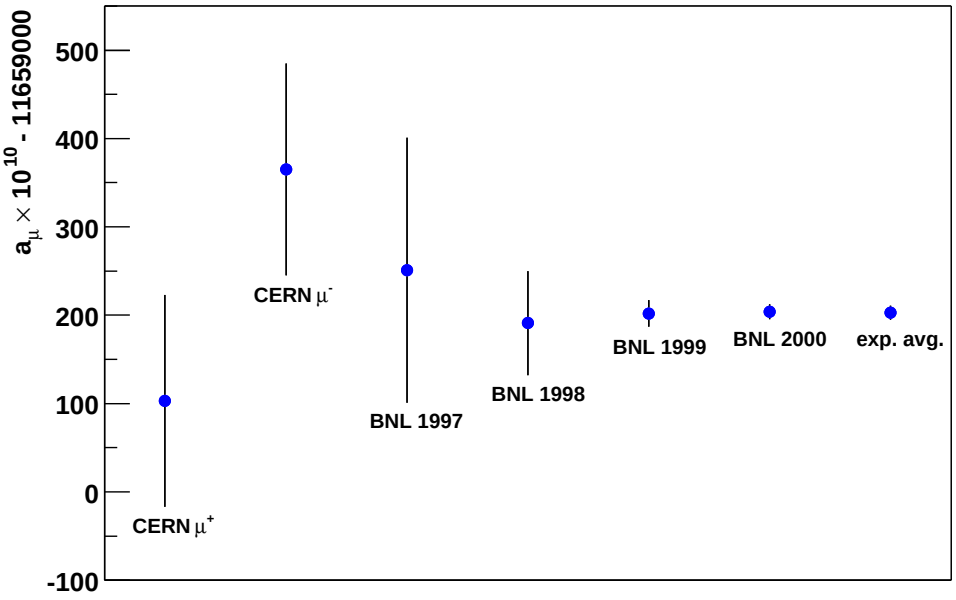



Figure 1.6: Comparison of $a_\mu$ results from various experiments and data sets.

# Chapter 2

# Standard model contributions to $a_\mu$

This chapter describes the calculation of the theoretical value of $a_\mu$ in the context of the standard model. The calculation is divided into terms that arise from QED, electroweak interactions, and hadronic physics.

## 2.1 Quantum electrodynamics

By far the largest contribution to $a_\mu$ comes from interactions described by QED, which involve only leptons and photons. This term is expressed as a power series expansion in the number of loops in the Feynman diagrams; each loop leads to a factor of $\left(\frac{\alpha}{\pi}\right)$. The leading-order term in this expansion is the Schwinger term that was mentioned in Chapter 1, which contains a single virtual photon. Kinoshita and his collaborators have evaluated the series through order $\left(\frac{\alpha}{\pi}\right)^4$ for the electron ([23] and references therein). The calculation for the muon shares a large set of diagrams in common with the electron; only interactions that contain more than one lepton flavor must be computed separately. These evaluations have also been performed by Kinoshita's group, with a full calculation through order $\left(\frac{\alpha}{\pi}\right)^4$ and an estimate of the next coefficient ([24] and references therein). The results of these calculations are summarized in Table 2.1. Diagrams with one and two loops were all evaluated analytically, while numerical methods were employed for the higher-order processes. The indicated uncertainties for the coefficients take into account both numerical errors and contributions from diagrams that were not evaluated. In all, more than 1000 diagrams were evaluated over a period of decades; this calculation represents a clear *tour de force*.

Hughes and Kinoshita use these expansion coefficients to compute a numerical value [25]:[1]

$$a_\mu^{QED} = 116\,584\,70.57(0.18)(0.05) \times 10^{-10} \ . \tag{2.1}$$

The first error estimate is propagated from the coefficients while the second comes from the uncertainty in the fine structure constant $\alpha$. Combining these two uncertainties in quadrature, the overall relative error is 16 ppb, which is entirely insignificant on the scale of the 0.7 ppm experimental uncertainty. The value for $\alpha$

[1] Kinoshita and Nio have recently uncovered [26] a small numerical error in the evaluation of a few of the order $\left(\frac{\alpha}{\pi}\right)^4$ terms. This error modifies $\alpha^{-1}$ and $a_\mu^{QED}$ by a few parts per billion (ppb). It is small enough that it does not have any impact on the comparison of the theoretical and experimental values of $a_\mu$. The authors do not directly give a revised result, referring instead to a more detailed paper in preparation. Consequently, the uncorrected QED results are given here.

| Order | Electron | | | Additional component for muon | | |
|---|---|---|---|---|---|---|
| | Diagrams | Coefficient | Value | Diagrams | Coefficient | Value |
| $\left(\frac{\alpha}{\pi}\right)$ | 1 | 0.5 | $1.16 \times 10^{-3}$ | 0 | 0 | 0 |
| $\left(\frac{\alpha}{\pi}\right)^2$ | 7 | -0.328478965 | $-1.77 \times 10^{-6}$ | 2 | 1.0943370 | $5.90 \times 10^{-6}$ |
| $\left(\frac{\alpha}{\pi}\right)^3$ | 72 | 1.17611(42) | $1.47 \times 10^{-8}$ | 25 | 22.8676(33) | $2.87 \times 10^{-7}$ |
| $\left(\frac{\alpha}{\pi}\right)^4$ | 518 | -1.434(138) | $-4.17 \times 10^{-11}$ | 471 | 127.00(41) | $3.70 \times 10^{-9}$ |
| $\left(\frac{\alpha}{\pi}\right)^5$ | (not evaluated) | | | (estimate) | 570(140) | $3.85 \times 10^{-11}$ |

Table 2.1: Coefficients of QED terms for $a_e$ and $a_\mu$ in the power series expansion. The values are taken from [23] and [24]. The number of diagrams includes only those that were actually evaluated, rather than being determined by inspection to be negligible.

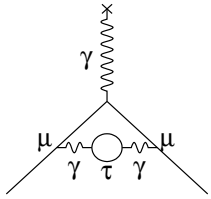



Figure 2.1: Feynman diagram for vacuum polarization by a $\tau$ loop.

that is used in this evaluation is determined from $a_e$, the anomalous magnetic moment of the electron, and it is

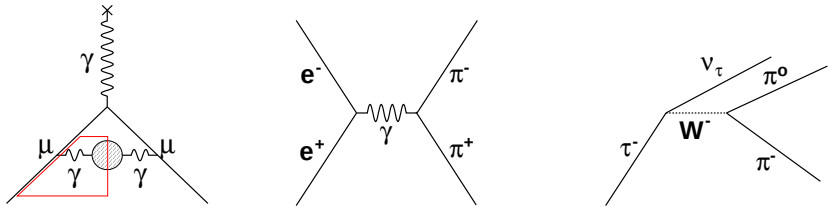
$$\alpha^{-1} = 137.035\,999\,58(52) \ .$$

In other words, rather than regarding the experiment as a test of the theory for the electron, it is assumed that they agree, and the corresponding value of $\alpha$ is extracted.

One of the two-loop QED diagrams, vacuum polarization by a $\tau$ loop, is illustrated in Figure 2.1. The form of its contribution to $a_\mu$ is instructive because it is typical of the effect of a heavy particle [27]:

$$a_\mu^{\tau VP} = \left[\frac{1}{45}\left(\frac{m_\mu}{m_\tau}\right)^2 + O\left(\frac{m_\mu^4}{m_\tau^4}\ln\frac{m_\tau}{m_\mu}\right)\right]\left(\frac{\alpha}{\pi}\right)^2 \ . \tag{2.2}$$

The factor $\left(\frac{m_l}{m_X}\right)^2$ generally appears for this sort of diagram, where $X$ represents the mass associated with the heavy loop and $l$ is the lepton of interest. The magnetic moment of the muon is therefore more sensitive to these effects than that of the electron by a factor of order $\left(\frac{m_e}{m_\mu}\right)^2 = 4.3 \times 10^4$. The stability of the electron has permitted $a_e$ to be measured approximately 200 times more precisely than $a_\mu$. However, the muon is clearly still a more suitable probe for the effects of physics beyond the standard model, which would in many cases include new vacuum polarization loops of this form. Although new physics contributions to the $\tau$'s magnetic moment would be quite large, it is so short-lived that $a_\tau$ cannot be measured with any useful




(a) (b) (c)

Figure 2.2: Feynman diagrams for (a) hadronic vacuum polarization, (b) dominant hadron production process $e^+e^- \to \pi^+\pi^-$, and (c) equivalent $\tau$ decay process $\tau^- \to \nu_\tau \pi^- \pi^o$.

precision. The muon therefore provides an excellent compromise between mass and stability.

## 2.2 Hadronic vacuum polarization

Hadronic vacuum polarization processes, as illustrated in Figure 2.2(a), provide the second-largest contribution to $a_\mu$. Because QCD is nonperturbative at low energies, this term is difficult to calculate from first principles. Fortunately, it can be related to experimental data through the dispersion relation [13]

$$a_\mu^{HVP} = \left(\frac{1}{4\pi^3}\right) \int_{4m_\pi^2}^{\infty} \sigma_H(s) K(s) ds \; . \tag{2.3}$$

$\sigma_H$ is the inclusive cross section for hadron production in $e^+e^-$ collisions, and $K(s)$ is a kinematic weighting factor given by

$$K(s) = x^2\left(1-\frac{x^2}{2}\right) + (1+x)^2\left(1+\frac{1}{x^2}\right)\left[\ln(1+x) - x + \frac{x^2}{2}\right] + \frac{1+x}{1-x}x^2 \ln x \tag{2.4}$$

$$x = \frac{1-\beta_\mu}{1+\beta_\mu} \tag{2.5}$$

$$\beta_\mu = \sqrt{1-4\frac{m_\mu^2}{s}} \; . \tag{2.6}$$

Intuitively, this expression isolates the portion of the vacuum polarization diagram drawn in a box in Figure 2.2(a), splitting the hadronic state in half and connecting the virtual photon to incoming $e^+$ and $e^-$ lines. The resulting real hadron production diagram is shown in Figure 2.2(b). The leading-order behavior of $K(s)$ is proportional to $\frac{1}{s}$; this property is illustrated by the graph of $3\frac{s}{m_\mu^2}K(s)$ versus $s$ in Figure 2.3, which trends asymptotically to 1 for large $s$. The cross section $\sigma_H(s)$ itself also decreases roughly as $\frac{1}{s}$. Consequently,

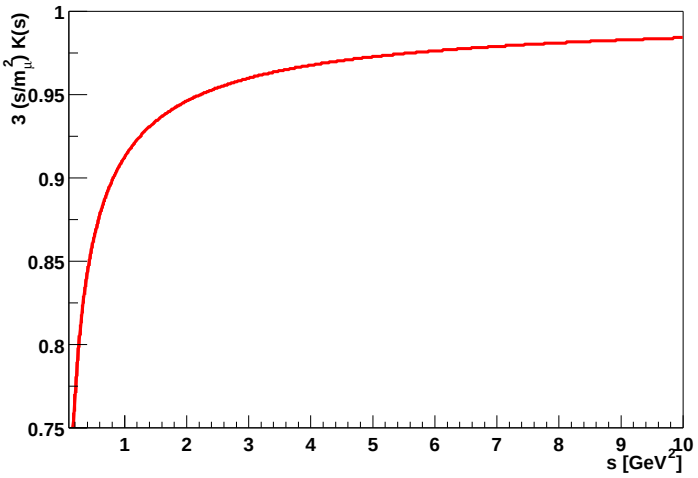



Figure 2.3: Graph of $3\frac{s}{m_\mu^2}K(s)$ versus $s$, illustrating that it asymptotically approaches 1 at high energies.

| Final state | Energies | $\Delta a_\mu^{HVP}$ |
|---|---|---|
| $\pi^+\pi^-$ | $< 1.8$ GeV | $498.85 \pm 4.95 \pm 1.92$ |
| $\pi^+\pi^-\pi^o\pi^o$ | | $16.73 \pm 1.32 \pm 0.20$ |
| $\pi^+\pi^+\pi^-\pi^-$ | | $13.95 \pm 0.90 \pm 0.23$ |
| $\omega$ | | $36.94 \pm 0.84 \pm 0.80$ |
| $\phi$ | | $36.94 \pm 0.84 \pm 0.80$ |
| Others | | $24.01 \pm 1.38 \pm 0.11$ |
| All | $> 1.8$ GeV | $59.44 \pm 1.89 \pm 0.10$ |
| Total | All | $684.7 \pm 6.0 \pm 3.6$ |

Table 2.2: Contributions to $a_\mu^{HVP}$ from various hadronic final states; the values are from [4].

the integral is dominated by low-energy data; more than 90 percent of the entire contribution comes from energies $\sqrt{s} < 1.8$ GeV.

Davier and collaborators have tabulated the world's hadron production cross section measurements and performed the required numerical integration to obtain $a_\mu^{HVP}$ [4]. For energies below 2 GeV, they use exclusive measurements, in which cross sections are determined separately for each final state. The dominant final state, accounting for 73 percent of the total hadronic contribution, is $\pi^+\pi^-$. A number of experiments have published low-energy cross section measurements for this process, but the most precise by far are from the CMD-2 detector at the VEPP-2M collider in Novosibirsk, Russia, which boast a systematic uncertainty of 0.6 percent [28]. At intermediate energies between 2 GeV and 5 GeV they use inclusive measurements of the ratio $R(s)$, the normalized cross-section for hadron production defined by

$$R(s) = \frac{\sigma(e^+e^- \to \text{hadrons})}{\sigma(e^+e^- \to \mu^+\mu^-)} \,. \tag{2.7}$$

Above 5 GeV, they rely on the predictions of perturbative QCD, which are essentially determined by the number of accessible quark degrees of freedom. The contributions of various final states are given in Table 2.2. They obtain the result

$$a_\mu^{HVP;e^+e^-} = 684.7(6.0)(3.6) \times 10^{-10} \ . \tag{2.8}$$

The first uncertainty is propagated from the experimental cross section measurements, and the second refers to the QED radiative corrections that are applied to strip the input data to their bare tree-level values, thereby avoiding double-counting of higher-order contributions. Hagiwara and collaborators performed a similar evaluation [29] based on the same input data, and they obtained

$$a_\mu^{HVP;e^+e^-} = 683.1(5.9)(2.0) \times 10^{-10} \ , \tag{2.9}$$

which is in good agreement. There are also many previous computations of $a_\mu^{HVP}$, but these are the only two that employ the most recent, very precise, data from CMD-2.

It is also possible to determine effective cross sections for the two- and four-pion final states from branching fractions in $\tau$ decay. Assuming that isospin symmetry holds, the conserved vector current (CVC) hypothesis may be used to relate the two-pion production process $e^+e^- \to \pi^+\pi^-$ to the decay $\tau^- \to \nu_\tau \pi^- \pi^o$ shown in Figure 2.2(c). The relationship is [30]

$$\sigma^{I=1}_{e^+e^-\to\pi^+\pi^-} = \frac{4\pi\alpha^2}{s} v_{\pi^-\pi^o} \tag{2.10}$$

The $I = 1$ superscript excludes the small isoscalar component of the cross section that arises from $\omega$ and $\rho$ intermediate states. $v_V$ refers to the $\tau$ spectral function for the final state $\nu_\tau V$, which is defined by

$$v_V(s) = \frac{m_\tau^2}{6|V_{ud}|^2 S_{EW}} \frac{B(\tau^- \to \nu_\tau V)}{B(\tau^- \to \nu_\tau e^- \bar{\nu}_e)} \frac{dN_V}{N_V\, ds} \left[ \left(1 - \frac{s}{m_\tau^2}\right)^2 \left(1 + \frac{2s}{m_\tau^2}\right) \right]^{-1} \tag{2.11}$$

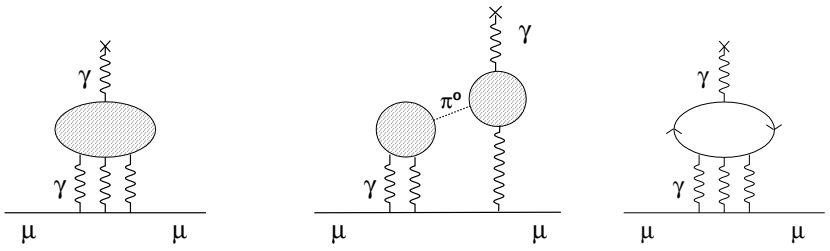
$S_{EW} = 1.0194$ accounts for electroweak radiative corrections. Isospin is not a perfect symmetry; it is clearly broken by electromagnetic interactions and by the mass difference between the $u$ and $d$ quarks. Corrections for isospin breaking must therefore be carefully computed and applied to the data; they are at the level of 3 percent. Davier and collaborators have performed this analysis, which is dominated by data from ALEPH [30]. Their result, which uses $\tau$ data for the two- and four-pion final states and $e^+e^-$ data otherwise, is [4]

$$a_\mu^{HVP;\tau} = 709.0(5.1)(1.2)(2.8) \times 10^{-10} \ . \tag{2.12}$$

The first two uncertainties are defined as they were for the $e^+e^-$ result and the third refers to the isospin-breaking correction.

Unfortunately, this value does not agree with the $e^+e^-$ results. The difference for the $\pi^+\pi^-$ channel on its own is

$$a_\mu^{HVP;\pi^+\pi^-;e^+e^-} - a_\mu^{HVP;\pi^+\pi^-;\tau} = -21.2 \pm 6.8 = (-3.0 \pm 1.0)\% \tag{2.13}$$




(a) (b) (c)

Figure 2.4: Feynman diagrams for (a) the general case of hadronic light-by-light scattering, (b) four-point loop interaction, where the particle in the loop may be a free quark or a meson, and (c) interaction involving two three-point loops connected by exchange of a pseudoscalar such as a $\pi^o$.

which amounts to three standard deviations. In addition, other evidence suggests that the discrepancy is not simply a statistical fluctuation. There is a clear energy dependence in the difference between the cross sections derived from the $e^+e^-$ and $\tau$ data. Consequently, an observable with a different kinematic weighting factor is influenced differently; one such quantity is the total branching fraction $B(\tau^- \to \nu_\tau \pi^- \pi^o)$. It was computed from $e^+e^-$ scattering data via CVC relations and the result was compared against the direct $\tau$ decay measurement; they differ by 4.6 standard deviations. None of the cross-checks determine whether the fault is in the $e^+e^-$ or the $\tau$ data, so for the moment both must be viewed with suspicion.

The current difference between $e^+e^-$ and $\tau$ data is 2 ppm; this is a rather unsatisfying situation, because it complicates the interpretation of a 0.7 ppm experimental result. There is, fortunately, some hope for a relatively quick resolution of the problem. High-energy $e^+e^-$ colliders can use initial-state radiation of a photon to permit a "radiative return" to lower energies, searching for final states such as $e^+e^- \to \gamma\pi^+\pi^-$. This analysis is currently being performed with data from the KLOE detector at Frascati [31] and from the BaBar *B* meson factory [32]. Finally, Blum reports [33] that a lattice QCD calculation is in progress and that it may be possible to improve its accuracy to the level of the usual data-driven approach.

Higher-order hadronic vacuum polarization processes can be related to experimental data in the same way as the leading order. They were most recently computed by Krause, who obtained [34]

$$a_\mu^{HOHVP} = -10.1(0.6) \times 10^{-10} \; . \tag{2.14}$$

## 2.3 Hadronic light-by-light scattering

Hadronic light-by-light scattering, in which four photons couple to a hadronic intermediate state, is shown in Figure 2.4(a). This process cannot easily be related to experimental data. Consequently, it is necessary to calculate this contribution in the context of a low-energy hadronic model. The dominant subdiagram is shown in Figure 2.4(b), where there are two distinct vertices at which a $\pi^o$ couples to two photons.

Consequently, the most important part of the calculation is the determination of the form factor $F_{\pi\gamma\gamma}$. There are, however, additional contributions from similar diagrams with other intermediate states and also from four-point loop diagram topologies as shown in Figure 2.4(c).

There have been two full evaluations of the hadronic light-by-light scattering term. Bijnens, Pallante, and Prades used the Extended Nambu-Jona-Lasinio (ENJL) model [35] for energies below 0.5 GeV. This model integrates out gluonic degrees of freedom in QCD, replacing gluon exchange interactions with local four-quark scattering diagrams. It also expands the Lagrangian as a power series in the inverse of the number of colors $\frac{1}{N_c}$. It preserves the chiral symmetry structure of QCD as well as its spontaneous breaking, though it does not naturally include confinement. The choice of the ENJL model is validated by its prediction of $a_\mu^{HVP}$, which was calculated [36] by de Rafael. He obtained the value $a_\mu^{HVP} = 670 \times 10^{-10}$, which falls within 15 percent of the much more precise result from the data-driven approach. At higher energies, the calculation shifts to a vector meson dominance (VMD) model, in which the fundamental degrees of freedom are vector mesons such as the $\rho$ and $\omega$. The result is [37]

$$a_\mu^{HLBL} = -9.2(3.2) \times 10^{-10} \ . \tag{2.15}$$

Hayakawa, Kinoshita, and Sanda performed a similar calculation. They rely primarily on the hidden local symmetry (HLS) variant of VMD. HLS preserves gauge invariance and chiral symmetry, which simple VMD models sacrifice. They also use the ENJL model for some contributions. They obtained [38]

$$a_\mu^{HLBL} = -5.2(1.8) \times 10^{-10} \ . \tag{2.16}$$

Consequently, there appeared to be two independently derived results which provided mutual confirmation. However, Knecht and Nyffeler rechecked the dominant pion-pole contribution, finding [39]

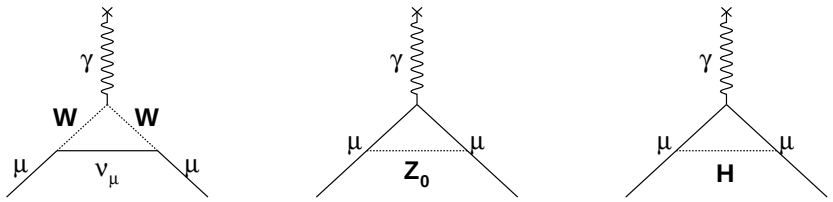
$$a_\mu^{HLBL;\pi^o} = +5.8(1.0) \times 10^{-10} \ . \tag{2.17}$$

This result is of comparable magnitude to previous calculations, but the sign is positive rather than negative. They found that this sign was not dependent on the details of the form factor that they used, but was more generally applicable. Separately, Knecht and collaborators employed effective field theory arguments [40] to argue that the sign must be positive. Following a period of debate, several errata were released. Hayakawa and Kinoshita found that they had misunderstood a phase convention used for the Levi-Civita symbol $\varepsilon_{ijkl}$ in the algebraic manipulation program that they had employed. Following a Dutch tradition [41], it implicitly contained a factor of $i$. The corrected value is [42]

$$a_\mu^{HLBL} = +8.96(1.54) \times 10^{-10} \ . \tag{2.18}$$

Bijnens and collaborators found a sign mistake in their calculation of a normalization factor [43], and corrected their value to

$$a_\mu^{HLBL} = +8.3(3.2) \times 10^{-10} \ . \tag{2.19}$$




(a) Charged current (b) Neutral current (c) Higgs boson exchange

Figure 2.5: Feynman diagrams for the lowest-order weak corrections to $a_\mu$.

Finally, another independent calculation by Blokland, Czarnecki, and Melnikov confirmed [44] the sign of the pion pole part.

Ramsey-Musolf and Wise have criticized the possible model dependence of all of these approaches, preferring a chiral perturbation theory calculation. They obtain [45]

$$a_\mu^{HLBL} = (+1.3^{+5}_{-6} + 3.1\tilde{C}) \times 10^{-10} \,, \tag{2.20}$$

where $\tilde{C}$ is a low-energy constant representing non-leading order terms that have not been calculated. $\tilde{C}$ has a natural scale of order 1, but a value of +3 or -3 would not be unusual, and there is no clear way to further constrain it. It is therefore difficult to extract a useful value from this calculation.

## 2.4 Weak and electroweak processes

The Feynman diagrams for the leading-order weak interaction contributions to $a_\mu$ are illustrated in Figure 2.5. They are topologically similar to the Schwinger diagram in QED, but they contain virtual $W$, $Z$, and Higgs bosons rather than photons. The Higgs exchange diagram in Figure 2.5(c) is included only for completeness; one-loop contributions from the standard model Higgs are entirely negligible. These diagrams were evaluated in the early 1970s and found to give ([46] and references therein)

$$a_\mu^{EW1} = \frac{5G_\mu m_\mu^2}{24\sqrt{2}\pi^2}\left[1 + \frac{1}{5}(1-\sin^2\theta_W)^2 + O\left(\frac{m_\mu^2}{m_{W,H}^2}\right)\right] \tag{2.21}$$

Numerically, $a_\mu^{EW1} = 19.5 \times 10^{-10}$. This value is 1.7 ppm of the QED contribution and is therefore quite significant on the scale of the experimental measurement.

Initially, one would expect that two-loop electroweak contributions would be suppressed by a factor of $\frac{\alpha}{\pi}$ and therefore negligible. However, Kukhto and collaborators found [47] that many of these two-loop diagrams have large coefficients involving a factor of $\ln\left(\frac{m_{W,Z}}{\mu}\right)$, where $\mu$ is a scale much lighter than $m_W$ or

$m_Z$. A significant subset of these processes contain closed fermion loops; in this case, $\mu$ is the mass of a light fermion such as an electron or a $u$ or $d$ quark. Their contributions were evaluated by Czarnecki, Krause, and Marciano [48]. These authors also evaluated the full set of diagrams, including those containing only boson loops [49]. The inclusion of the two-loop diagrams decreased the total electroweak contribution to

$$a_\mu^{EW} = 15.1(0.4) \times 10^{-10} \ . \tag{2.22}$$

The quark loops in these calculations were evaluated with a simple constituent quark model. That is, the quarks were treated as free particles and the effects of hadronization were folded into their masses, which were chosen so that $2m_u + m_d \approx m_p$. While it easy to perform calculations in the context of this model, it does not reproduce properly many of the essential features of low-energy QCD. Because the other types of hadronic terms have proven troublesome recently, this choice was recently re-examined by Knecht, Peris, Perrottet and de Rafael [50] and also by Czarnecki, Marciano, and Vainshtein [46]. Both groups applied more sophisticated hadronic models and found very small changes in the result. The updated prediction is [46]

$$a_\mu^{EW} = 15.4(0.1)(0.2) \times 10^{-10} \ . \tag{2.23}$$

The first uncertainty is related to hadronic physics and the second arises from the Higgs boson mass. Three-loop (and higher-order) contributions were found to be entirely negligible. Indeed, they are even smaller than one would expect because of a fortunate cancellation between the quark and lepton sectors. An uncertainty of $0.02 \times 10^{-10}$ is assigned to account for them.

## 2.5 Summary of standard model contributions

The complete value of $a_\mu$ is

$$a_\mu^{SM} = a_\mu^{QED} + a_\mu^{HVP} + a_\mu^{HOHVP} + a_\mu^{HLBL} + a_\mu^{EW} \ . \tag{2.24}$$

The corrected hadronic light-by-light result from Bijnens and collaborators [43] will be employed because of its more conservative estimate of the systematic uncertainty from model dependence. At this point, it is premature to choose between the two competing values for $a_\mu^{HVP}$, so two standard model predictions will be given:

$$a_\mu^{SM;e^+e^-} = 11\,659\,168.8(7.7) \times 10^{-10} \tag{2.25}$$

$$a_\mu^{SM;\tau} = 11\,659\,193.2(6.8) \times 10^{-10} \ . \tag{2.26}$$

# Chapter 3

# Possible contributions to $a_\mu$ from new physics

The comparison of theoretical and experimental values for $a_\mu$ is interesting, no matter what the outcome. If the values differ, then the comparison provides evidence for physics beyond the standard model. If they agree, then the result constrains any proposed speculative extensions, assuming that there are no fine-tuned cancellations between different varieties of new physics. The expected contributions to $a_\mu$ from many such extensions have been analyzed. This chapter identifies three examples that are of some interest: muon compositeness, supersymmetry, and a muon electric dipole moment, analyzing the effect that each would have on $a_\mu$.

## 3.1 Muon compositeness

The vast spectrum of hadrons is explained by the quark model; in the same way, fundamental "preons" [51] might be able to account for the existence of multiple generations of leptons. The magnetic moments of the proton and the neutron are dramatically different from 2 because they are composed of quarks. One would therefore expect a perturbation to $a_\mu$ if the muon is formed of smaller constituents.

An initial model is represented by the tree-level Feynman diagram in Figure 3.1(a). The associated contribution to $a_\mu$ is linear [53] in the ratio of the muon mass to the characteristic scale $\Lambda = \frac{1}{r}$. Unfortunately, models with such linear terms are untenable. Even the CERN III measurement of $a_\mu$ was sufficient to require $\Lambda > 2000$ TeV. Furthermore, the simple model leads to a self-energy term that would enhance the muon mass to an unphysical value. Any reasonable compositeness model must be constructed to cancel out these effects. One natural way to achieve this cancellation is to build a chirally symmetric wavefunction in which the left-handed and right-handed interaction terms exactly balance [53].

Once the linear contribution has been removed, substructure affects $a_\mu$ in three ways. First, each vertex at which a muon interacts with another particle is multiplied by a form factor $\left(1 + \frac{q^2}{\Lambda^2}\right)$ to account for the spatial extent of the charge distribution. Second, the muon may enter excited states in which the constituents have acquired relative orbital angular momentum. Finally, there may be contact interactions among the constituents that do not correspond to the usual exchange of gauge bosons. Feynman diagrams representative

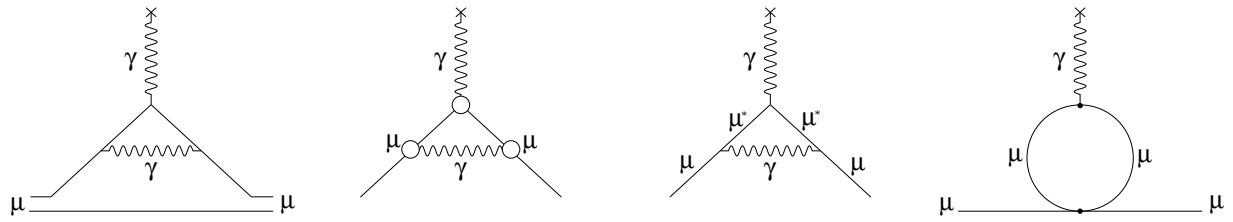



Figure 3.1: Feynman diagrams [52] for (a) the leading-order effect of compositeness, which must be canceled out in a workable model; (b) a form factor at each $\mu\gamma$ interaction vertex; (c) excited lepton states; (d) four-fermion contact interactions.

of each of these categories of effects are shown in Figures 3.1(b)-(d). The numerical results are clearly dependent on the details of the model; however, in each case the contribution to $a_\mu$ is proportional [52] to $\left(\frac{m_\mu}{\Lambda}\right)^2$. It is likely that, after summing over the diagrams, the coefficient is of order 1. Consequently,

$$a_\mu^\Lambda = O\left[\left(\frac{m_\mu}{\Lambda}\right)^2\right] . \tag{3.1}$$

A measurement of $a_\mu$ to 0.7 ppm is then capable of constraining $\Lambda$ at the level of 4 TeV. This limit is comparable to one obtained from studies of the cross sections for $e^+e^- \to \mu^+\mu^-$ and $q\bar{q} \to \mu^+\mu^-$ scattering [51]. However, it has the advantage of being less model-dependent; it makes no assumptions about substructure in particles other than the muon [54].

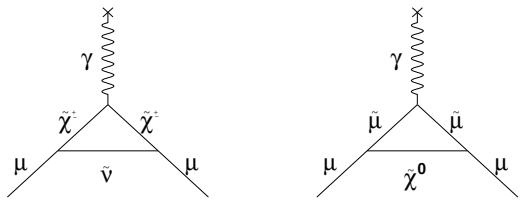
## 3.2 Supersymmetry

The masses of the fundamental particles and the couplings associated with their interactions are free parameters of the standard model, determined by spontaneous symmetry breaking. However, they appear to conspire to give a dramatic cancellation in their effects on the Higgs boson mass. This cancellation is known as the "gauge hierarchy problem." Each fermion $f$ that couples directly to the Higgs causes a radiative correction to its mass of [55]

$$\Delta m_H^2 = \frac{|\lambda_f|^2}{16\pi^2}\left[-2\Lambda_{UV}^2 + 6m_f^2 \ln\left(\frac{\Lambda_{UV}}{m_f}\right) + \cdots\right] . \tag{3.2}$$

The corresponding perturbation for a boson field $S$ has a similar form, but with the opposite sign:

$$\Delta m_H^2 = \frac{\lambda_S}{16\pi^2}\left[\Lambda_{UV}^2 - 2m_S^2 \ln\left(\frac{\Lambda_{UV}}{m_S}\right) + \cdots\right] . \tag{3.3}$$




(a) (b)

Figure 3.2: Feynman diagrams for the lowest-order supersymmetric contributions to $a_\mu$.

In these expressions, $\Lambda_{UV}$ is a high-momentum cutoff parameter that regulates the integration, preventing divergence. It is expected to fall near the Planck scale $m_P = 2.4 \times 10^{18}$ GeV, so the corrections to $m_H^2$ are tremendous even if the couplings $\lambda_f$ and $\lambda_S$ are quite small. These corrections must be reduced by a very large factor because $m_H$ is thought to be of the same order as the $W$ and $Z$ boson masses. It is possible for the contributions from fermions and bosons to cancel, but only if the masses and couplings are chosen very carefully. It would be extremely unlikely for suitable values to fall out by chance.

Supersymmetry provides a way around this fine-tuning problem by postulating a symmetry between bosons and fermions. Every boson has a fermion partner, and every fermion has a boson partner. The theoretical motivation for supersymmetry is compelling, but there is no experimental evidence for the partner particles. The symmetry must therefore be broken by the addition of "soft" terms to the Lagrangian; otherwise, the masses of the paired fermions and bosons would be the same, and the partners would have been seen long ago. The as-yet-undiscovered partners are named by prepending *s-* to the names of the matter fermions; "quark" becomes "squark" and "lepton" becomes "slepton". For the bosons, *-ino* is appended, turning the photon into a "photino" and the $W$ into a "wino."

Supersymmetry leads to two Higgs doublets; one gives mass to the upper half of each generation (the $u$, $c$, and $t$ quarks), and the other to the lower half (the $d$, $s$, and $b$ quarks). Of the four Higgs states, two are electrically neutral, one is positively charged, and one is negatively charged. The ratio

$$\tan\beta = \frac{v_2}{v_1} \tag{3.4}$$

of the vacuum expectation values of the two Higgs doublets is an important parameter describing the nature of supersymmetry. Its scale should be approximately set by the ratio of the masses $\frac{m_t}{m_b} \approx 40$.

The lowest-order supersymmetric contributions to $a_\mu$ are shown in Figure 3.2. They are identical in form to the dominant standard model electroweak corrections. The symbol $\tilde{\mu}$ represents the smuon while $\tilde{\nu}$ stands for the sneutrino. The photino, zino, wino, and Higgsinos are not mass eigenstates; they are rearranged into linear combinations called neutralinos and charginos, represented by $\tilde{\chi}^o$ and $\tilde{\chi}^{\pm}$. To illustrate the scale of these contributions, the masses of the supersymmetric particles can be taken to be degenerate at $\tilde{m}$,

yielding [56]

$$|a_\mu^{SUSY}| \approx \left(\frac{\alpha(M_Z)}{8\pi\sin^2\theta_W}\right)\left(\frac{m_\mu}{\tilde{m}}\right)^2\left(1-\frac{4\alpha}{\pi}\ln\frac{\tilde{m}}{m_\mu}\right)\tan\beta \quad (3.5)$$

$$\approx 13\times10^{-10}\left(\frac{100\text{GeV}}{\tilde{m}}\right)^2\tan\beta\,. \quad (3.6)$$

The experimental precision of 0.7 ppm corresponds to an absolute scale of $8\times10^{-10}$. If $\tan\beta$ is as large as 40, then $a_\mu$ is sensitive to mass scales $\tilde{m}$ of up to 800 GeV.

More realistic supersymmetry models do not assume degenerate masses. The Minimal Supersymmetric Standard Model (MSSM) describes a very general form of supersymmetry breaking. It has 105 independent free parameters–masses, phases, and mixing angles–beyond those of the ordinary standard model [55]. Fortunately, only a small number of them are involved in the calculation of the leading contribution to $a_\mu$. In addition to $\tan\beta$, they are [57]

- $M_1$, $M_2$, and $\mu$, the mass scales associated with the photino, wino/zino, and Higgsino, respectively;
- $m_{\tilde{\mu}_L}$ and $m_{\tilde{\mu}_R}$: the mass scales of the left- and right-handed sleptons; and
- $A_\mu$, a parameter that describes the mixing of left- and right-handed smuons.

Of these, $M_1$, $M_2$, $\mu$, and $A_\mu$ are complex, but their phases are expected to be small. A number of recent analyses have determined constraints [58, 59] on these parameters using the 0.7 ppm result described in this thesis. The set of parameters can be further reduced by postulating a mechanism that is responsible for supersymmetry breaking. For example, by assuming unification with the gravitational force at the Planck scale, one constructs a supergravity model, in which there are only four relevant parameters remaining [60].

## 3.3 Electric dipole moment

A nonzero electric dipole moment (EDM) would cause a precession of the radial component of the muon's spin. To account for an EDM, two new terms [61] must be added to Equation 1.17:

$$\Delta\vec{\omega}_a = -2d_\mu\vec{\beta}\times\vec{B} - 2d_\mu\vec{E}\,. \quad (3.7)$$

The anomalous precession frequency that is measured experimentally is the magnitude $\omega_a = |\vec{\omega}_a|$, which is sensitive to the combined effects of the magnetic and electric dipole moments. The electric field is very small compared to the magnetic field, so the term involving it may be neglected, and the muons are quite relativistic, so $\beta\approx1$. Therefore,

$$\omega_a = B\sqrt{\left(\frac{e}{m_\mu}a_\mu\right)^2+(2d_\mu)^2}\,. \quad (3.8)$$

Neglecting $d_\mu$ leads to a perturbation to the observed value of $a_\mu$, not to its true value. The magnitude of this perturbation is shown in Figure 3.3.

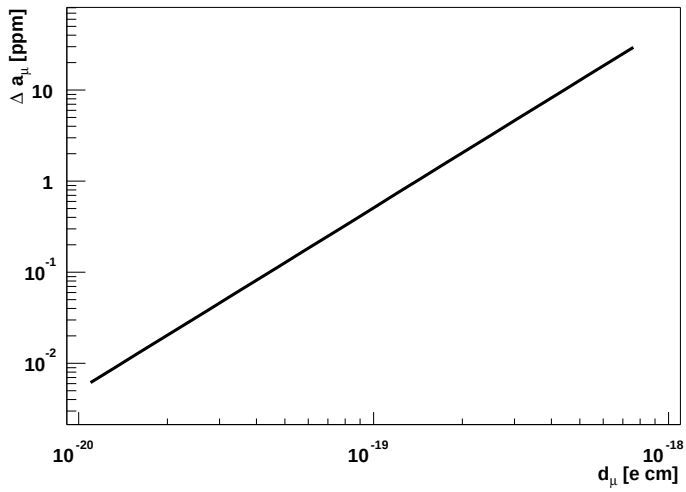



Figure 3.3: Perturbation to the measured value of $a_\mu$ as a function of the electric dipole moment $d_\mu$.

A nonzero EDM for an elementary particle violates both parity ($P$) and time reversal ($T$) symmetries. The EDM operator may be written as [62]

$$\vec{D} = \int \vec{r}\rho(\vec{r})d\vec{r}\,, \tag{3.9}$$

which is clearly a vector quantity. When the particle is at rest, the only vector property associated with the particle is its spin, so $\vec{D}$ must be proportional to the spin. Spin is antisymmetric with respect to $T$ and symmetric under $P$, while $\vec{D}$ is the opposite: symmetric under $T$ and antisymmetric under $P$. $CPT$ is an unbroken symmetry in any quantum field theory, so the $T$ violation is equivalent to a $CP$ violation. In the standard model, the only source of $CP$ violation is the complex phase in the CKM mixing matrix. The resulting EDM has been calculated [63], yielding $d_\mu^{SM} = -3\times10^{-36}\ e$ cm. This value is very small; the first nonvanishing order in this calculation involves diagrams with three loops.

A number of models [62], including supersymmetry, include additional $CP$ violating diagrams that produce an EDM at one-loop order. Still, to the extent that these models treat electrons and muons identically, they are constrained by a simple scaling from the electron EDM. Assuming muon-electron universality, the EDM simply scales in proportion to the mass of the particle. Consequently, the current limit of the electron EDM provides an indirect, model-dependent bound for the muon EDM [61]

$$d_\mu < 9.1\times10^{-25}\ e\ \text{cm}\ (90\%\ \text{CL})\,. \tag{3.10}$$

A model can only produce a larger muon EDM if a violation of the scaling argument can be found. One such path is provided by the lepton flavor violation implied by neutrino mixing, which may lead to a value as large as $5\times10^{-23}\ e$ cm [64].

The current direct experimental limit comes from a search for oscillations in the average vertical direc-

tion of the decay positron momentum in the CERN III $(g-2)$ experiment. It is [65]

$$d_\mu \leq 1.05 \times 10^{-18}\ e \text{ cm } (95\% \text{ CL}) \,. \tag{3.11}$$

A similar analysis of the data from the present BNL experiment is ongoing. It is expected to lead to an improvement in sensitivity of approximately a factor of 5, to the level of $2 \times 10^{-19} e$ cm. The most important consideration in this analysis is the elimination of the false vertical precession signal that results from any offset between the beam centroid and the positron detectors [66].

A dedicated muon EDM measurement has been proposed [67]. Rather than choosing a beam momentum to cancel the dependence on the electric field, it will be chosen to remove the effect of $a_\mu$. Consequently, the only source of spin precession will be from the EDM. The projected sensitivity of this experiment is around $10^{-24}$ $e$ cm, which is comparable to the indirect limit in Equation 3.10.

## 3.4 Other "new physics" effects

It must be emphasized that the three topics addressed in this chapter are by no means the only models that have been analyzed in terms of $a_\mu$. Supersymmetry was chosen as an example because it is rather popular among theorists, and the other models are simple and pedagogically interesting. Other possibilities include leptoquarks, radiative muon mass generation, anomalous gauge boson properties, and compact extra dimensions. A review of some of these scenarios has been written by Czarnecki and Marciano [56].

# Chapter 4

# Experimental apparatus

This chapter describes the equipment that is used to measure $a_\mu$ in BNL experiment 821. It discusses the preparation of the muon beam, the storage ring in which it is trapped, and the detectors and electronics that observe its decay. A comprehensive guide to the experimental hardware is the technical design report [68]. Detailed papers have been prepared about many of the individual components, and references to these works are provided where available.

## 4.1 Accelerator, target and beamline

The Alternating Gradient Synchrotron (AGS) [69] at BNL was used to accelerate protons to an energy of 24 GeV. The beam was divided into twelve bunches, which were extracted one at a time into the primary beamline at 33 ms intervals. Each of these bunches was used to fill the storage ring with muons. The cycle time for acceleration and extraction was 2.7 s; each bunch had an root mean square (RMS) width on the order of 25 ns. The beam intensity, which increased over the course of the run, reached $5 \times 10^{12}$ protons per fill.

The beamline is shown in Figure 4.1. The protons are transported to a pion production target, which

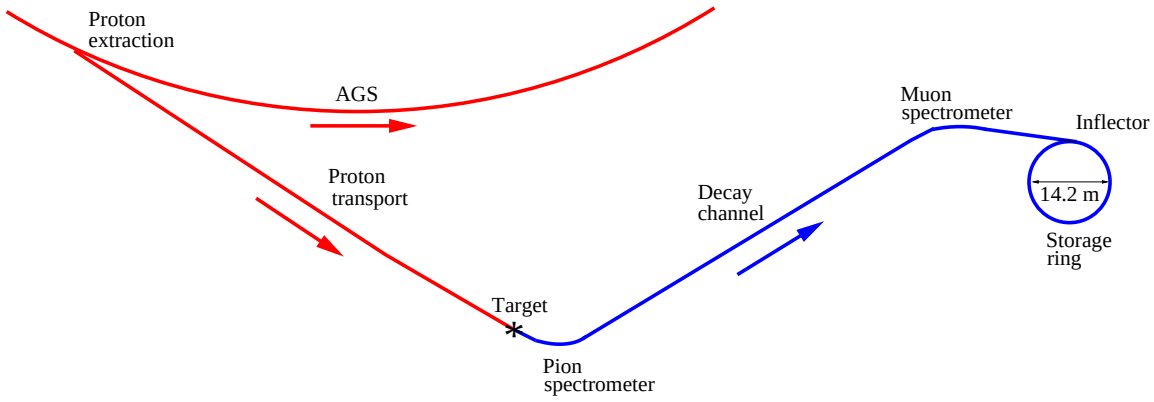



Figure 4.1: Diagram of the primary and secondary beamlines leading from the AGS to the experiment.

is a set of rotating water-cooled nickel cylinders. The secondary beamline consists of two spectrometers separated by a long straight channel. The first spectrometer selects positively-charged pions at 1.7 percent above the "magic momentum" of 3.09 GeV/$c$. The pions then enter a 72 m straight channel where approximately half decay by $\pi^+ \to \mu^+\nu_\mu$. Because all neutrinos are left-handed, conservation of angular momentum implies that the resulting muons are highly polarized. The kinematics of the decay lead to muons whose energy is less than that of the parent pion. The last stage of the beamline is a second spectrometer that selects particles at the magic momentum, so most of the residual pions are removed at this point. A Čerenkov counter was filled with isobutane at a range of pressures to measure the concentration of various particle species at the end of the beamline. At this point, the beam is approximately 30 percent muons, 30 percent pions, 30 percent positrons, and 10 percent protons [70]. The proton fraction is an estimate rather than a measurement, since the counter was not able to reach the pressure that would be required.

Extraneous particles from the other bunches circulating in the AGS are prevented from reaching the storage ring during the measuring period after the main beam pulse by a sweeper magnet placed in the pion decay channel. It is turned off at the nominal injection time, then immediately ramped up to a field of 50 mT over the following 15 $\mu$s. It decays with an $RC$ time constant of 700 $\mu$s. The peak field is sufficient to give a 5 mrad angular deflection, which reduces the transmission of the beamline by a factor of approximately 10. The sweeper magnet was not installed until midway through the run; approximately 20 percent of the data set was collected before it had been commissioned.

## 4.2 Storage ring

The muon storage ring [71] is a C-shaped magnet with a dipole field of 1.45 T and a radius of 7.11 m. It has a steel yoke that is excited by niobium-titanium superconducting coils that are cooled to 4.2 K and operated at a current of 5200 A. A cross section view of the magnet appears in Figure 4.2. The uniformity of the magnetic field is at the level of 1 ppm over the muon storage region after averaging over the azimuthal coordinate. The yoke steel is of conventional quality, but the pole pieces are vacuum-cast from extremely pure iron, with a carbon content of less than 0.004 percent. The magnet poles are partially decoupled from the yoke by a 1.5 cm air gap, in which wedge-shaped iron shims are placed. These shims were positioned and the thickness of the edges of the poles was adjusted in order to minimize higher multipoles of the field. The final handle was provided by currents on a printed circuit board attached to the surface of the poles. Also, the use of ferromagnetic material in equipment located near the storage region was systematically avoided.

The beam circulates within an aluminum vacuum chamber that is placed in the gap between the magnet poles. The vacuum chamber is scalloped so that the decay positrons encounter a minimal thickness of material along their path to the detectors. Consequently, the decay positrons lose less energy than they would with a conventional design like the one used in the CERN III experiment. More importantly, the amount of energy lost is much less correlated to the decay angle, so the effective energy resolution of the system is improved. The vacuum was maintained in the range of $10^{-7}$ torr by a set of turbomolecular, ion, and cryogenic pumps. This level of vacuum was necessary to avoid spontaneous discharges of the

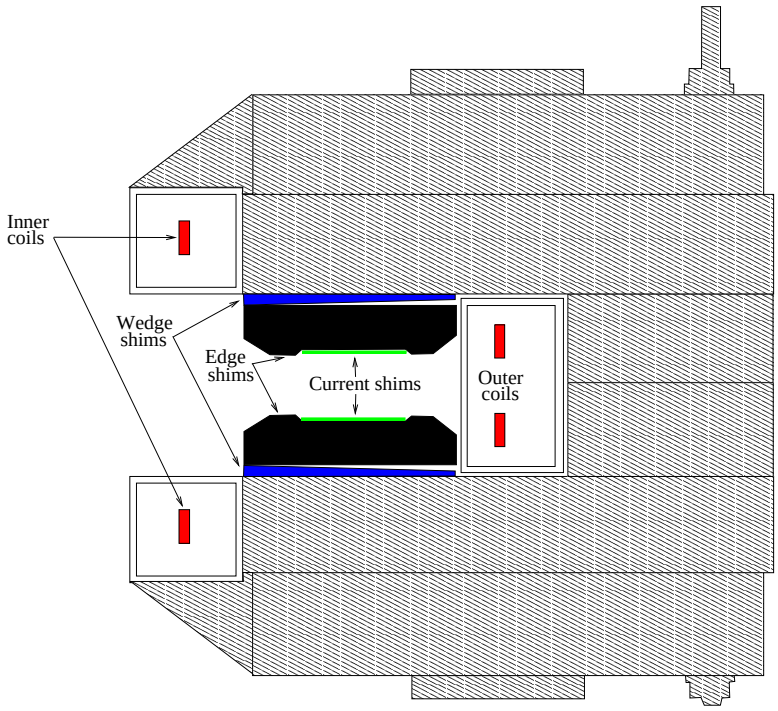



Figure 4.2: Cross section of the storage ring magnet, indicating the positions of the coils and the various shimming mechanisms. The size of the edge shims is greatly exaggerated.

high-voltage devices that are installed inside the vacuum chamber.

The beam is injected into the storage region through an inflector magnet [72]. The inflector is 1.7 m in length, and it is placed inside a hole in the yoke, where it cancels out the field of the main magnet so that the beam can enter the ring without being deflected. The inflector's superconducting coils are arranged in a truncated double $\cos\theta$ geometry [73], which efficiently traps most of its fringe field. The remainder of the flux is contained by a continuous sheet of type II superconductor. This shield is cycled above and below its critical temperature after the main magnet has been energized but before the inflector coils are ramped up, "pinning" the field lines on its exterior. The construction of the inflector presented a significant technical challenge that required compromises. First, the magnet coils are closed on both ends, so the beam must pass through this material, causing significant losses. Second, it was not possible to make the inflector aperture large enough to fill the accepted phase space of the storage ring. The position of the inflector is shown in Figure 4.3, along with the locations of the other devices described in this section.

The injection point at the end of the inflector is at a radius 76 mm larger than the central orbit. A magnetic kicker [74] is fired as the beam is injected in order to place it onto a stored orbit; otherwise, it would simply make one turn and collide with the inflector. The kicker consists of three 1.76 m long current loops, each independently driven with a 4200 A pulse of 400 ns full width. The cyclotron period of the ring is 149.2 ns, so the required 10 mrad kick is delivered over a number of turns. Each of the electrical modulators is effectively an *RLC* circuit, charging a capacitor to a potential of 95 kV and discharging it at

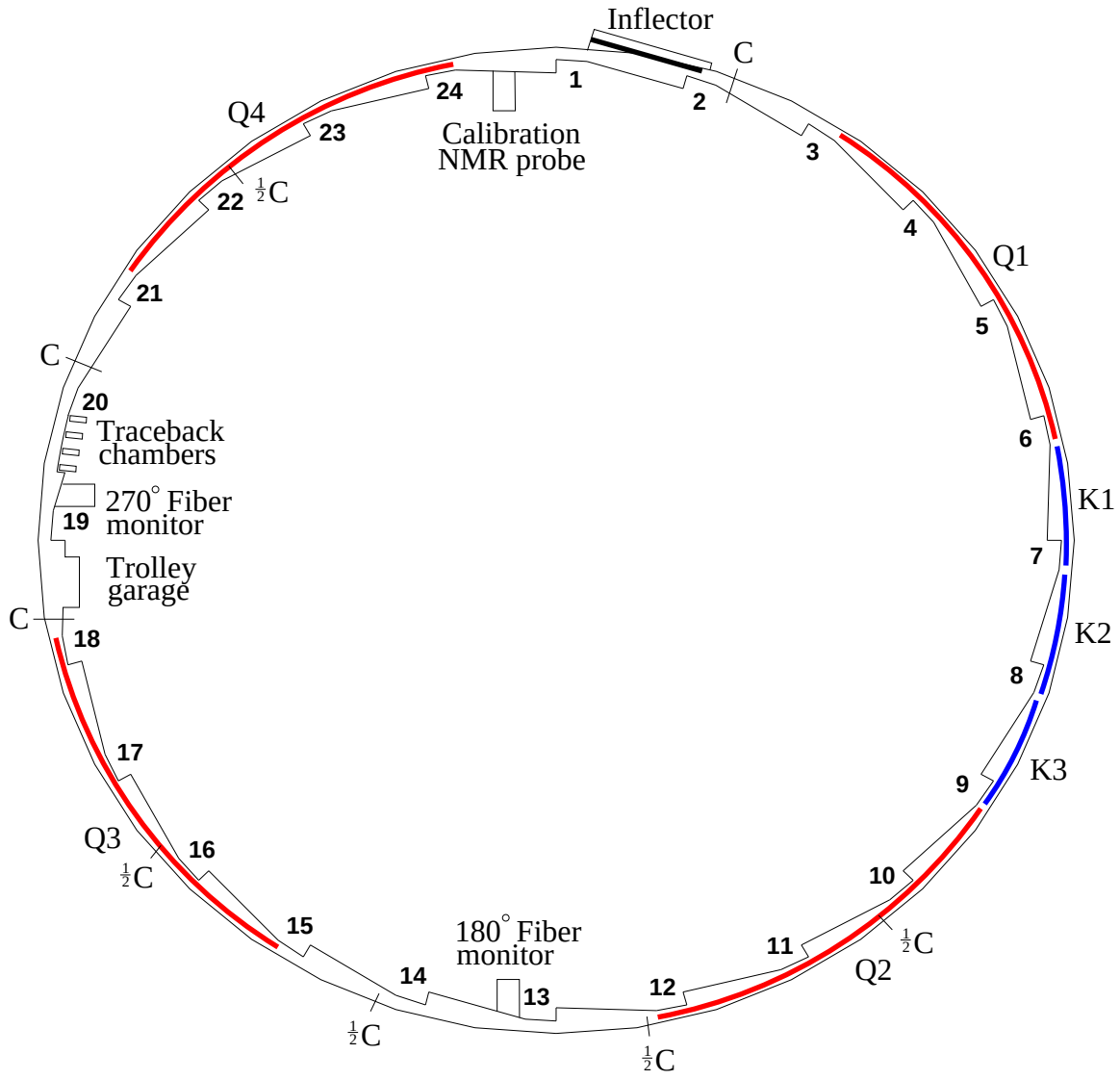



Figure 4.3: Overhead view of the equipment in the storage ring. The numbers 1 through 24 indicate the positions of the detector stations. Q1-Q4 are the electric quadrupole electrodes and K1-K3 the muon kicker plates. C is a collimator, and $\frac{1}{2}$C is a half-collimator.

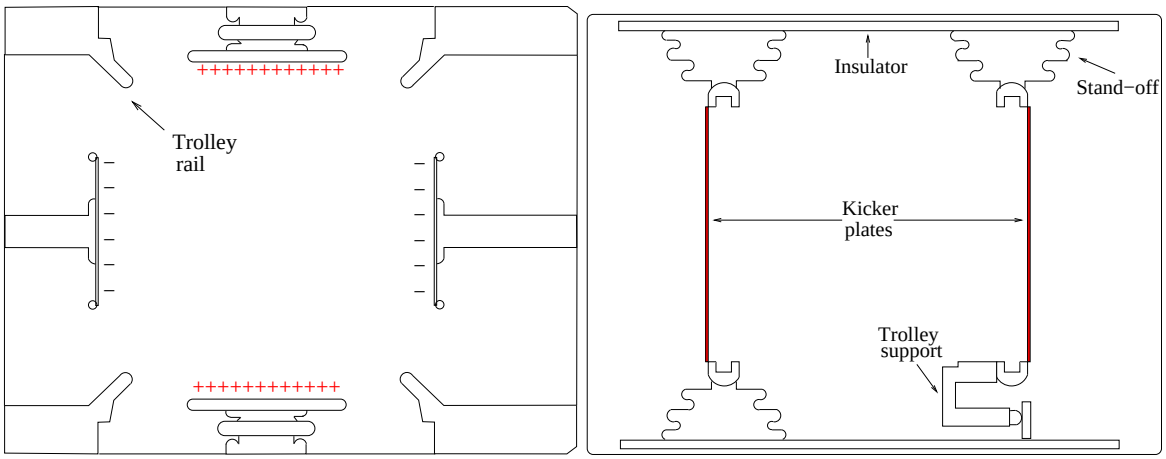



Figure 4.4: Placement of (a) the electric quadrupole plates, and (b) the kicker magnet plates, in the vacuum chamber.

the proper moment through the kicker plates. The high voltage is switched by deuterium thyratron tubes. The placement of the kicker plates within the vacuum chamber is shown in Figure 4.4(b).

Circular collimators are suspended inside the vacuum chamber at eight points around the ring, defining the storage aperture to have a diameter of 9 cm. The collimators are made of copper and are only 3 mm thick; this thickness is a small fraction of a radiation length. Consequently, the primary effect of the collimators is through multiple scattering. When a muon strikes a collimator, it receives a random kick and is typically lost from the ring several turns later. To minimize interference with the beam on its first turn, while it is still offset from the ideal central orbit, either the inner or outer half of many of the collimators has been removed. There are three full collimators, four outer half-collimators, and one inner half-collimator [75, 76].

In a perfect magnetic dipole field all orbits are circular, so there is no need for an additional confining field in the radial dimension. A particle arrives back at its initial radial position after a turn regardless of its initial conditions, provided only that they are not so extreme as to send it into an obstruction along the way. However, a muon with a nonzero initial vertical angle will continue to spiral vertically without bound. Consequently, a vertically focusing electric quadrupole field is created [77] by applying a potential of $\pm$ 24 kV to electrodes secured inside the vacuum chamber as illustrated in Figure 4.4(a). The quadrupole electrodes cover 43 percent of the circumference of the storage ring. The applied charge configuration is clearly defocusing in the radial dimension; the resulting beam dynamics will be discussed in Chapter 5. The voltage on the electrodes is pulsed; they are only charged for a period of about 1 ms beginning just before the beam is injected. This periodic discharge releases trapped low-energy electrons that circle in tight orbits in the electric and magnetic fields, ionizing the residual gas in the vacuum chamber and eventually causing a spark.

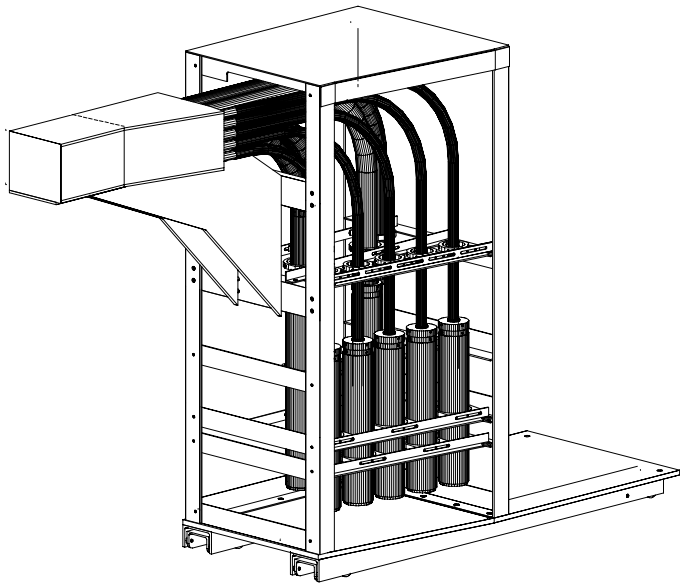


Figure 4.5: Mechanical drawing of a detector station, including the calorimeter block, light guides, photomultiplier tubes, and mechanical support structure. (Reproduced from [78].)

## 4.3 Positron detectors

The laboratory energy of the decay positrons is measured using lead/scintillating fiber calorimeters [78]. An electromagnetic shower is created as the positrons pass through the dense lead plates. Plastic scintillating fibers, glued in a honeycomb pattern in grooves that were cut in the plates, produce light as charged particles pass through them. The light is captured by four acrylic light guides and transmitted to four photomultiplier tubes positioned below the plane of the storage ring. In response to the incident light, the photomultipliers produce an electrical signal. The photomultipliers are shielded from the magnet's fringe field by several layers of steel and $\mu$-metal. A detector station is shown in Figure 4.5. There are twenty-four of these stations spaced at equal intervals around the inside of the storage ring.

When the beam is injected into the storage ring, there is a prompt flash of light in the detectors. This burst of activity arises from particles that scatter from the inflector or the magnet steel, pions that decay shortly after entering the storage ring, and positrons that lose energy through synchrotron radiation and spiral into a detector. To protect the photomultipliers from being damaged by the flash, they are gated off for several microseconds after injection by reversing the voltages applied to two of the dynode stages. A few detectors positioned near the inflector have to be gated off for as long as 45 $\mu$s because of delayed effects of the injection losses, which liberate neutrons from nuclei in the magnet steel. These neutrons diffuse, thermalize, and are captured, producing a large number of low-energy photons in the region around the detectors.

The final calorimeter design was characterized in June 1998 in a test beam area at the AGS. At this time, the accelerator was operated in its resonant extraction mode; the beam consisted of a slow stream of particles over two seconds rather than a sequence of tight bunches. The beamline selected a narrow range of

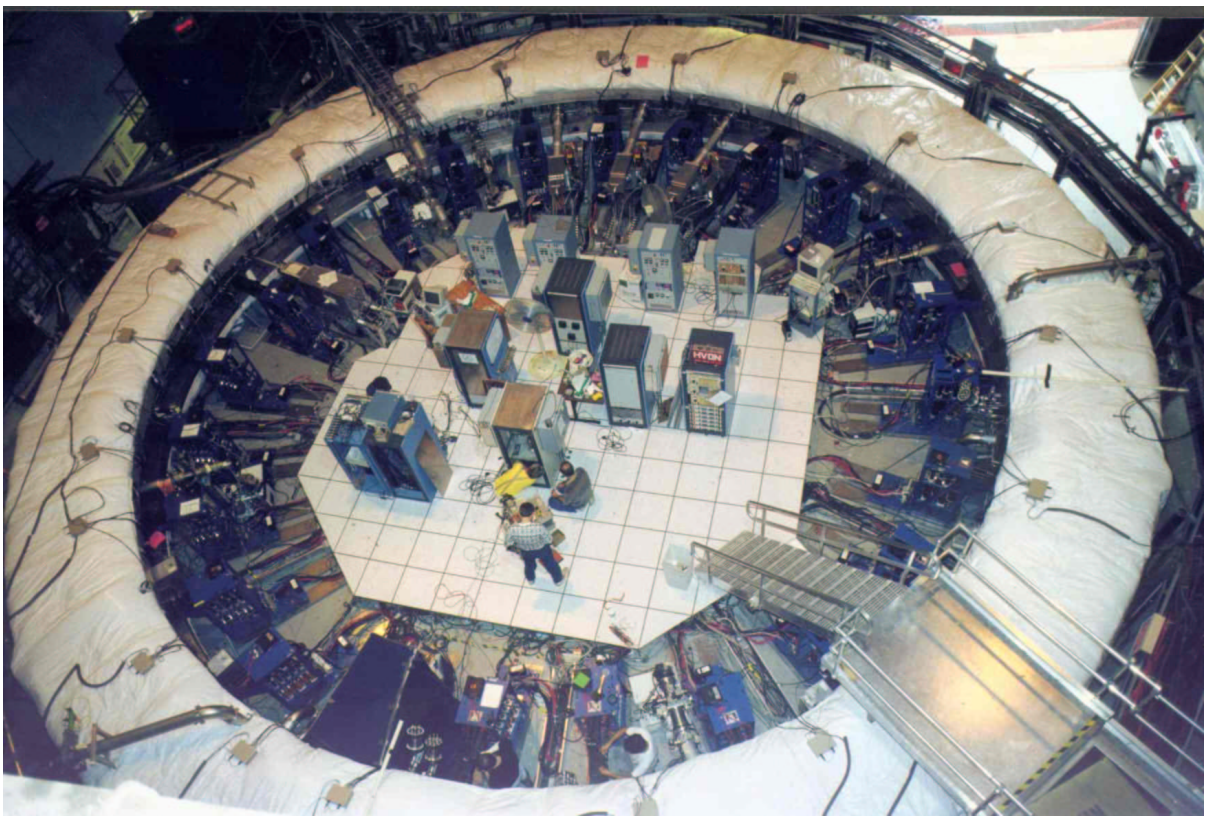


Figure 4.6: Photograph showing the storage ring from overhead. As a reference point, station 16 is located immediately under the stairs. (Photograph by K. Jungmann.)

momenta with width $\Delta p/p \approx 3$ percent. An air Čerenkov counter was used to tag electrons, separating them from the pions that were also in the beam. The charge output was measured as a function of beam energy; it deviated from a linear form by less than 1 percent up to 3 GeV, falling short by 2.5 percent at 4 GeV. The detector energy resolution was also measured, yielding approximately $8.0\%/\sqrt{E\ [\mathrm{GeV}]}$.

Eleven of the stations are equipped with front scintillating detectors (FSDs). These devices consist of five 2.8 cm $\times$ 23 cm strips of scintillator with a thickness of $\frac{3}{8}$ in. The strips are arrayed in a radial orientation over the front of the calorimeter block, providing information about the vertical position of entering particles. Each strip is coupled through an acrylic light guide to a photomultiplier.

Similarly, position sensitive detectors (PSDs) are mounted on five of the stations [79]. Like the FSDs, these devices are arrays of scintillator tiles, but with much finer segmentation. Each PSD tile is 7 mm wide and 8 mm thick; they are arranged into two planes. The first layer measures in the horizontal dimension with 33 tiles, and the second layer measures in the vertical with 20 tiles. Each tile is coupled through a wavelength-shifting optical fiber to a part of a multi-channel photomultiplier or hybrid photodiode. Consequently, this large set of channels can be read out by a single device. Unfortunately, difficulties in understanding efficiencies and noise problems limited the PSDs to a diagnostic role.

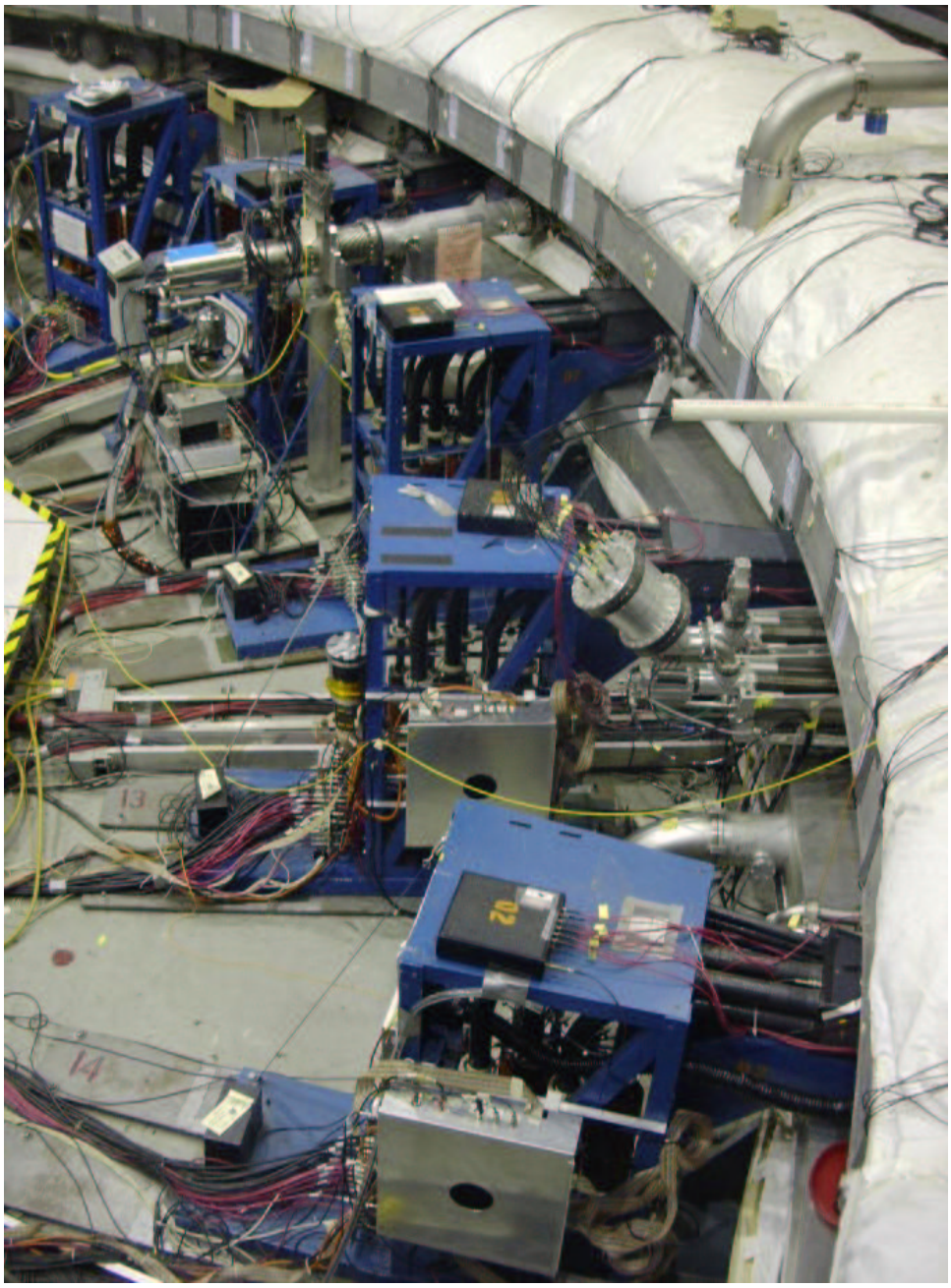



Figure 4.7: Photograph showing a portion of the inside of the storage ring including detector stations 10 through 14. The device located between stations 11 and 12 is a cryogenic vacuum pump; the protrusion between stations 13 and 14 is a fiber beam monitor. The black boxes on top of each detector are the final splitters for the laser fibers. The aluminum boxes mounted on the sides of stations 13 and 14 contain the PSD preamplifiers and discriminators. The cabling on top of the white thermal insulation connects to the fixed fixed NMR probes. (Photograph by G. Onderwater.)

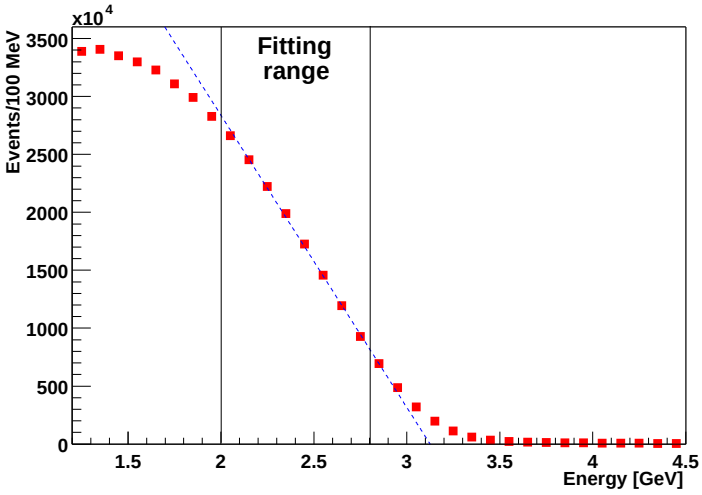



Figure 4.8: Typical measured calorimeter energy distribution with an endpoint fit superimposed.

## 4.4 Calorimeter calibration

The initial calibration of the photomultiplier gains was performed as a part of the June 1998 test beam program [80]. Each calorimeter was positioned end-on so that the beam struck the center of a single quadrant. The gains were iteratively adjusted to give the same average pulse area in each photomultiplier tube. The responses of the top and bottom halves in the usual side-on position were then verified.

A blue light-emitting diode (LED) is located near the joint between the calorimeter block and the light-guides so that it shines into all four of them. The fraction of light that goes to each quadrant is determined entirely by its physical position and orientation, which are stable on long time scales. This LED was used to check the calibration after transporting the detectors from the test beam area to their data-taking positions inside the storage ring. It was then used to correct the photomultiplier gains for the effect of the fringe field of the storage ring magnet. Most stations did not require any significant adjustment, but a few required gain changes of up to 15 percent.

Very precise calibration information can be determined *in situ* from the positron energy spectra. A typical spectrum is shown in Figure 4.8. The indicated portion of the falling edge of this spectrum, in which the number of counts declines from 80 percent to 20 percent of its peak value, can be fit by a straight line. This line is then extrapolated to find the endpoint, the point at which the line intercepts the horizontal axis. This point corresponds to approximately 3.1 GeV. In the August 1998 run, the endpoint spectra for the top and bottom halves of the calorimeters were determined; the RMS of the distribution of detector gains was 2.7 percent. Consequently, the use of the LED to maintain the calibration was effective at this level.

The final calibration was performed at the beginning of the 2000 running period using the energy spectrum method. Data were collected for each of the four photomultipliers on each station separately; the high voltage supply to the others was turned off. The resulting endpoints for all of the front tubes (those closest to the vacuum chamber scallop) were averaged, and this average was defined to be the ideal endpoint. The

gains of all of the front tubes were then iteratively adjusted to match the ideal value. The same procedure was used to calibrate the back tubes to their own ideal value.

It is necessary to establish that the detector gain calibration remains constant as a function of time after injection. The stored muon intensity and the neutron background both change from early to late times, and the photomultiplier gating circuit introduces a clear potential source of gain shifts. A simple estimate shows that an uncorrected gain change of 0.2 percent shifts the fitted value of $\omega_a$ by approximately 0.1 ppm; the exact value depends on the details of the functional form of the gain change and the background terms in the form used to fit the time spectrum. A pulsed ultraviolet nitrogen laser is located outside the experimental area. The light from this laser is split in several stages into many optical fibers. Two of these fibers lead to reference photodiodes in quiet locations. The others are routed to the calorimeters and FSDs. A fiber is coupled to each segment, where the ultraviolet light excites the scintillator much as a charged particle would. For about twenty minutes out of each eight hour shift during the data collection period, the laser was fired every other fill in parallel with the beam data. Firings alternated among four points in time with respect to beam injection. These points were changed for each laser run, providing a map of any possible gain changes. At the same time, any rate-dependent shifts in the determination of the pulse time were also measured. No overall trends were seen in the gain change data; from early to late times, the observed gain changes in each detector were generally a few tenths of a percent, without a preferred sign. Unfortunately, the scatter in the data points is much greater than is allowed by statistics, and the responsible mechanism has not been identified. Consequently, the gain stability must be established using the ordinary positron energy spectrum rather than the laser data. On the other hand, timing shifts were clearly limited to less than 4 ps from early to late times, with none of the difficulties that affected the gain analysis [81]. This level of timing shift corresponds to a 0.02 ppm systematic uncertainty in $\omega_a$.

## 4.5 Magnetic field probes

The magnetic field is measured using pulsed nuclear magnetic resonance (NMR) [82]. This technique effectively measures the spin precession frequency $\omega_p$ of protons at rest in the field; this frequency is related to the magnetic field by

$$\omega_p = \mu_p B \, . \tag{4.1}$$

In the ground state, the protons are polarized parallel to the applied field. An excitation pulse at a reference frequency near $\omega_p$ is applied to the sample, rotating its polarization by about 90 degrees. The oscillations of the magnetization are then measured as it relaxes back to the ground state.

The details of this apparatus will be described in Chapter 9 in parallel with the analysis of the data that it generated. In brief, the heart of the system is a trolley containing 17 NMR probes that is driven around the interior of the vacuum chamber every few days during the data collection run. The geometry of these probes permits the determination of the multipoles of the field in the muon storage region up to the decupole. They are calibrated at the beginning and end of the data collection period against an absolute standard probe [83]. Between trolley runs, the field is continuously tracked by 360 fixed NMR probes that are scattered around

the storage ring. They reside outside the storage region, between the pole tips and the vacuum chambers.

A subset of the fixed probes provides input to a real-time feedback system that controls the magnet power supply. This system maintains a constant average dipole field at the level of 1 ppm as a function of time, correcting for slow drifts that might be caused by temperature or line voltage changes. For a period of a few days during the data collection period, this feedback system failed, and excursions of up to 50 ppm in the average field were observed.

## 4.6 Beam monitoring detectors

There are a number of diagnostic detectors placed along the beamline. Calibrated ionization chambers are used to measure the beam intensity at various points, and segmented wire chambers give its horizontal and vertical profiles. Information from these detectors was used to iteratively tune beamline magnet settings to optimize the transmission. The final detector encountered by the muon beam before it enters the storage ring is known as the $t_0$ counter. This device, positioned just outside the hole in the magnet yoke, is simply a scintillator paddle coupled to a low-gain photomultiplier. Its output is used to determine the start time of each fill; all decay times are defined to be relative to it.

A set of four straw tube wire chambers positioned in front of detector station 20 allows a subset of the decay positrons to be traced back into the storage region [84]. Each chamber has three layers of straws that measure in the horizontal dimension followed by three layers that measure vertically. The decay vertex cannot be identified for any individual event, so tracking proceeds to the point of maximum radius. The resulting distribution can be related statistically to the true muon distribution.

Scintillating fiber beam monitors can be plunged into the storage region to measure the muon distribution directly at early times after injection. The fibers are 0.5 mm thin; there are seven of them oriented vertically and seven horizontally, spaced at intervals of 1.3 cm. These measurements cause the beam to be scattered and lost quickly, so they are only available for occasional special studies. The monitors are located at two strategic positions in the storage ring, at 180 and 270 degrees from the injection point.

## 4.7 Electronics and data acquisition system

The output signals from all of the photomultipliers for the calorimeters and FSDs are transmitted through RG-213 coaxial cable from the detector to the counting room. The cables are approximately 50 m in length; they were cut to fit, so the length varies from one station to another. The analog signals from the PSDs are very small; to avoid noise pickup, they are discriminated with electronics located adjacent to the station, and digital signals are transmitted to the control room.

Two types of electronics are used to digitize and store nearly all of the experimental data: waveform digitizers (WFDs) and multihit time-to-digital converters (MTDCs). Both of these devices were constructed or modified for the experiment by the Boston University Electronics Design Facility. They are VME modules, so their data can be transferred over a backplane using a standardized protocol [85]. A small amount of

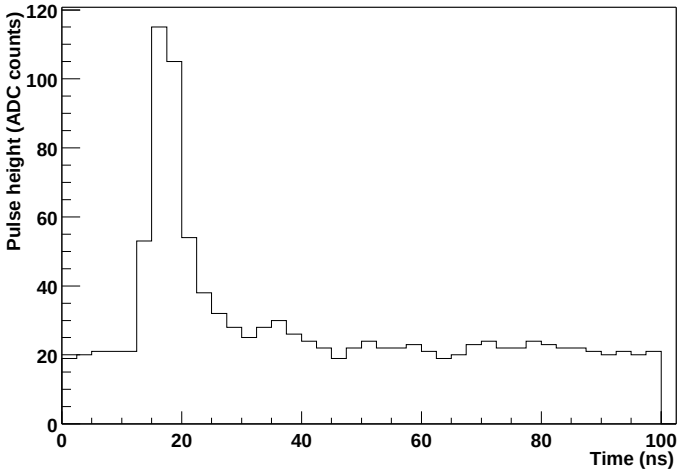



Figure 4.9: Digitized signal from a calorimeter for a typical 1.8 GeV positron.

data, mostly related to an alternative readout system for the laser calibration data, is recorded using CAMAC ADCs and TDCs.

Each WFD module [86] contains four independent flash analog-to-digital converters (ADCs) that sample the input signal every 5 ns. However, most are operated in a bridged mode in which the same input signal is recorded by two channels on opposite phases of the 200 MHz clock, leading to an effective sampling period of 2.5 ns. The bridging card performs a sum of the signals from its five analog inputs and distributes it to the flash ADCs; it also contains a discriminator circuit with a programmable threshold for each input and for the sum. The samples are written to an onboard memory; 64 kB of memory is provided for each channel. Half of this memory is used to store ADC samples; the other half holds the associated times and the input discriminator status. A zero-suppression feature selects regions of time around pulses that trigger the sum discriminator; samples outside these regions are not recorded.

The WFDs are the primary recording instrument for the calorimeters. Four of the inputs on the bridging card are used for the analog signals from each of the quadrants of a single calorimeter. The fifth input is used for a triangle-wave marker pulse that is triggered just before each fill to permit the data from the two bridged channels to be aligned. A typical digitized calorimeter pulse appears in Figure 4.9. WFDs are used in a similar fashion to store signals from other sources as well, including the $t_0$ counter, the reference photodiodes in the laser calibration system, and the muon kicker waveform. They are used in non-bridged mode to digitize the electric quadrupole waveforms and also the signals from the fiber beam monitors when they are inserted.

Each MTDC [87] has 24 independent digital inputs. In response to a logic pulse, it records the time in 1.25 ns units together with the channel number. The boundaries of the time bins are potentially rate-dependent, so some channels are connected to external derandomizer circuits that place the input pulse in the center of a 5 ns bin in a stable way. The MTDC is the primary digitization mechanism for the FSDs and

the traceback chambers. The calorimeter signals are also connected to MTDCs, but these data have been almost completely unused.

The WFDs and MTDCs are divided among six enclosures ("crates"), each of which also contains a single-board computer and a VME-to-VME interconnection module. Approximately 60 $\mu$s before every fill, all of the electronics is armed by a start signal derived from the AGS control system. A timer generates a stop signal 1200 $\mu$s later, disarming the electronics. At this point, the data from each of the WFDs and MTDCs is copied to the computer in the same crate. Between accelerator cycles, the data for the 12 fills is transferred over the VME-to-VME interconnect from each crate to an event builder computer, which writes it to Digital Linear Tape (DLT) media. A fraction of the accelerator cycles, typically 10 percent, is also sent to a computer workstation for immediate online analysis. This data acquisition system, which is described more fully in [88], uses software that is based on UNIDAQ [89].

The output from the NMR probes in the magnetic field measurement system is recorded with a completely independent data acquisition system. Both systems record a timestamp from a common 100 Hz timer that permits the data to be correlated later. All of the electronics used in both systems is locked in phase to the same reference clock, which is derived from a Stanford Research Systems FS700 LORAN-C receiver. Its specifications claim long-term stability at the level of $10^{-12}$ and short-term stability of $10^{-10}$. Even better, the clock synchronization between the two parts of the measurement means that any drift outside of these specifications would cancel out in the ratio $\frac{\omega_a}{\omega_p}$.

# Chapter 5

# Beam dynamics

During the measuring period, the phase space distributions of the stored muon beam are not static in either the longitudinal or the transverse dimension. This chapter explores some aspects of their dynamics. It does not attempt to provide a complete course in accelerator physics; only those aspects of the subject that may affect the measurement of $\omega_a$ are considered.

## 5.1 Cyclotron motion

In the absence of any focusing forces, muons circle the storage ring at a radius that is determined by their momentum:

$$r = \frac{p}{eB} \, . \tag{5.1}$$

The cyclotron period is then

$$\tau_c = \frac{2\pi r}{\beta} = \gamma \frac{2\pi m_\mu}{eB} \, . \tag{5.2}$$

The last step follows from the relativistic definition $\vec{p} = \gamma m \vec{\beta}$. Somewhat surprisingly, higher-momentum particles take longer to complete an orbit than lower-momentum particles. This effect is relativistic; all particles move at essentially the same speed $c$, so the primary effect of additional momentum is an increased radius of curvature. At the magic momentum and for the magnetic field used in the experiment, $\tau_c = 149.2$ ns.

The beam is injected in a bunch whose RMS width is of order 25 ns, so it initially occupies only a small portion of the available longitudinal space. Because the beam contains a range of momenta $\Delta p/p \approx 0.6\%$, the bunch spreads out over time, eventually filling the storage ring in azimuth. This debunching occurs with a characteristic time constant of approximately 30 $\mu$s. The effect of the bunched beam on the observed positron spectrum is shown in Figure 5.1. At early times after injection, there is clearly a large modulation. This so-called "fast rotation" adds a small inconvenience to the determination of $\omega_a$, but it permits the radial distribution to be measured from the data. The magnetic field varies over the storage volume, so measurement of this distribution is essential to a precise determination of the average magnetic field experienced by a muon. Two methods have been employed to extract this distribution. One

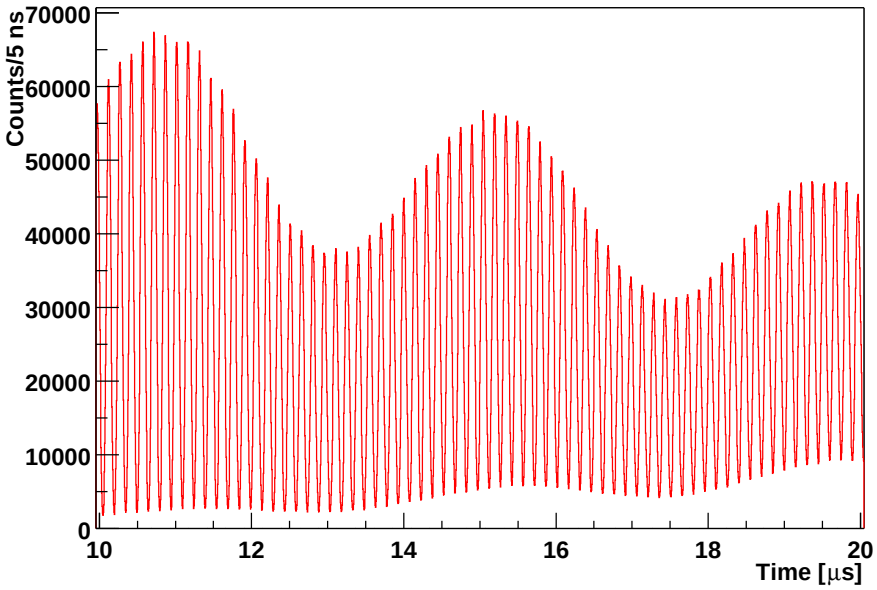



Figure 5.1: Count rate as a function of time soon after injection in detector 24, illustrating the bunched nature of the injected beam. The fast rotation structure is visible together with an envelope determined by the slower modulation from the $(g-2)$ precession.

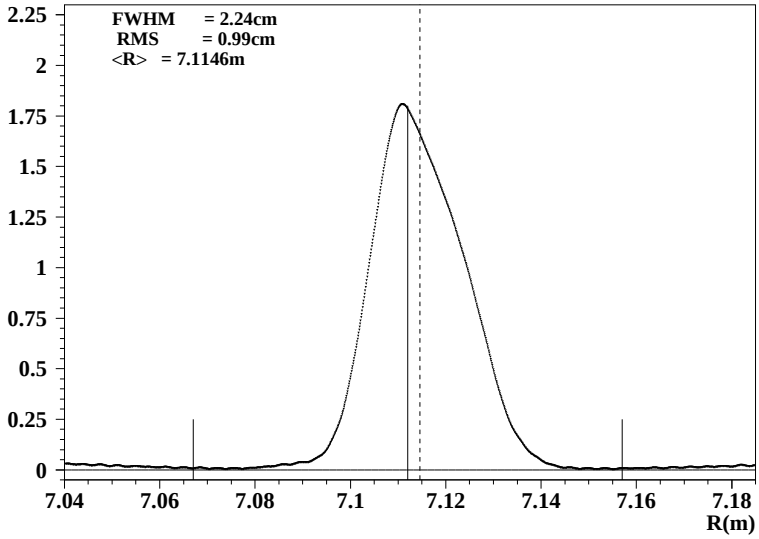



Figure 5.2: Radial muon distribution determined from fast rotation analysis using the Fourier transform method. The centroid of the distribution is at 711.46 m. The lines indicate the edges of the storage region defined by the collimators. (Reproduced from [90].)

approach uses a variant of the Fourier transform [91] that compensates properly for the missing data while the detectors are gated off. The other method performs a $\chi^2$ minimization between the measured time spectrum and the prediction of an assumed momentum distribution. The two techniques agreed on the radial profile shown in Figure 5.2 whose mean value is 711.46 m, slightly different from the central radius of 711.2 cm that corresponds to the magic momentum.

## 5.2 Betatron oscillations

The electric quadrupole field leads to harmonic oscillations known as betatron motion about the momentum-dependent central radius. To analyze this motion, a conventional accelerator coordinate system will be used. For each muon, $x = r - R_0$ is the radial distance and $y$ the vertical distance away from the fixed central orbit defined by the center of the collimators and quadrupole electrodes; the longitudinal path length is $s$.

The strength of the electric quadrupole field is characterized by $n = \frac{\partial E_r}{\partial r}$. The leading-order electric and magnetic fields are given by[1]

$$\vec{B} = B_0\hat{y} \tag{5.3}$$
$$\vec{E} = \kappa(x\hat{x} - y\hat{y}) \; . \tag{5.4}$$

The net force on the muon, including centrifugal and Lorentz forces, is then

$$F_x = \frac{\beta p}{r} - e\beta B_0 + e\kappa x \tag{5.5}$$
$$F_y = -e\kappa y \; . \tag{5.6}$$

For small oscillation amplitudes, the centrifugal force term may be approximated by a linear form:

$$\frac{\beta p}{r} = \frac{\beta(P_0 + \delta p)}{R_0 + x} \approx \frac{\beta P_0}{R_0}\left(1 - \frac{x}{R_0}\right) + \frac{\beta\,\delta p}{R_0} \; . \tag{5.7}$$

The equations of motion may be simplified by introducing a quantity $n$ called the field index, defined by

$$n = \frac{\kappa R_0}{\beta B_0} \; , \tag{5.8}$$

yielding

$$F_x = \gamma m\ddot{x} = -\omega_c^2(1-n)x + \frac{\beta\,\delta p}{R_0} \tag{5.9}$$
$$F_y = \gamma m\ddot{y} = -\omega_c^2 n y \; . \tag{5.10}$$

[1] This derivation loosely follows [92].

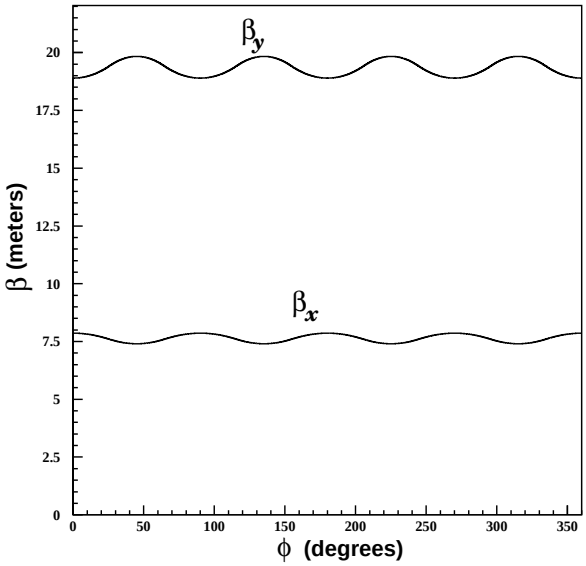



Figure 5.3: Radial and vertical β-functions for the storage ring. (Reproduced from [92].)

The solutions to these equations represent the betatron oscillations of single particles in the beam:

$$x = x_e + A_x \cos(\omega_x t + \phi_x) \tag{5.11}$$

$$y = A_y \cos(\omega_y t + \phi_y) \; . \tag{5.12}$$

Each particle has its own betatron amplitude and phase; they are determined by the position and angle at which it was injected. The oscillation frequencies

$$\omega_x = \omega_c \nu_x = \omega_c \sqrt{1-n} \tag{5.13}$$

$$\omega_y = \omega_c \nu_y = \sqrt{n} \tag{5.14}$$

are well-defined provided that $0 < n < 1$. In this range, the combined effect of the magnetic dipole and electric quadrupole fields is to focus the beam in both dimensions. The quantities $\nu_x$ and $\nu_y$ are called the radial and vertical tune. They represent the number of betatron wavelengths per turn. For the quadrupole voltages that are employed in the experiment, $n = 0.137$, so $\nu_x = 0.929$ and $\nu_y = 0.370$. The corresponding frequencies are $\omega_x/2\pi = 6.2$ MHz and $\omega_y/2\pi = 2.5$ MHz.

The previous derivation assumes that the $n$ value is constant as a function of $s$. The quadrupole electrodes only cover 43 percent of the azimuthal space, so this is clearly only an approximation. The primary effect of the real lattice structure is to multiply the solution by a β-function [77]; a phase advance term also appears:

$$x(s) = x_e + A_x \sqrt{\beta_x(s)} \cos\left[\nu_x \frac{s}{R_0} + \phi_x(s)\right] \tag{5.15}$$

$$y(s) = A_y \sqrt{\beta_y(s)} \cos\left[\nu_y \frac{s}{R_0} + \phi_y(s)\right] \; . \tag{5.16}$$

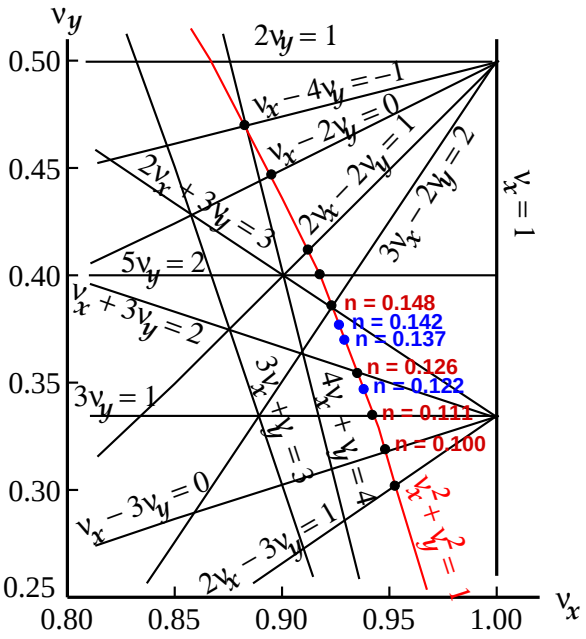



Figure 5.4: Resonance lines in the tune plane. Each of the straight black lines indicates a relation between $\nu_x$ and $\nu_y$ that places individual particles on closed orbits so that they continually experience the same field perturbations. The red curve expresses the constraint $\nu_x^2 + \nu_y^2 = 1$ that defines the values that can be reached by a weak-focusing system. The chosen operating point is $n = 0.137$. (Reproduced from [77].)

Because of the four-fold symmetry structure of the quadrupole electrode placement, the $\beta$-functions are approximately constant, as shown in Figure 5.3. Consequently, the smoothed approximation is a good one, and the solutions in Equations 5.11 and 5.12 constitute a useful description of the betatron motion of the beam.

## 5.3 Resonances, muon losses, and scraping

Losses of muons from the storage ring complicate the measurement of $\omega_a$ in two ways. If the fractional loss rate is not constant, then the exponential decay is distorted. Consequently, it is more difficult to obtain an acceptable fit to the recorded spectrum. If the mechanism that causes losses is correlated in any way with the muon spin direction, then an early-to-late phase change causes a systematic shift in $\omega_a$.

Muon losses are driven by imperfections in the magnetic and electric fields. Although the magnetic field is extremely uniform in the radial and vertical dimensions, it varies in azimuth by several hundred ppm. Losses are enhanced by resonances, which occur when betatron orbits are nearly closed. Near a resonance, a single particle passes through the same region of space many times, experiencing the same field perturbations each time. The condition that defines a resonance is

$$l\nu_x + m\nu_y = jN \tag{5.17}$$

where $l$, $m$, and $n$ are nearly integers and $N$ is the level of symmetry associated with the perturbation of interest. For example, the pattern of regions containing and excluding quadrupole electrodes repeats four times around the ring, so $N = 4$ for field perturbations inherent in the design of the quadrupoles. However, $N = 1$ for any perturbations that are located at a single position in azimuth. The order of a resonance is defined to be $|l| + |m|$. The lowest-order resonance lines in the $(\nu_x, \nu_y)$ plane are shown in Figure 5.4. By definition,

$$\nu_x^2 + \nu_y^2 = 1 \ , \tag{5.18}$$

so this curve is superimposed on the resonance lines. Clearly, the chosen value of $n = 0.137$ lies far any of these resonances.

In order to minimize losses at late times, the beam is scraped against the collimators by charging the quadrupole electrodes asymmetrically at very early times after injection. The particles nearest to the edge of the accepted phase space distribution are removed before the measuring period begins so that they are not lost later. While the top plates are charged to the full 24 kV, the bottom plates operate at only 17 kV during this period. In the radial dimension, the beam is shifted by placing -17 kV on the Q2 inner and Q4 outer plates (indicated in Figure 4.3) while all others are at the usual -24 kV. The scraping period lasts for 16 $\mu$s; the voltages then relax to a symmetric distribution with a time constant of 5 $\mu$s.

## 5.4 Coherent betatron oscillations

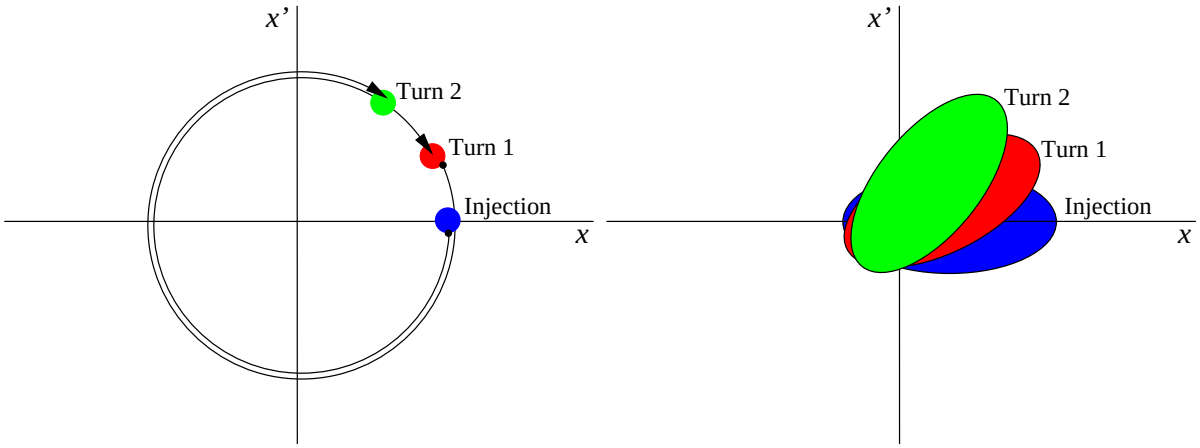
The betatron oscillations described in Section 5.2 refer to the motion of a single particle. There are also collective modes, known as coherent betatron oscillations, in which the centroid of the distribution of particles at a given azimuthal position oscillates radially and vertically at frequencies

$$\omega_{CBO} = \omega_c - \omega_x = (1 - \sqrt{1-n})\omega_c \tag{5.19}$$

$$\omega_{VCBO} = \omega_c - \omega_y = (1 - \sqrt{n})\omega_c \ . \tag{5.20}$$

Numerically, $\omega_{CBO}/2\pi = 470$ kHz and $\omega_{VCBO}/2\pi = 4.3$ MHz. Meanwhile, the anomalous precession frequency is $\omega_a/2\pi = 229.1$ kHz. The proximity of $\omega_{CBO}$ to $2\omega_a$ leads to a potentially dangerous observational resonance. Indeed, it will be shown in Section 8.1 that the largest systematic uncertainty in the measurement of $\omega_a$ results from this condition.

The origin of the coherent effects may be seen in Figure 5.4, which shows the motion of an ellipse representing a distribution of particles with a single momentum value in $(x, x')$ phase space. In a radial betatron period, each particle follows a circular trajectory centered at $(x_e, 0)$ and ends up where it began. The cyclotron period is slightly shorter than the radial betatron period, so the particle makes only a partial phase-space orbit in the time required to go around the ring; after a second turn, it lags behind even more, and so on. These snapshots of the phase space distribution at intervals of a cyclotron period correspond to the view from a single azimuthal position in the ring. This relation is clearest in the context of a tightly bunched beam where the distribution can only be observed once per turn as the bunch passes. Consequently, as sampled at a single azimuthal position, the distribution appears to be rotating backwards at the difference




(a) (b)

The radial betatron motion of (a) a single particle, and (b) a distribution of particles, illustrating why the CBO appears at the beat frequency $\omega_c - \omega_x$. Each turn around the ring, the phase space distribution makes a partial cycle, so it appears to rotate backwards at the distance frequency. The centroid and width of the projections of the distribution onto the $x$ and $x'$ axes clearly move with each turn as well.

frequency $\omega_c - \omega_x = \omega_{CBO}$. The argument continues to apply after the beam has debunched; each azimuthal slice can be considered separately.

The centroid of the radial profile associated with each snapshot of the distribution oscillates at $\omega_{CBO}$ because the phase space is filled somewhat asymmetrically. If viewed at a different azimuthal position, the same motion of the centroid would be observed but with a different phase; the CBO phase varies continuously through the entire range of $2\pi$ around the ring. A useful intuitive picture of this "swimming" mode is that of a "hula hoop" being shaken around the storage region. The amplitude of this oscillation is a few millimeters. There is also a small modulation of the width of the beam. The frequency associated with this "breathing" mode is twice that associated with the centroid motion because a reflection across the central orbit $x = x_e$ does not affect the width.

The acceptance of the calorimeters is a function of decay vertex radius as well as azimuth and energy. Consequently, the modulation of the average radius leads to a modulation of the counting rate at the level of a few percent. When particles come from a larger radius, they arrive at the inner edge of the storage region having curved around more in the field, so their angle with respect to the central orbit is greater. There are two competing mechanisms [93] that translate this angular shift into a change in acceptance:

- Geometry mechanism: particles that come from a larger radius are more likely to miss the detector. Their momentum is directed less in the azimuthal direction (towards the detector) and more inward along the radial direction. The acceptance is therefore greatest for *smaller* radial decay positions.

- Obstacle mechanism: particles that come from a larger radius arrive at obstructions such as the quadrupole electrodes and kicker plates with more of a perpendicular momentum component and

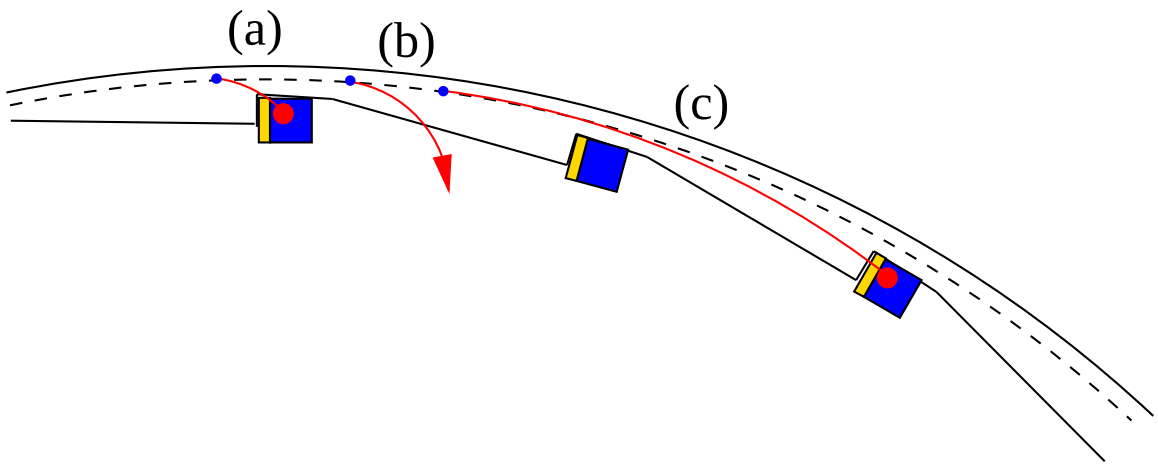


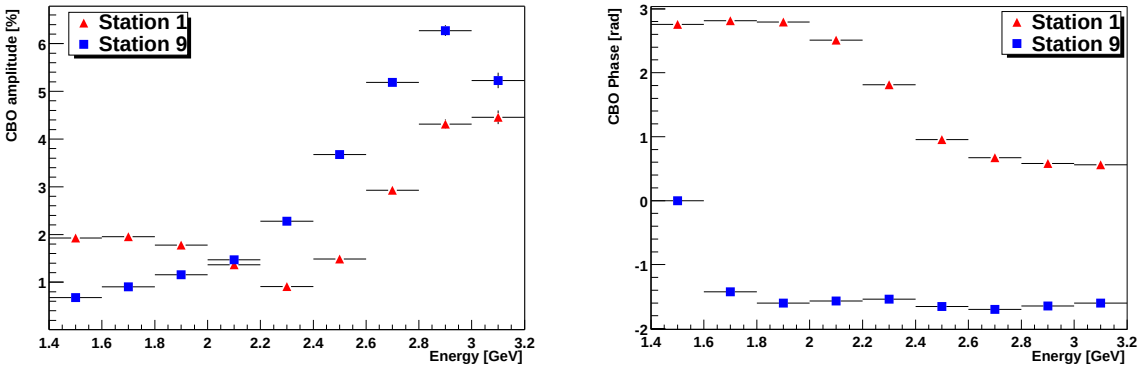

Figure 5.5: Trajectories followed by several decay positrons: (a) an accepted low-energy positron; (b) a low-energy positron that is not accepted; (c) a high-energy positron, nearly all of which are accepted.

pass through them more quickly. Consequently, they lose less energy, and the probability of an electromagnetic shower initiated outside the detector is smaller. This model predicts that the acceptance should be maximized for *larger* radial decay positions.

The schematic trajectories of decay positrons of various energies are shown in Figure 5.5. Lower-energy positrons follow a short path with a large curvature from the decay vertex, and they often exit the vacuum chamber between detectors and are missed. Higher-energy positrons, on the other hand, drift for several meters along a long arc before reaching a detector, so their geometric acceptance is essentially 100 percent. For low-energy positrons, the geometry mechanism is very important in determining the radial dependence of the acceptance. It is irrelevant for higher-energy positrons, because their geometric acceptance is complete, regardless of radius. Meanwhile, lower-energy positrons pass through obstacles like the kicker plates at an angle near the perpendicular in any event; small changes in the angle do not have a large effect on the amount of material encountered. Higher-energy positrons pass through at an oblique angle where such changes are important. Therefore, the obstacle mechanism describes the radial acceptance function for the higher-energy positrons and the geometry mechanism accounts for the lower-energy positrons.

Graphs of the observed CBO amplitude and phase versus detected positron energy in Figure 5.6 demonstrate the transition between these two regimes. They show the result of fits to the experimental positron arrival time spectrum for detector stations 1 and 9. In each case, the CBO phase shifts at a characteristic energy, indicating a transition from the dominance of the geometry mechanism to the obstacle mechanism. Station 1 is a typical detector; for it, the transition takes place at approximately 2.2 GeV. Station 9 is located just downstream of the kicker plates. Because the kicker plates are the thickest obstruction in the ring, they change the balance between the two mechanisms and move the phase transition down to lower energies.

The spin precession phase is also modulated directly at the CBO frequency. This phase modulation is quite important; the $(g-2)$ phase is strongly correlated with $\omega_a$ in a fit, so this effect couples $2\omega_a$ to $\omega_{CBO}$. The mean drift time from the decay point to the calorimeter increases with radius. Consequently, the CBO




(a) (b)

Figure 5.6: Energy dependence of (a) the amplitude and (b) the phase of the coherent betatron motion, illustrating the transition from one radial acceptance regime to the other. Station 1 is a typical detector, while station 9 is located just downstream of the kicker plates.

oscillation of the mean radius leads directly to a drift time oscillation. However, this effect is estimated to account for only approximately 20 percent of the observed phase modulation. A more significant issue is the CBO modulation of the $x'$ distribution. The average radial angle oscillates with an amplitude of up to a milliradian. This oscillation effectively represents a net torque on the ensemble of muons. The torque rotates the spin, and thus the phase, just as it rotates the momentum.

## 5.5 Pitch and electric field corrections

Equation 1.18 allows the determination of $a_\mu$ from $\omega_a$ in terms of the magnetic field. However, in the derivation of this equation, a term proportional to $\vec{\beta}\cdot\vec{B}$ has been neglected. The magnetic field is generally perpendicular to the momentum, so this approximation is justified. However, vertical betatron motion gives each particle a component of momentum parallel to the vertical magnetic field. A small correction to the measured frequency is required to account for this term. The correction was evaluated by Granger and Ford [94] and by Farley [95]. In the relevant limit, in which $\omega_y$ is very different from $\omega_a$, their result is

$$(\Delta\omega_a)_{pitch} = \frac{1}{2} < \psi_y^2 > \ , \tag{5.21}$$

where $\psi_y$ is the vertical pitch angle of the muon. The spectrum of pitch angles resulting from the vertical betatron oscillations is determined using a tracking simulation program that has been calibrated to measurements made with the fiber beam monitors and the fast rotation spectrum and checked against preliminary results from the traceback chambers. It gives $\sqrt{< \psi_y^2 >} = 0.79$ mrad, so this correction to $\omega_a$ is 0.31 ppm [96, 97].

Field and Fiorentini [98] also derive a formula for the pitch correction. In addition, they calculated the analogous correction that arises from radial betatron motion. This "yaw correction" is negligible because it does not depend on the dipole component of the magnetic field; at leading order, it is proportional to the quadrupole moment, which is very small.

By choosing a "magic momentum," the dependence of the $(g-2)$ frequency on the electric field is removed, permitting a term proportional to $\vec{\beta}\times\vec{E}$ to be dropped from Equation 1.18. However, the particles in the beam actually have a range of momenta around this value, so another correction must be applied to account for the effect of the electric field on these particles. This correction, which turns out to be 0.45 ppm, is computed with the same tracking simulation program [96, 97]. It performs the integration

$$(\Delta\omega_a)_E = \frac{\oint \left(\frac{E_r}{B_z}\right)\left(1-\frac{1}{a_\mu\beta^2\gamma^2}\right)\vec{\beta}\cdot d\vec{l}}{\oint \vec{\beta}\cdot d\vec{l}}\ . \tag{5.22}$$

In summary, the combined pitch and electric field correction is $0.76 \pm 0.03$ ppm.

# Chapter 6

# Preparation of $\omega_a$ data

The signals from the calorimeters are recorded as digitized waveforms like the one shown in Figure 4.9. This chapter describes the process used to translate these waveforms into the quantities of interest for physics analysis, which are the decay positron energy and the time relative to beam injection. It begins by explaining two features of the data, overlapping pulses and a time-varying baseline, that complicate the extraction of these two parameters. It describes the pulse fitting algorithm in detail, and it gives the results of tests of the algorithm with simulated overlapping pulses.

## 6.1 Overlapping pulses

When two pulses with energies $E_1$ and $E_2$ precisely overlap, they are indistinguishable from a single pulse with an energy $E_1 + E_2$. As they move apart, the pulse shape widens, though the integrated pulse area remains constant. Two distinguishable peaks emerge once the pulses are separated by more than their individual widths. Pulses at various separation times are illustrated in Figure 6.1.

The probability of observing two pulses that overlap within a fixed time resolution window $\Delta t_{res}$ to give an apparent energy $E_{tot} = E_1 + E_2$ is given by the convolution product [99]

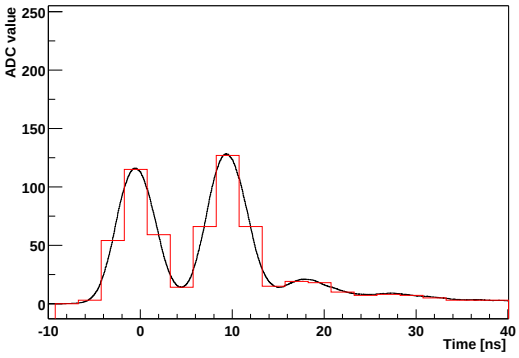
$$P_2(E_{tot},t) = \int_0^{E_{tot}} P_1(E_1,t)P_1(E_{tot}-E_1,t)\,dE_1 \ . \tag{6.1}$$

In this expression, $P_1(E,t)$ is the probability of finding a single pulse with energy $E$ within $\Delta t_{res}$ of $t$:

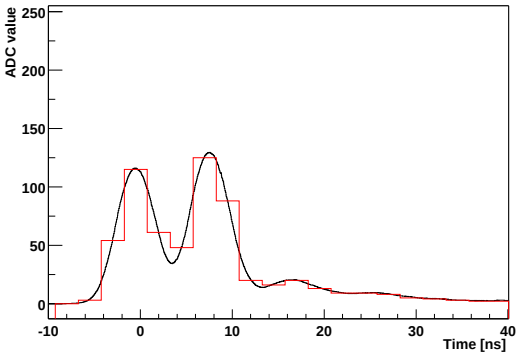
$$P_1(E,t) \propto e^{-t/\tau}[1 - A(E)\cos(\omega_a t + \phi_a(E))] \ . \tag{6.2}$$

Therefore, $P_2(E_{tot},t) \propto e^{-2t/\tau}$; that is, the lifetime of overlapping pulses ("pileup") is half that of the single-particle pulses from which they are formed.

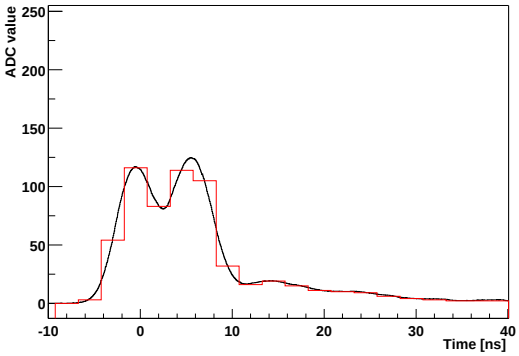
The observed spin precession phase depends on positron energy because higher-energy positrons follow a longer path from the decay vertex to the detector and therefore arrive later than lower-energy positrons. Pileup pulses carry the phase of the lower-energy positrons from which they are composed rather than



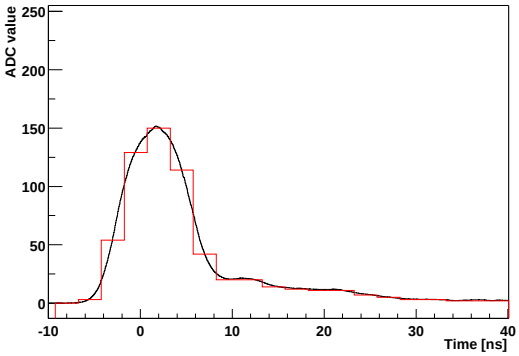

(a) $\Delta t = 10$ ns



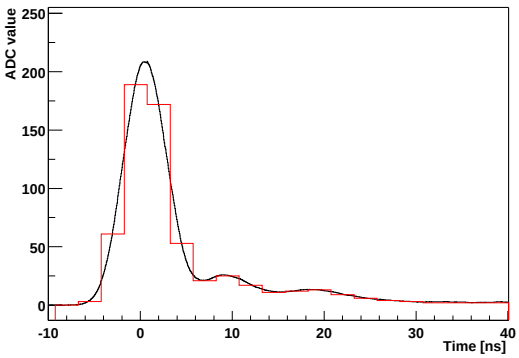

(b) $\Delta t = 8$ ns



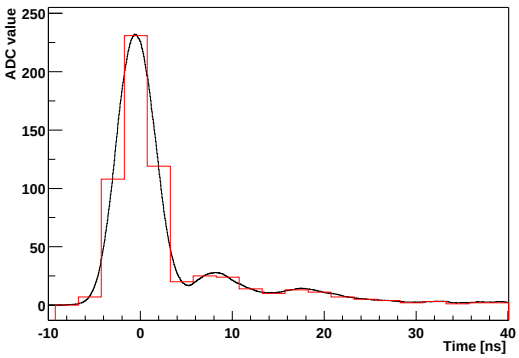

(c) $\Delta t = 6$ ns




(d) $\Delta t = 4$ ns




(e) $\Delta t = 2$ ns




(f) $\Delta t = 0$ ns

Figure 6.1: Two simulated overlapping pulses of equal energy are shown at various time separations. The analog pulse shape appears together with a simulated digitized waveform.

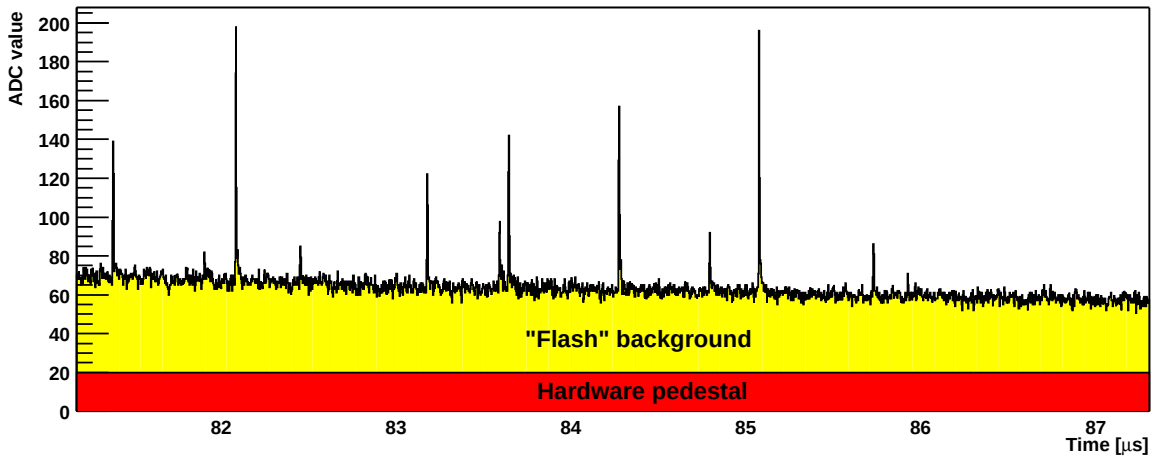



Figure 6.2: Illustration of the time dependence of the pedestal.

the phase of the high-energy positron from which they appear to arise. The concentration of overlapping pulses changes with the time relative to injection, so the average observed phase changes with time as well. This effect leads directly to a systematic error in the fitted frequency $\omega_a$. The goal of the pulse processing algorithm is therefore to separate overlapping pulses cleanly when it is possible to do so. The behavior of the algorithm in cases where separation is not possible needs to be understood so that a correction may be applied at a later stage of the analysis.

## 6.2 Pedestal shifts

The baseline ADC value on which the decay positron signals are superimposed, known as the "pedestal," varies as a function of time after injection. The waveform recorded by a single calorimeter at early times after injection in a typical fill is shown in Figure 6.2. The sharp spikes are decay positrons, while the pedestal shift arises from the delayed flash from neutron interactions in the vicinity of the detectors. The pedestal as a function of time after injection in a particular fill is described well by

$$p(t) = p_o + A_p \left( \frac{t}{1\ \mu\text{s}} \right)^{-n_p} . \tag{6.3}$$

The constant $p_o$ is set in the WFD hardware; it is deliberately nonzero so that pulses that cross through ground can be recorded. The exponent $n_p$ is approximately 1.2. The overall scale $A_p$ varies from fill to fill with the incident pion intensity; also, it is greatest in stations near the inflector. The pulse processing algorithm must remove the effect of the time-varying pedestal, not allowing it to artificially increase the energy that is assigned to pulses at early times.

## 6.3 Pulse fitting procedure

The design of the pulse processing algorithm is predicated on the assumptions that

- all positron signals from a particular calorimeter have essentially the same shape, and
- this shape scales linearly with the energy of the pulse.

It iteratively adjusts the assumed time, energy, and pedestal for each pulse in a digitization interval ("island") to match the average pulse shape to the measured ADC samples. When several pulses are present in the same island, their times and energies are varied simultaneously to achieve the optimal least-squares fit.

The procedure begins by locating distinct peaks within the interval to determine how many pulses to include in the initial fit. It then minimizes

$$D = \sum_{i \in \text{samples}} [S_i - P - \sum_{j \in \text{pulses}} A_j f_i(t_j)]^2, \tag{6.4}$$

where $S_i$ are the measured samples and $f_i(t)$ is the average pulse shape. $t_j$, $A_j$ and $P$ are the fit parameters, representing the times and amplitudes of each pulse and the overall pedestal. The quality of the fit is evaluated; if there are locations where the discrepancy between the model and the measured samples is too great, it tries again with a new model that includes additional pulses. The detailed structure of this process is shown in the pseudocode outline in Figure 6.3.

Within the collaboration, there are two programs, known as `G2OFF` and `G2Too`, that implement this basic procedure. The analysis described in this thesis is based on `G2Too` [100], which is written in the C++ programming language in the context of the ROOT framework [101]. `G2OFF` [102, 103] is written in Fortran and relies on older CERN libraries. The two pulse fitting subprograms are conceptually similar; however, they share absolutely no code, and they differ in a substantial number of relevant details. The existence of both programs has proven to be mutually beneficial; by cross-checking their results, each has uncovered defects in the other, permitting them to evolve together.

## 6.4 Average pulse shape

The time of a pulse is defined by its peak position, which varies from 0 to 5 ns with respect to the rising edge of the 200 MHz clock that drives the WFD. The pulse shape has a different appearance depending on where it falls with respect to this clock. Figure 6.4 shows "snapshots" of a pulse at different time offsets. Snapshots like these are stored for the average pulse shape at 100 time offsets. Each snapshot holds 24 samples, starting 5 samples before the peak of the pulse and going through 18 samples after. The pulse shape snapshots from different time offsets are interleaved to show the overall pulse shape, illustrated for station 1 in Figure 6.5. The two curves in this figure correspond to the two independent phases of the WFD, which clearly see a slightly different pulse shape. The pulse fitting procedure deals with these shapes consistently, always associating each sample with the average pulse shape from the same phase.

- Split up long digitization intervals.
- For each sub-interval:
  - Make initial guesses for number of pulses and the times of each pulse.
  - Repeat:
    - Remove pulses whose parameters have drifted to unreasonable values.
    - Perform minimization to fit times, energies, and pedestal.
    - If there is a location where an additional pulse is needed, add it.

    ...until no new pulse was added (maximum of 50 iterations)
  - Repeat:
    - Remove pulses according to strict criteria.
    - If pulses were removed, then perform minimization to fit times, energies, and pedestal.

    ...until no pulses were removed (maximum of 50 iterations)

Figure 6.3: Outline of the overall structure of the pulse processing algorithm in `G2Too`.

The pulses that are used to build the pulse library must be aligned in time and normalized to have unit area. The area normalization is almost irrelevant, since the pulse shape is not a strong function of energy. Consequently, a simple algorithm is used: the area is computed simply by adding together the 16 pedestal-subtracted samples starting 4 samples before the peak of the pulse. The time alignment, on the other hand, is particularly critical; if an incorrect procedure is used to determine the time, then the resulting average pulse shape will be distorted.

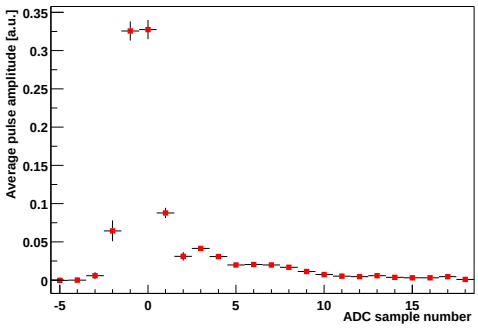
The first step in estimating the time of each pulse is to compute its "pseudotime"

$$\tau = (2.5\text{ ns})\left(i_{max} + \frac{2}{\pi}\tan^{-1}\frac{S_{i_{max}} - S_{i_{max}-1}}{S_{i_{max}} - S_{i_{max}+1}}\right), \tag{6.5}$$

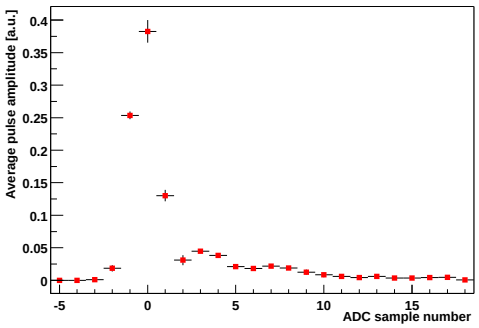
where $i_{max}$ is the index of the maximum sample and $S_i$ is the value of the $i$th sample. The argument of the inverse tangent measures the relationship between the three central samples; as the pulse moves later in time, the later samples increase relative to the earlier ones. It should be noted that both the pedestal and the scale of the pulse cancel out in this calculation. The true time of the pulse is in general a nonlinear function $t(\tau)$ of the pseudotime. However, for any reasonable pulse shape, $t(\tau)$ is a monotonically increasing function. The distribution of pseudotimes $N(\tau)$ can be measured, and the distribution $N(t)$ of true times is known to be uniform. Modulo 5 ns, the function $t(\tau)$ is then determined to be

$$t(\tau) = (5\text{ ns})\frac{\int_0^\tau N(\tau')d\tau'}{\int_0^{5\text{ ns}} N(\tau')d\tau'}. \tag{6.6}$$

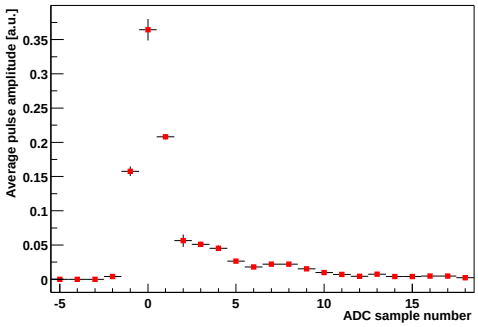
This function is shown for a typical detector in Figure 6.6. Building the average pulse shape library requires



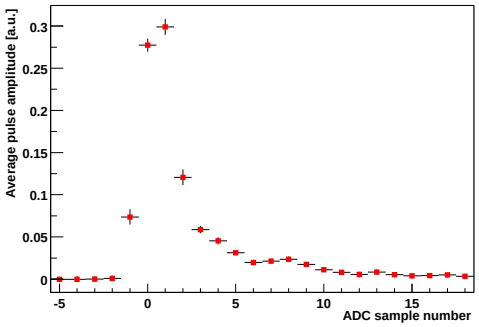

(a) $\Delta t$ = 0 ns



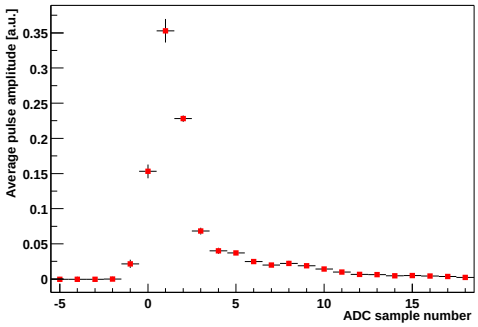

(b) $\Delta t$ = 1 ns




(c) $\Delta t$ = 2 ns




(d) $\Delta t$ = 3 ns




(e) $\Delta t$ = 4 ns

Figure 6.4: Five snapshots from the pulse shape library for station 1. Each snapshot shows how a pulse at a given time offset $\Delta t$ from the 200 MHz clock boundary is sampled (on average) by the WFD. The pulse shape library for each detector contains 100 of these snapshots.

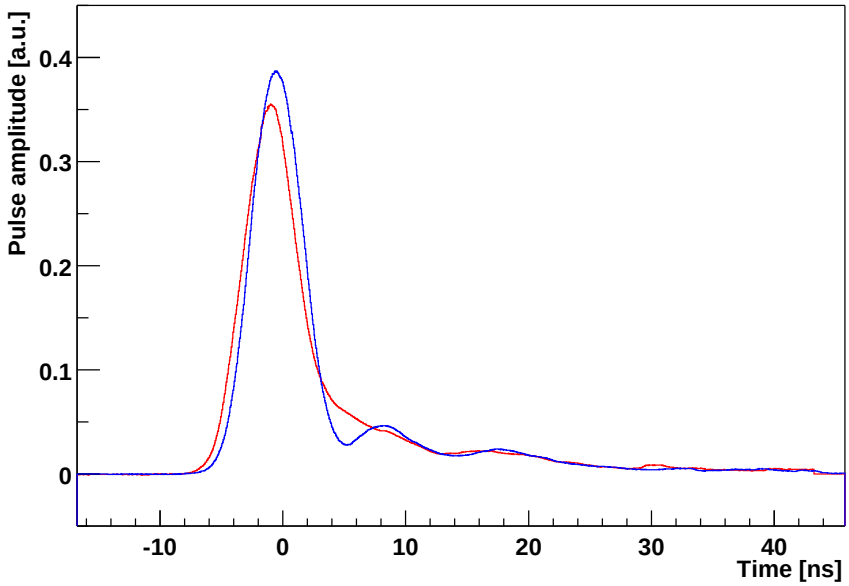



Figure 6.5: Average pulse shapes for detector station 1. Phase 0 is shown in blue and phase 1 is shown in red; alternatively, phase 0 is the shape with more pronounced oscillatory behavior.

two passes through the data. During the first pass, the distribution $N(\tau)$ of pseudotimes is constructed. The second pass actually builds the pulse library.

## 6.5 Breaking up long digitization intervals

The WFD employs a simple analog threshold to determine when to record data; it does not attempt to compensate for pedestal shifts. During the periods of highest intensity, the signals from some stations exceed this threshold for more than 100 $\mu$s, leading to very long waveforms. It is impractical to fit all of the pulses residing on these long continuous digitizations simultaneously: the number of fit parameters becomes too large, and the pedestal is no longer approximated well by a constant over this time range. Consequently, the interval must be split up.

It is only safe to break up an interval at points where the gap between adjacent pulses is large enough that they can be treated independently. Otherwise, energy that is associated with one pulse will also be counted with the next, causing a rate-dependent change in the effective gain of the pulse fitting process. The first step in locating these safe points is to estimate the pedestal as a function of time within the interval; once the pedestal is known, quiet periods can be located. The interval is initially broken into segments of up to 64 ADC samples, and the pedestal associated with each sub-interval is computed following the algorithm given in Figure 6.7. This algorithm takes the pedestal to be the average of the lowest few samples in the region, where “few” is defined by matching the RMS width of the chosen set of samples to the known noise level in the WFD. Less than 1 percent of the area of a pulse follows the peak by more than 60 ns.

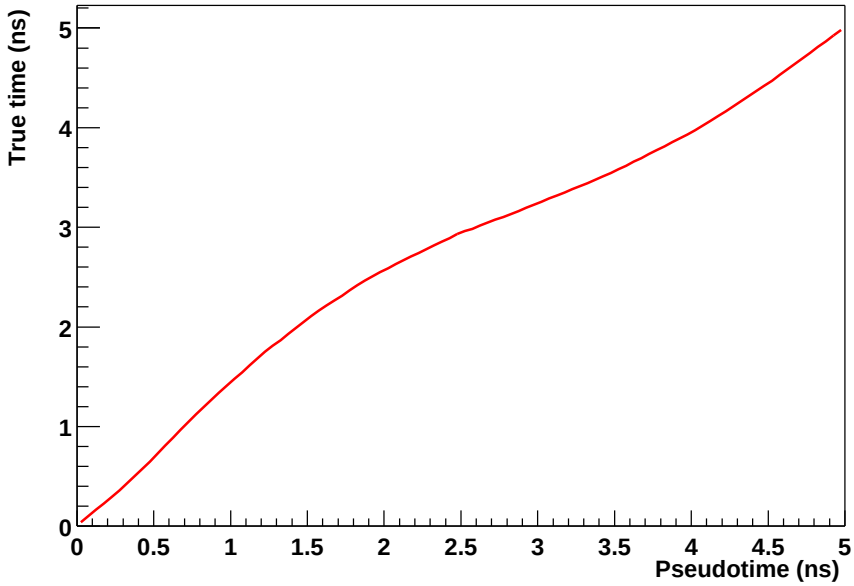



Figure 6.6: Relationship between pseudotime $\tau$ and true time $t$.

Consequently, the final break points are placed where the preceding 60 ns are consistent with the estimate of the pedestal, again within the known noise level of the WFD.

## 6.6 Minimization procedure

In order to remove any possible dependence on the length of the digitization interval, only a subset of the ADC samples is included in the fit to the model of Equation 6.4. A total of 15 samples are used for each pulse that is included in the fit, ranging from seven samples before the assumed peak time through seven samples after it. Samples that equal 0 or 255, which are the underflow and overflow values of the WFD,

- Compute a histogram of the sample values. That is, count the number of occurrences of each integer between 0 and 255 in the samples in the interval.
- Let $T$ step from 0 to 255. At each point:
  - Compute the mean and the RMS of the subset of the histogram less than $T$.
  - If the RMS is less than 0.25, set $P$ to be the mean, and continue to the next point.
  - If the RMS is greater than 0.25, define the pedestal to be the last value set as $P$, and stop.

Figure 6.7: Pseudocode describing the algorithm that makes an initial guess for the pedestal of a subset of a digitization interval. This pedestal value is used only to locate points where the interval may be broken safely into smaller constituents that will be treated independently.

are eliminated; there is no way to know what voltage level they actually represent. When several pulses are included in the fit, any of them may trigger the inclusion of a sample. The set of included samples is chosen before the minimization process begins based on the initial estimate of the pulse times, and it is kept constant during the fit.

The most obvious approach to the fitting problem treats $t_j$, $A_j$ and $P$ as independent fit parameters and allows a nonlinear minimization engine to vary them all. A refinement of the process recognizes that only the times $t_j$ need to be included directly as fit parameters; the optimal $A_j$ and $P$ may be calculated analytically from them. The partial derivatives of $D$ in Equation 6.4 are

$$\frac{\partial D}{\partial P} = -2 \sum_{i \in \text{samples}} [S_i - P - \sum_{j \in \text{pulses}} A_j f_i(t_j)] \text{ and} \tag{6.7}$$

$$\frac{\partial D}{\partial A_j} = -2 \sum_{i \in \text{samples}} f_i(t_j)[S_i - P - \sum_{j \in \text{pulses}} A_j f_i(t_j)] \, . \tag{6.8}$$

The minimum occurs at the point where $\frac{\partial D}{\partial P} = \frac{\partial D}{\partial A_j} = 0$. This condition defines a system of linear equations which may be written in matrix-vector form as $AX = B$ where

$$X = \begin{pmatrix} P \\ A_0 \\ A_1 \\ A_2 \\ \vdots \end{pmatrix} , \tag{6.9}$$

$$A = \sum_{i \in \text{samples}} \begin{pmatrix} 1 & f_i(t_0) & f_i(t_1) & f_i(t_2) & \cdots \\ f_i(t_0) & f_i(t_0) f_i(t_0) & f_i(t_0) f_i(t_1) & f_i(t_0) f_i(t_2) & \cdots \\ f_i(t_1) & f_i(t_1) f_i(t_0) & f_i(t_1) f_i(t_1) & f_i(t_1) f_i(t_2) & \cdots \\ f_i(t_2) & f_i(t_2) f_i(t_0) & f_i(t_2) f_i(t_1) & f_i(t_2) f_i(t_2) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} , \text{ and} \tag{6.10}$$

$$B = \sum_{i \in \text{samples}} \begin{pmatrix} S_i \\ S_i f_i(t_0) \\ S_i f_i(t_1) \\ S_i f_i(t_2) \\ \vdots \end{pmatrix} . \tag{6.11}$$

For the typical case where there is only one pulse present in the interval being fit, there is only a single fit parameter: the time of that single pulse. For efficiency, a very simple one-dimensional minimizer is used to optimize for the time in this special case. It assumes that its initial guess is near a parabolic minimum. At each iteration, it computes $D$ at three points ($t_0 - 0.2$ ns, $t_0$, and $t_0 + 0.2$ ns), calculates the parameters of the parabola that passes through these points, and determines the position of the minimum. Up to three

iterations are allowed. It reports convergence when the time changes from one iteration to the next by less than 0.02 ns, and it reports failure if it ever changes by more than 1 ns. If more than one pulse is included in the fit, or if the simple minimizer fails, then MINUIT [104] is used instead. It has been verified that, for the case of a single pulse, the simple minimizer and MINUIT yield the same result.

## 6.7 Including and excluding pulses

Initially, the pulse finder searches for local maxima that exceed a software threshold of 40 ADC counts above the estimated pedestal. Their times are determined using the same pseudotime method that is used in the computation of the average pulse shapes. Through the iterative process of adding and removing pulses from the fitting function, significantly smaller pulses will eventually be found and included.

The decision to add a pulse to the fit is based on local fit quality following a minimization step. Each group of three consecutive samples is examined. The criteria for adding a pulse are (all must be satisfied):

- The residual area must be at least 30 units (approximately 200 MeV).
- The average squared residual of the three samples must exceed 9.
- There must be no other pulse within 2.63 ns.

The initial guess for the time of a pulse that is added is based on the pseudotime computed from the residuals.

There are two sets of criteria for removing pulses that are applied at different stages in the fitting process. At intermediate stages, only pulses that have physically unreasonable parameters are removed. If any of the following properties is met, the pulse is removed:

- The area is less than 20 units (approximately 140 MeV).
- The time is more than 5 ns before the beginning of the interval being fit or more than 5 ns after the end.
- The pulse is within 1.75 ns of another pulse.

In the final stage, stricter criteria are applied:

- The area must be greater than 40 units (approximately 275 MeV).
- The time must be within the interval being fit.
- There may not be another pulse within 3.5 ns.

The area and time separation thresholds may appear to be “magic” numbers. In fact, they were obtained with a simple Monte Carlo simulation, which will be described in Section 6.9. The chosen thresholds arose from an iterative application of this simulation. They were first set to very low values and then progressively raised until no spurious pulses were reconstructed.

## 6.8 Cancellation of very low-energy pulses

There are a large number of low-energy pulses below the software threshold of 275 MeV. Depending on where these small pulses occur, whether close to a large pulse or far away from it, the effect they have on the large pulse varies. When a small pulse is close to a large pulse, it will be absorbed in the large one, thereby raising the energy of the large pulse and possibly shifting its time. If a small pulse is far away from the main pulse, it will be absorbed in the pedestal, thereby raising the pedestal by some amount, which in turn lowers the energy of the main pulse. The number of samples included in the fit determines the magnitude of the effect on the pedestal.

In an ideal case, the energy contributed by small pulses exactly cancels out when averaged over many fills. The proof of this result follows from two assumptions:

- Small pulses are equally likely to appear at any time separation from the large pulse. This is certainly true for suitably small time separations.
- The pulse time does not vary in the fit. This is only an approximation because the pulse time is allowed to float; however, the effect of a small pulse on the pulse time is quite small.

The expression in Equation 6.4 may be simplified for the typical case of a single pulse over the software threshold:

$$D = \sum_{i \in \text{samples}} [S_i - P - A f_i(t)]^2. \tag{6.12}$$

Then the solution for the minimum is

$$A = \frac{(\sum_i S_i)(\sum_i f_i(t)) - N_{samples} \sum_i S_i f_i(t)}{(\sum_i f_i(t))^2 - N_{samples} \sum_i (f_i(t)^2)} \; . \tag{6.13}$$

If a sample $S_k$ is perturbed by an amount $\delta S_k$ then the fitted area is perturbed by

$$\delta A_k = \delta S_k \frac{\sum_i f_i(t) - N_{samples} f_k(t)}{(\sum_i f_i(t))^2 - N_{samples} \sum_i (f_i(t)^2)} \; . \tag{6.14}$$

If the perturbation $\delta S_k$ is the same for all samples, then $\sum_k \delta A_k = 0$.

In spite of this cancellation, the presence of small pulses causes a degraded energy resolution. Consequently, the energy resolution changes from early to late times. A systematic uncertainty will be assigned for this effect, which is known as “unseen pileup.”

## 6.9 Tests with simulated waveforms

The pulse fitting algorithm has been tested with a Monte Carlo simulation that samples the average pulse shape to generate raw WFD events containing a set of pulses of known times and areas. It is used to

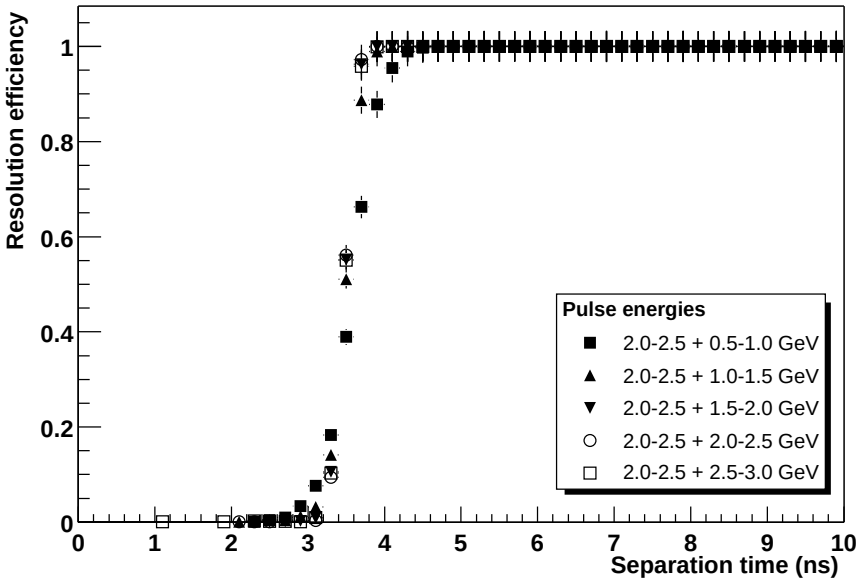
superimpose two pulses from the pulse shape library, fit the resulting waveform, and compare the result with the known input parameters. The essential idea is illustrated by the double pulses shown in Figure 6.1, which were generated using this simulation.

Figure 6.8(a) shows the probability that the pulse fitter reconstructs two pulses as a function of the separation in time between the two pulses, broken down into energy bins. All energies follow a similar pattern; for separation times less than 3 ns, essentially no double pulses are reconstructed. For separation times greater than 4 ns, two pulses are essentially always found. In the region between 3 and 4 ns, there is a gradual transition between these two states. Because two pulses are not allowed to have a reconstructed time separation of less than 3.5 ns, many double pulses found in this transition region have misreconstructed times. The individual energies of the two pulses are similarly mistreated. It is found that the individual times and energies can only be trusted for pulses with a separation of at least about 5 ns.

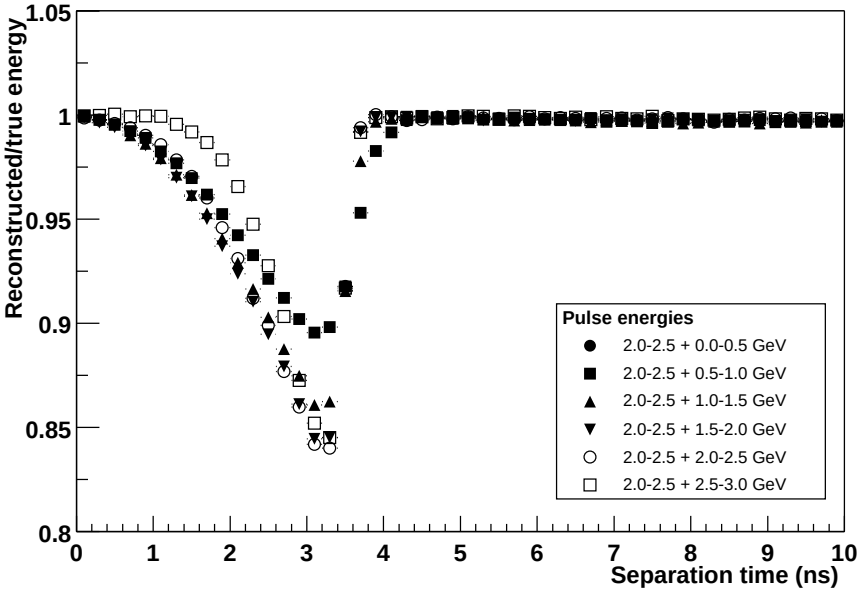
The energy-weighted average time and the sum of the energies are treated predictably by the algorithm, even for pulses in the transition region. Figure 6.8(b) shows the ratio of the sum of the energies in all pulses reconstructed by the fitter to the sum of the true energies of the two pulses created by the simulation. It deviates from unity by as much as 15 percent in the transition region; averaged over all unresolvable pileup events, the ratio is 0.94. This shift necessitates a correction at a later phase of the analysis. Meanwhile, the difference between the energy-weighted average time of all pulses reconstructed by the fitter and the energy-weighted true time of the two pulses deviates from 0 by a maximum of 150 ps in the transition region, which is not enough to require a correction.

The "soft" deadtime of 3 ns to 5 ns implied by Figure 6.8(a) makes it difficult to quantitatively model the expected pileup distribution. To simplify the situation, an artificial deadtime of at least 5 ns is applied. When two pulses are found at a separation of less than this deadtime, they are combined into a single pulse. The principle of this deadtime application is that the constructed pileup should look as much as possible like the inseparable overlapping pulses. The energy of the combined pulse is therefore the sum of the input energies, multiplied by an energy-dependent nonlinearity factor, while its time is the energy-weighted average of the times of its constituents. The nonlinearity factors–which, as previously stated, average to 0.94–are computed for all combinations of overlapping pulse energies $E_1$ and $E_2$. Figure 6.9 shows the average correction factor in two dimensions versus these energies; the largest fractional deviations from linearity are seen for the smallest pulses. The factor also depends strongly on the separation $\Delta t$ between the two pulses that are combined. Figure 6.10 illustrates this point by showing the distribution of correction factors obtained for a pair of 2 GeV particles over the entire range of time separation values. Histograms like this one are constructed for every pair of energy ranges $E_1$ and $E_2$; each time pulses are combined, a value is drawn at random from the appropriate distribution.

The level of cancellation of unseen pileup was also estimated using the simulation. Figure 6.11 shows the ratio of the reconstructed to the true energy as a function of the location of a small pulse with respect to the main pulse. Of course, the relevant quantity is the average over the time bins in each of these plots: the extent by which it differs from unity measures the amount of non-cancellation of unseen pileup. These average deviations are approximately 0.3 percent.




(a) Reconstruction efficiency




(b) Reconstructed / true energy

Figure 6.8: Reconstruction efficiency and ratio of reconstructed to true energy for two simulated overlapping pulses.

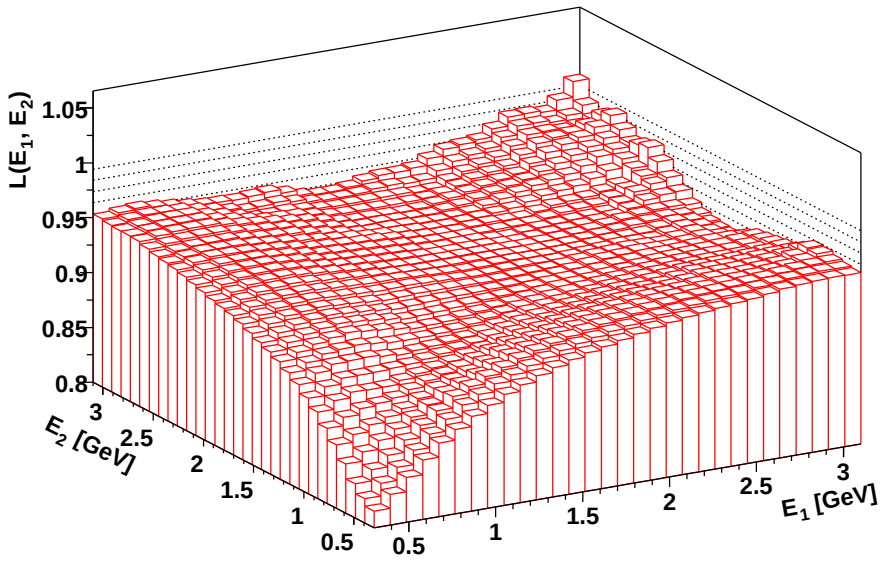



Figure 6.9: Average pulse fitting nonlinearity factor $E_{reconstructed}/E_{true}$ versus pulse energies $E_1$ and $E_2$.

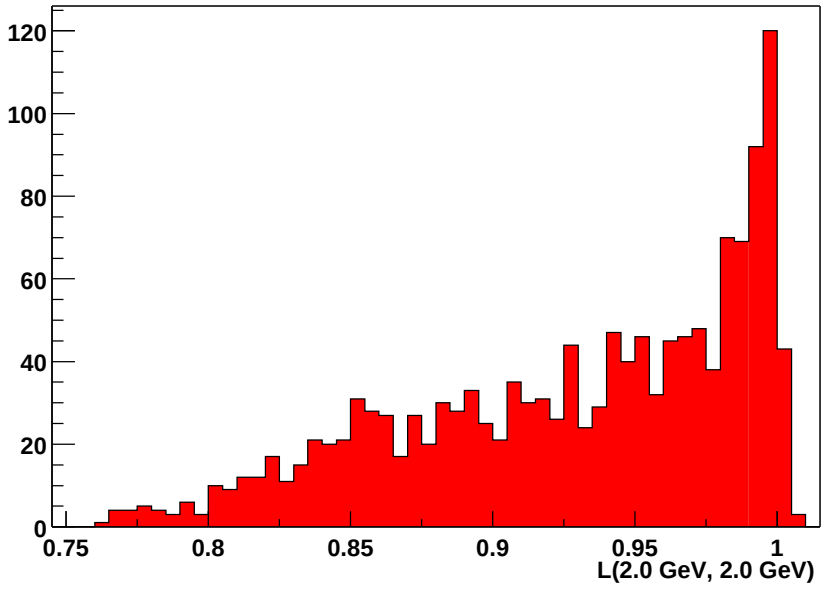



Figure 6.10: Distribution of nonlinearity factors $E_{reconstructed}/E_{true}$ for 2.0 GeV + 2.0 GeV pulses

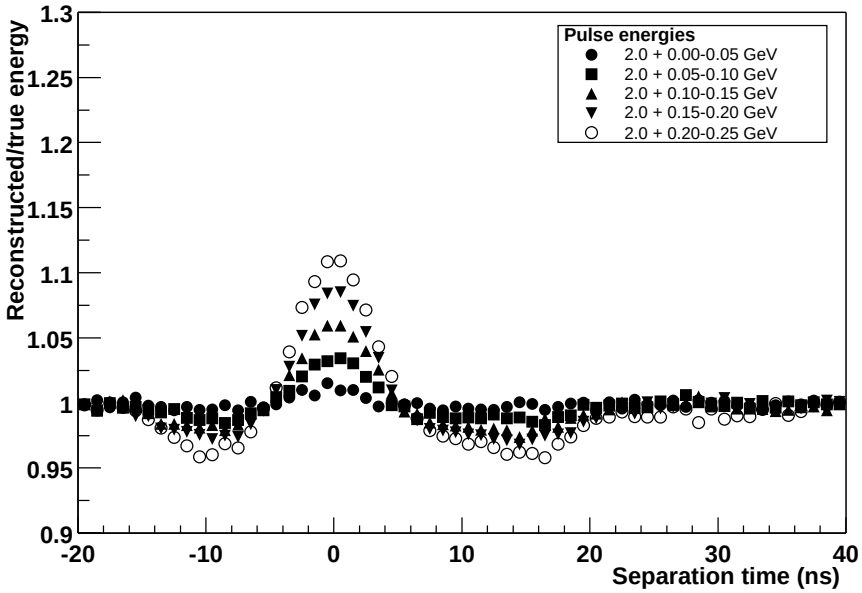



Figure 6.11: Ratio of the reconstructed to the true energy for a 2 GeV primary pulse distorted by the addition of a very small secondary pulse, as a function of the time separation of the two pulses. Integrated over all time separations, the energy contributed by the secondary pulses is canceled at the level of $10^{-3}$.

## 6.10 Pileup construction

The applied artificial deadtime determines the concentration of pileup in the data sample. For the final result quoted in this thesis, an artificial deadtime of 7.5 ns was used; it was checked against a result that used a 5 ns deadtime, and good agreement was observed. Knowing this parameter, the distribution of overlapping pulses may be constructed by looking for out-of-time coincidences. Each pulse is followed by a 7.5 ns deadtime window; the probability of a second pulse arriving in that window is equal to the probability that it lands in any other sufficiently nearby 7.5 ns window. The distribution is measured by counting pulses in the nearby window as a function of energy and time.

The WFD has a configurable hardware threshold, $E_{HW}$, which was set for each station to a value between 0.9 and 1.4 GeV. Because of the time variation of the pedestal, the effective threshold is not time independent; these are the values to which the threshold rises at very late times after injection. A "trigger pulse" is defined to be a pulse whose energy exceeds this late-time threshold. The WFD always records at least 15 ns before and 25 ns after each trigger pulse, creating an effectively threshold-free region there. Two 7.5 ns collection windows are placed in this region as shown in Figure 6.12; pulses that fall in these windows are known as "shadow pulses." They may have any energy above the pulse fitter's reconstruction threshold $E_{SW}$.

A simplified construction procedure uses only shadow pulses from the window after the trigger pulse, ignoring the window that precedes it. Again, the probability of finding a second pulse in this window is equal

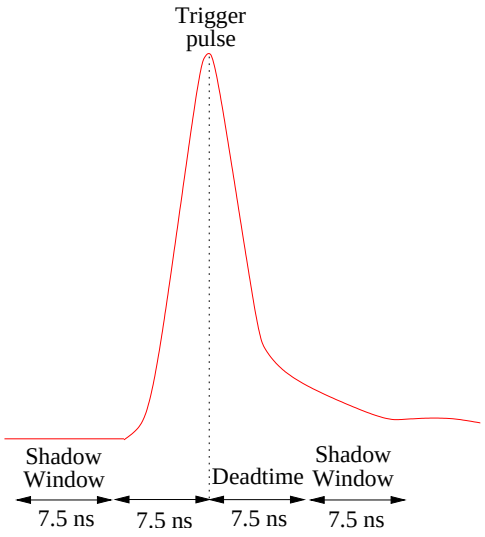



Figure 6.12: Placement of the shadow windows relative to a trigger pulse.

to the probability of finding the two pulses overlapped within 7.5 ns. If they had overlapped, the two single pulses would be lost and replaced by a combined pulse. The energy of the merged pulse would be the sum of the individual energies, multiplied by a nonlinearity adjustment factor taken from the same distribution that is used to apply the artificial deadtime. The spectrum of trigger pulses will be called $S_T(E,t)$, the spectrum of shadow pulses $S_S(E,t)$, and the spectrum of combined pulses $D_{TS}(E,t)$. In these terms, the pileup is given by

$$P_{TS}(E,t) = D_{TS}(E,t) - S_T(E,t) - S_S(E,t)\ . \tag{6.15}$$

The plane of shadow energy versus trigger energy is shown in Figure 6.13. The area in this plane that is directly constructed by definition has $E_T > E_{HW}$. In order to cover a larger fraction of the plane, the part where $E_S < E_{HW}$ is duplicated with $E_T$ and $E_S$ interchanged as indicated by the arrow. Practically, this duplication means that events with $E_S < E_{HW}$ are each counted twice. The area in the plane that is still left uncovered after this reflection corresponds to two pulses whose energies are both below the hardware threshold $E_{HW}$ and are therefore never seen. Their energies can add up to values as large as about $2E_{HW}$; consequently, pileup construction is only complete for energies above twice the hardware threshold.

The procedure that is actually employed in this analysis uses both of the windows before and after the trigger pulse. This symmetrization avoids even the appearance of a phase shift in the overlap construction procedure. At first glance, it appears to systematically overcount by a factor of 2 because the collection windows cover a total of 15 ns rather than 7.5 ns. However, each pulse is only allowed to be used once; if a pulse has been a shadow, it may not subsequently be a trigger. Consequently, pairs of pulses where both energies exceed $E_{HW}$ are counted properly. Because of the width covered by the windows, twice as many pairs including one trigger candidate and one non-trigger candidate are counted as in the single window method. Therefore, there is no need to artificially multiply these events by 2 as before, and all pairs of energies are counted properly. However, even the symmetrized procedure is only completely effective for

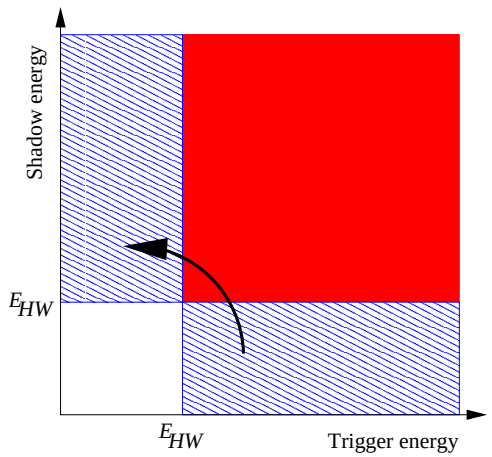



Figure 6.13: Plane of $E_S$ versus $E_T$, showing which regions are included in the constructed pileup spectrum.

energies of at least $2E_{HW}$.

The pileup construction procedure may only be extended to lower energies by eliminating the dependence on the trigger pulses. By combining shadow pulses with other shadow pulses rather than with trigger pulses, the barrier of the hardware threshold may be overcome. An additional spectrum $D_{SS}(E,t)$ is built, containing artificially combined shadow pulses. This spectrum, like the others, is a two-dimensional histogram with 150 ns wide time bins and 200 MeV wide energy bins. The shadow pulses that are combined are always drawn from the same time bin, but from unrelated WFD triggers that are typically in different fills. Indeed, at late times where there are few overlapping pulses, shadow pulses that arrive separated by many fills (even by days or weeks) may be combined. To avoid double-counting, pre-shadow pulses are never combined with post-shadow pulses; they are treated as two separate streams of events until they are matched up. The spectrum

$$P_{SS}(E,t) = D_{TS}(E,t) - S_S(E,t) \tag{6.16}$$

is valid for energies as low as twice the pulse fitter threshold. However, it is not correctly normalized; the time distribution of the trigger pulses is not folded out properly, so the efficiency with which $P_{SS}(E,t)$ is determined oscillates at $\omega_a$.

The solution to this problem is to combine the two methods, matching them in each time bin over the energy range where the original trigger-shadow method is valid. The original method sets the normalization, and the new technique determines the distribution for low energies:

$$P(E,t) = f_{TS/SS}(t) P_{SS}(E,t) \tag{6.17}$$

$$f_{TS/SS}(t) = \frac{\int_{E_L}^{E_U} D_{TS}(E,t)}{\int_{E_L}^{E_U} D_{SS}(E,t)} . \tag{6.18}$$

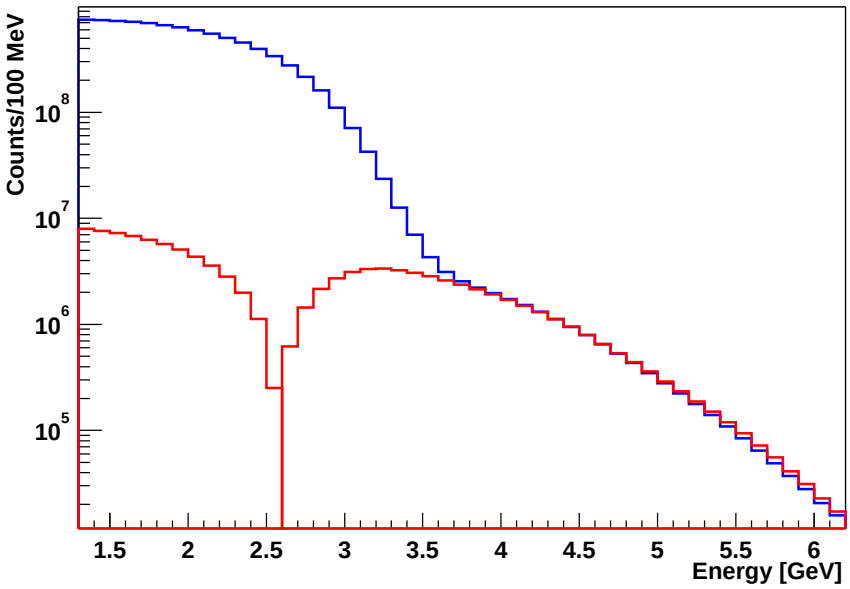



Figure 6.14: Energy distributions of normal (blue) and pileup (red) events. The pileup distribution crosses through zero at 2.6 GeV, becoming positive at high energies; its absolute value is shown.

The limits of integration are $E_L$=2.8 GeV and $E_U$=4.0 GeV. The lower limit is chosen to be just above twice the highest hardware threshold; the upper limit is somewhat arbitrary. In practice, the scale factor is fit to a function of the form

$$f_{TS/SS}(t) = N(1 + A\cos(\omega_a t + \phi_a)) \,, \tag{6.19}$$

and the function is evaluated at each point rather than using the actual data there.

A small correction is applied to the overlapping pulse spectrum to account for the fast rotation modulation. At very early times, the rate in the trigger pulse window differs substantially from the rate in the shadow pulse windows. The correction is given by

$$\frac{< R^2_{trigger}(t) >}{< R_{trigger}(t) \cdot R_{shadow}(t) >} \,, \tag{6.20}$$

where $R_{trigger}(t)$ is the average rate in a $\pm$10 ns window around time $t$, $R_{shadow}(t)$ is the average rate in the regions where shadow pulses are collected, and the brackets imply an average over a fast rotation cycle. The correction was calculated from the data itself in fine time bins. At times of interest, it is quite small; it is a 0.5 percent effect at 30 $\mu$s after injection and completely negligible by 40 $\mu$s.

The energy distribution of the constructed pileup is shown in Figure 6.14, together with the corresponding ordinary single-event spectrum. The portion of the spectrum at energies much higher than 3.1 GeV can only be the result of pileup, and it matches up well with the construction. At the very highest energies, the two curves begin to diverge again; this is where three-particle pileup begins to play a significant role.

## 6.11 Energy and time scale calibrations

The relationship between energy and pulse amplitude was not entirely constant over the course of the data collection period. Variations of order 5 percent result from deliberate recalibrations, drift in the high voltage power supply, and gradual deterioration of the photomultiplier tubes. Consequently, the data set was divided into approximately 50 time periods for each detector, and the endpoint of the energy spectrum was fit for each of these time periods. These energy calibrations were performed with the `G2OFF` data, but they were used to determine the relative calibration versus time for `G2Too` as well. The absolute energy scale was then set by forcing each detector to have a fitted $(g-2)$ asymmetry $A$ of 0.4 over the energy range from 2.0 to 3.2 GeV.

Finally, corrections were applied to account for changes in the energy scale as a function of time after injection. These corrections, which are of order 0.1 percent for times of interest for physics analysis, were computed to make the average energy of recorded pulses constant as a function of time. The method by which these corrections are obtained will be described in detail in Section 8.2, where a systematic uncertainty in $\omega_a$ is also attributed to them.

The signal cables to each detector position were cut to fit in each location, so the detectors do not naturally fall on the same absolute time scale. The $(g-2)$ precession phase $\phi_a$ was used to align all stations in time. Overall time offsets ranging from -10 ns to +100 ns were applied to force each station to have $\phi_a = 0$ over the energy range from 2.0 to 3.2 GeV.

The bunched nature of the beam causes a large modulation at early times after injection; the determination of the radial muon distribution from this modulation was discussed in Section 5.1. It is difficult to construct a functional form that describes this modulation with a statistically acceptable fit. Consequently, it is removed by adding random perturbations to the pulse times. For each detector in each fill, a random number is chosen uniformly over a range from -74.6 to +74.6 ns; this range corresponds to the cyclotron period. This random number is added to the time of the $t_0$ signal for that fill, which is in turn subtracted from the time associated with each positron pulse. Unfortunately, the use of random numbers introduces a dependence on the random number generator and its seed. The random number generator that is used [105] has a repetition period of $2^{19937}-1$. The final result is based on an average over data sets built with five different random seeds, and a systematic uncertainty will be attributed to this procedure.

# Chapter 7

# Determination of $\omega_a$

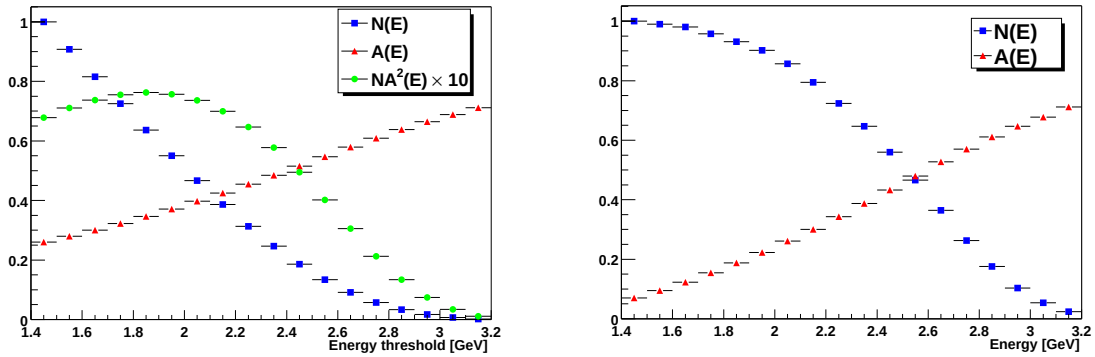
This chapter describes how the $(g-2)$ frequency $\omega_a$ is extracted from the distribution of decay positron events versus energy and time. This analysis uses a new method in which the data are divided into energy slices that are treated separately and then combined; the philosophy underlying this method is explained. The corrections to the functional form that are necessitated by CBO, pileup, and muon losses are identified. The quality checks that were employed to select a clean subset of the data are described. Finally, the results of the fitting procedure are shown.

## 7.1 Energy-binned analysis method

The conventional method of determining $\omega_a$ uses a single histogram constructed of the positron arrival times with respect to the injection time, including only positrons whose energies exceed a threshold. The statistical uncertainty of $\omega_a$ in this method is proportional to $\sqrt{NA^2}$, where $N$ is the number of counts and $A$ is the average $(g-2)$ asymmetry of the spectrum. As shown in Figure 7.1(a), $A$ increases as the energy threshold is raised, while $N$ decreases; $NA^2$ assumes its maximal value for a threshold between 1.8 and 2.0 GeV. Consequently, a conventional analysis would use a threshold near this value. Figure 7.1(b) shows $N$ and $A$ for narrow energy bands, rather than integrated for all energies over a threshold. Clearly, there is a significant amount of asymmetry even at energies below the optimal threshold. These data cannot simply be combined with the higher-energy positrons without diluting the overall asymmetry. However, they can be fitted separately, combining the resulting frequencies after the fact. Furthermore, the statistical precision of even the higher-energy data can be improved by dividing them into energy bins.

The energy-binning procedure is applied in this analysis, using bins of width 200 MeV, which is comparable to the energy resolution of the calorimeter. Each station is also treated separately, so the final result arises from a combination of $N_{stations} \cdot N_{E\ bins} = 22 \cdot 9 = 198$ individual fits (two stations are excluded for reasons that are explained in Section 7.5). The energy-binned approach provides an uncertainty that is 5 percent smaller than the conventional method when the same energy threshold of 2 GeV is used for both cases. Once data down to 1.4 GeV have been included, the uncertainty is reduced by 12 percent compared to a conventional analysis.

There are additional motivations for the energy-binned method:




(a) (b)

Figure 7.1: The quantities $N(E)$ and $A(E)$ plotted versus energy for (a) all energies greater than a threshold, and (b) energy bins. $NA^2$ is also shown in (a), illustrating its maximum between 1.8 and 2.0 GeV.

- Mapping $\omega_a$ as a function of positron energy provides a substantial consistency check. Unknown systematic errors that might bias the result differently for different energies can be excluded.
- Some known systematic biases, such as a time-dependent gain change, affect the conventional and energy-binned methods differently. A comparison of the results of the two methods tests both.
- Determining the energy dependence of parameters other than $\omega_a$ provides insight into subtle effects. The energy dependence of the CBO amplitudes and phases was discussed in Section 5.4; these data led to an understanding of how the CBO couple to the detector acceptance to affect the time spectrum.

The overall form of the function that is used to fit each energy bin in the data is

$$f(t) = e^{-t/\tau}\Lambda(t)N[C_1(t) - A(C_1(t) + C_2(t))\cos(\omega_a t + \phi_a + C_3(t))] + N_P(t)\ . \tag{7.1}$$

In addition to the terms in the ideal background-free function, this form includes parameters that account for CBO, overlapping pulses, and muon losses. These issues will be described in detail in the following sections. The frequency $\omega_a$ is further parameterized as

$$\omega_a = 2\pi[1 - (R + \Delta R)\cdot 10^{-6}](229.1\ \text{kHz})\ . \tag{7.2}$$

The quantity $R$ is therefore expressed in units of ppm, reducing numerical instabilities that might arise from sensitivity to trailing digits. The arbitrary offset $\Delta R = -14.744$ ppm was previously kept secret in order to facilitate a blind analysis. To eliminate possible psychological biases, the true values of $\omega_a$ and $\omega_p$ were never discussed within the collaboration until the entire analysis was finalized. Consequently, it was impossible for anyone to compute the physically significant quantity $a_\mu$ based on a preliminary result. These offsets were removed only after a unanimous secret ballot vote by the collaboration.

The form in Equation 7.1 was fit to the spectrum for each detector and each energy bin using a $\chi^2$ minimization carried out by the version of MINUIT [104] that has been incorporated into ROOT [101]. The $\chi^2$ is defined by

$$\chi^2 = \sum_k \frac{[N_k - f(t_k)]^2}{f(t_k)} \ , \tag{7.3}$$

where $k$ runs over the bins of the histogram that are included in the fitting range; the selection of this fitting range will be discussed in Section 7.6.

## 7.2 Coherent betatron oscillation terms

The coherent betatron motion of the beam leads to several physical effects. First, radial position oscillations modulate the overall count rate and the average drift time. Second, radial angle oscillations directly rotate the spin precession phase. These effects, which were discussed in Section 5.4, account for the factors $C_1(t)$ and $C_3(t)$ in Equation 7.1. They modulate the parameters $N$ and $\phi_a$, respectively, and are defined by

$$C_1(t) = 1 + e^{-t/\tau_{CBO}} E_{CBO}(t) A_{CBO1} \cos(\omega_{CBO} t + \phi_{CBO1}) \tag{7.4}$$

$$C_3(t) = e^{-t/\tau_{CBO}} E_{CBO}(t) A_{CBO3} \cos(\omega_{CBO} t + \phi_{CBO3}) \ . \tag{7.5}$$

Meanwhile, $C_2(t)$ arises purely mathematically from the combination of decay positrons with different energies into a single spectrum. This integration over energies is expressed by

$$f(t) = \int_{E_{low}}^{E_{high}} e^{-t/\tau} C_1(E,t) N(E) [1 + A(E) \cos(\omega_a t + \phi_a(E) + C_3(E,t))] \, dE \tag{7.6}$$

$$= e^{-t/\tau} \int_{E_{low}}^{E_{high}} C_1(E,t) N(E) \, dE \tag{7.7}$$

$$+ e^{-t/\tau} \int_{E_{low}}^{E_{high}} C_1(E,t) N(E) A(E)) \cos[\omega_a t + \phi_a(E) + C_3(E,t)] \, dE$$

$$\approx e^{-t/\tau} \int_{E_{low}}^{E_{high}} C_1(E,t) N(E) \, dE \tag{7.8}$$

$$+ e^{-t/\tau} \cos[\omega_a t + \phi_a(E) + C_3(E,t)] \int_{E_{low}}^{E_{high}} C_1(E,t) N(E) A(E) \, dE \ .$$

The observed spectrum is divided into two components. The first is not modulated at the anomalous precession frequency $\omega_a$, while the second is. The same overall rate modulation by the CBO $C_1(E,t)$ affects both components, but in one case it is multiplied by a weight $N(E)$ and in the other by $N(E)A(E)$. The asymmetry $A(E)$ is a strong function of energy, so the integrated effect of the CBO is different for the two

components. This difference introduces $C_2(t)$, defined by

$$C_2(t) = e^{-t/\tau_{CBO}} E_{CBO}(t) A_{CBO2} \cos(\omega_{CBO} t + \phi_{CBO2}) \; . \tag{7.9}$$

One of the original motivations for the energy-binned analysis method was to remove the $C_2(t)$ modulation of the $(g-2)$ asymmetry. However, even the combination of energies found in a 200 MeV energy bin, which includes additional energies that are folded in by the finite calorimeter resolution, is sufficient to produce this asymmetry modulation at a significant level. Consequently, it is necessary to account for this effect even in this analysis.

The function $E_{CBO}(t)$ describes the deviation of the envelope of the CBO from an exponential form. Physically, the envelope is determined by the dephasing of the oscillations of muons of different momenta. However, an experimental effect also contributes significantly; there is an admixture of slightly different CBO frequencies in the data, and a beating among these frequencies is observed in the envelope. The power supplies for the electric quadrupoles were turned off whenever the magnetic field mapping trolley was inserted. Afterwards, they were raised back to their nominal settings to within 0.1 kV, the precision of the front-panel indicators. Consequently, the *n* value of the ring was only constant at the 0.1 percent level. Also, early in the running period, the *n* value was deliberately adjusted slightly to move away from a resonance, and data from before and after this adjustment is included in the combined spectrum. The envelope was determined empirically by fitting the data in narrow time slices; the resulting lifetime is $\tau_{CBO} = 95$ $\mu$s, and the non-exponential part $E_{CBO}(t)$ is shown in Figure 7.2. The polynomial fit shown in that graph is used; it is cut off at the local maximum near 245 $\mu$s, so a simple exponential model is employed after this time.

Both the CBO frequency $\omega_{CBO}$ and its lifetime $\tau_{CBO}$ are found to depend on the decay positron energy as shown in Figure 7.3. The mechanism that leads to this relationship is not clear. These energy-dependent values are used in the fit, but a conservative systematic error will be attributed to the procedure.

For numerical stability, all three of the CBO amplitudes and phases are rewritten in terms of sine and cosine components. For example,

$$A_{CBO1C} = A_{CBO1} \cos\phi_{CBO1} \text{ and} \tag{7.10}$$

$$A_{CBO1S} = A_{CBO1} \sin\phi_{CBO1} \tag{7.11}$$

so that $C_1(t)$ becomes

$$C_1(t) = 1 + e^{-t/\tau_{CBO}} E_{CBO}(t) [A_{CBO1C} \cos\omega_{CBO} t + A_{CBO1S} \sin\omega_{CBO} t] \; . \tag{7.12}$$

The reformulation leads to a single unambiguous solution. In contrast, the presence of a phase parameter like $\phi_{CBO1}$ brings about an infinite number of identical optimal points at intervals of $2\pi$.

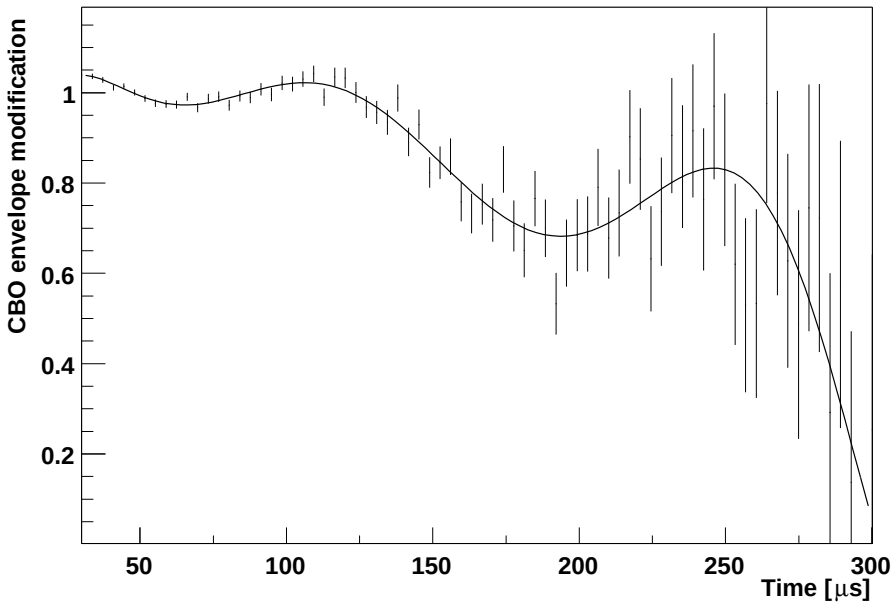
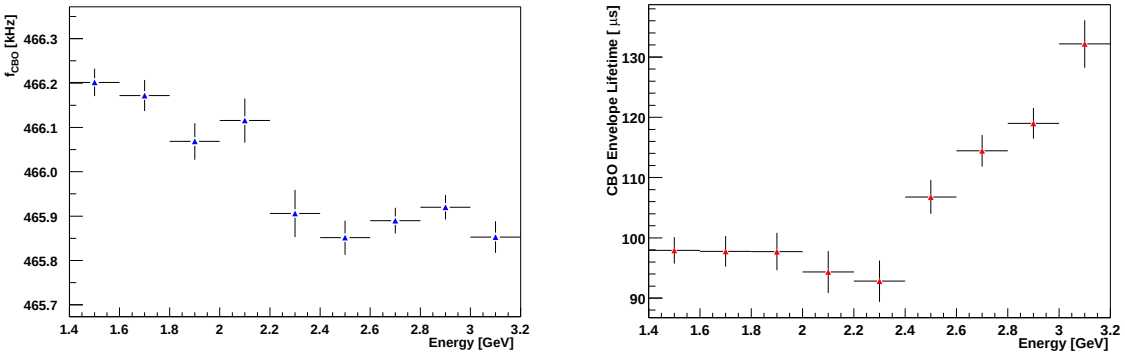



Figure 7.2: Non-exponential part of CBO envelope function $E_{CBO}(t)$, including a polynomial fit to the data.




(a) $f_{CBO}$ (b) $\tau_{CBO}$

Figure 7.3: Energy dependence of the CBO frequency $f_{CBO} = \omega_{CBO}/2\pi$ and lifetime $\tau_{CBO}$.

## 7.3 Overlapping pulse correction

The spectrum of overlapping pulses was constructed following the procedure described in Section 6.10. The pileup parameters for each station are determined by fitting these data by energy bin to a functional form. The resulting parameterization is then inserted into the ordinary fits to that station's data as $N_P(t)$ in Equation 7.1. This approach was chosen instead of the alternative of directly subtracting the constructed pileup from the ordinary data because it allows the simple definition of $\chi^2$ in Equation 7.3 to be used in the ordinary fits. After pileup has been subtracted, the data are no longer drawn from a Poisson distribution, and a much more complicated definition is necessary [106].

The distribution of overlapping pulses in each energy bin is described by a function of the form

$$\begin{aligned} f_P(t) &= e^{-2t/\tau} F_{FR}(t) \\ &\quad \left[1 + e^{-t/\tau_{CBO}} E_{CBO}(t) A_{CBO1} \cos(\omega_{CBO} t + \phi_{CBO1})\right]^2 \\ &\quad \left[N_P - A_{P1}\cos(\omega_a t + \phi_{P1}) - A_{P2}\cos(2(\omega_a t + \phi_{P2}))\right] . \end{aligned} \tag{7.13}$$

This function arises from the convolution product in Equation 6.1. It contains a non-oscillating component along with both the first and second harmonics of $\omega_a$.

The factor $F_{FR}(t)$ accounts for the enhancement of the pileup by the fast rotation; the randomization procedure averages away the modulation itself, but the effect of the bunched beam on the instantaneous rate remains. The calculation of this factor involves a time spectrum $N(t)$ divided into bins of width $\tau_c/30$. Then

$$F_{FR}(t) = \frac{< N^2(t) >}{< N(t) >^2} , \tag{7.14}$$

where the brackets imply an average over a fast rotation cycle. This enhancement factor is shown in Figure 7.4; it is quite small by 40 $\mu$s.

The form of Equation 7.13 is fit to the constructed pileup spectrum, of which an example appears in Figure 7.5. The fitting process involves nearly the same $\chi^2$ minimization procedure as is applied to the ordinary data. However, in this case the $\chi^2$ is defined by

$$\chi^2 = \sum_k \frac{[P_k - f_P(t_k)]^2}{(\sigma_P)_k^2} \text{ with} \tag{7.15}$$

$$\sigma_P = f_{TS/SS}(t)\sqrt{D_{TS}(E,t) + S_S(E,t)} . \tag{7.16}$$

The uncertainty $\sigma_P$ is an approximation, since it implies perfect knowledge of $f_{TS/SS}(t)$. The lifetime $\tau$ and the frequency $\omega_a$ are fixed in these fits, at values of $\tau = 64.4076$ $\mu$s and $R = 128.55$ ppm. The CBO frequency and lifetime are also fixed to the same values that are used in fits to the ordinary spectra.

Averaged over the 22 included stations, the results of these fits as a function of energy are shown in Figure 7.6. All of the parameters behave as would be expected from calculation or simulations, except that the $\chi^2$/dof for the lowest energy bins is slightly too high. The statistical uncertainty for the pileup spectrum is smallest at these energies, so the fits are sensitive to corrections such as muon losses that were omitted

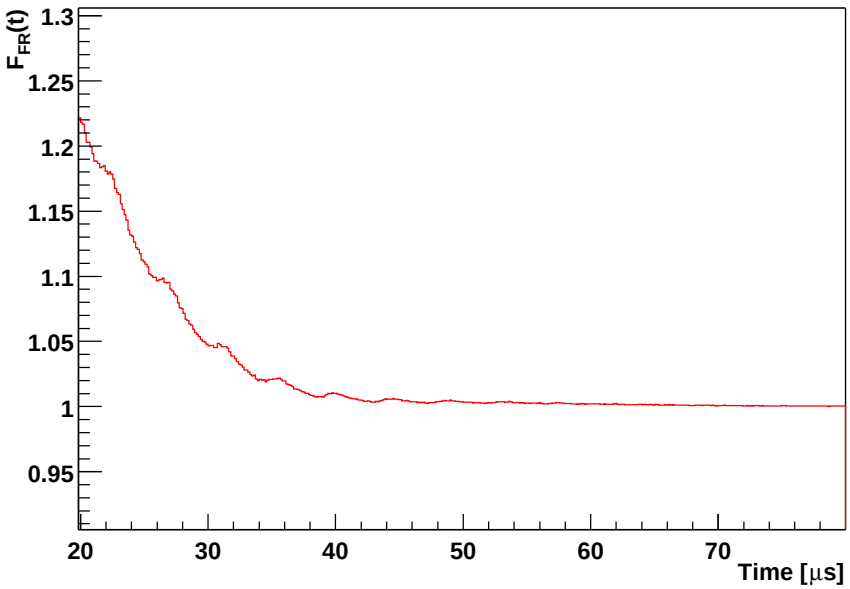



Figure 7.4: Fast rotation enhancement of pileup amplitude $F_{FR}(t)$.

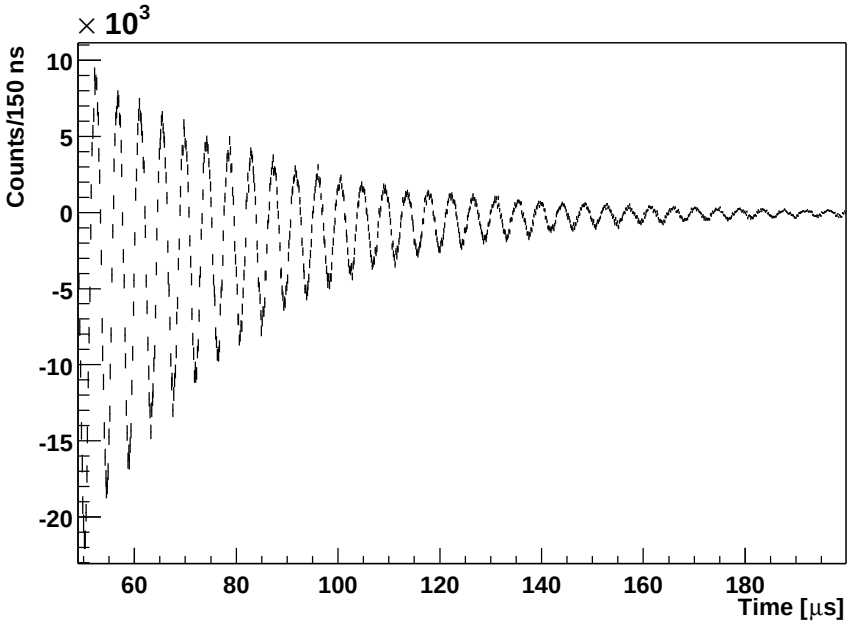



Figure 7.5: Constructed pileup spectrum for energies from 2.4 to 2.6 GeV, summed over detector stations. In this energy range, the non-oscillating component is small but the oscillating component is large.

from Equation 7.13.

## 7.4 Muon loss correction

In the absence of muon losses and, for simplicity, averaging over the $(g-2)$ cycle, muon decay is described by the differential equation

$$\frac{dN}{dt} = -\frac{N}{\tau} \,. \tag{7.17}$$

This is a separable first-order ordinary differential equation that is solved in the usual way:

$$\int_{N_o}^{N} \frac{1}{N'}\, dN' = \int_{t_o}^{t} -\frac{1}{\tau}\, dt' \tag{7.18}$$

$$\ln N_o - \ln N = -\frac{t_o}{\tau} + \frac{t}{\tau} \tag{7.19}$$

$$N(t) = N_o \exp\left[-(t-t_o)/\tau\right] \tag{7.20}$$

Muon losses simply add another mechanism by which muons disappear from the population, so the differential equation becomes

$$\frac{dN}{dt} = -N\left[\frac{1}{\tau} + f_\Lambda(t)\right] \,. \tag{7.21}$$

Here $f_\Lambda(t)$ is the fractional muon loss rate. This differential equation is solved as before:

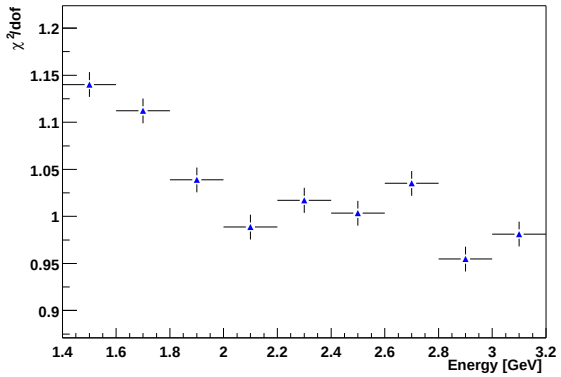
$$\int_{N_o}^{N} \frac{1}{N'}\, dN' = -\int_{t_o}^{t} \frac{1}{\tau} + f_\Lambda(t')\, dt' \tag{7.22}$$

$$\ln N_o - \ln N = -\frac{t_o}{\tau} + \frac{t}{\tau} - \int_{t_o}^{t} f_\Lambda(t')\, dt' \tag{7.23}$$

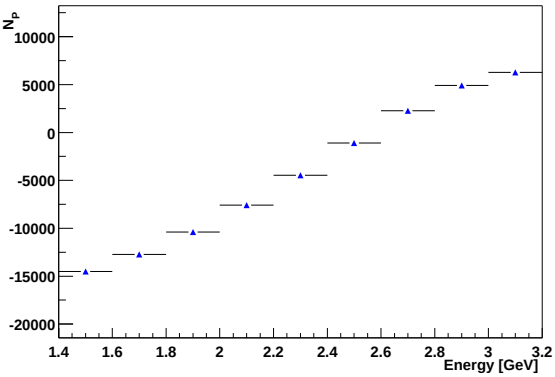
$$N(t) = N_o \exp\left[-(t-t_o)/\tau\right] \exp\left[-\int_{t_o}^{t} f_\Lambda(t')\, dt'\right] \,. \tag{7.24}$$

In this evaluation, $t_o$ may be chosen arbitrarily, even independently in the decay and muon loss terms. If a nonzero value is chosen, the only side effect is that $N_o$ no longer has a clear physical interpretation.

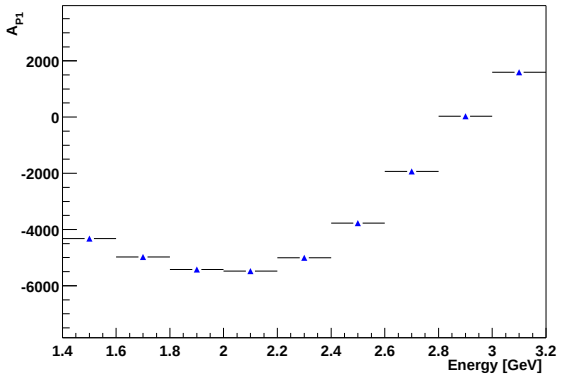
Muons with energies near 3.1 GeV typically pass through a calorimeter with only a minimum-ionizing energy deposit of approximately 100 MeV. Therefore, the muon loss rate is measured by searching for coincident signals in three consecutive FSDs, with less than 1 GeV of energy deposited in the first calorimeter, as illustrated in Figure 7.7 [107]. Because muon loss events are quite rare, there is a considerable background from accidental coincidences of three low-energy positrons. The distribution of these background events is determined and subtracted by looking for coincidences in an out-of-time interval, following in the spirit of the overlapping pulse correction discussed in the previous section. The procedure is complicated somewhat by the fact that the three detectors are independent; there are additional combinations to subtract, such as a



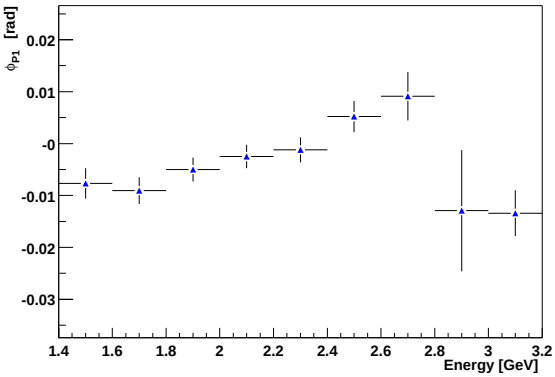

(a) $\chi^2$/dof



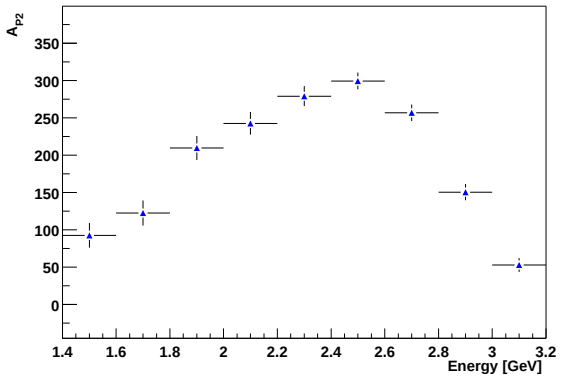

(b) $N_P$



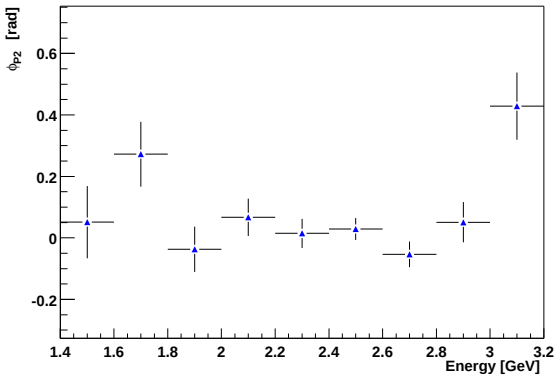

(c) $A_{P1}$




(d) $\phi_{P1}$




(e) $A_{P2}$




(f) $\phi_{P2}$

Figure 7.6: Fit parameters for constructed pileup spectrum by energy bin, averaged over detectors.

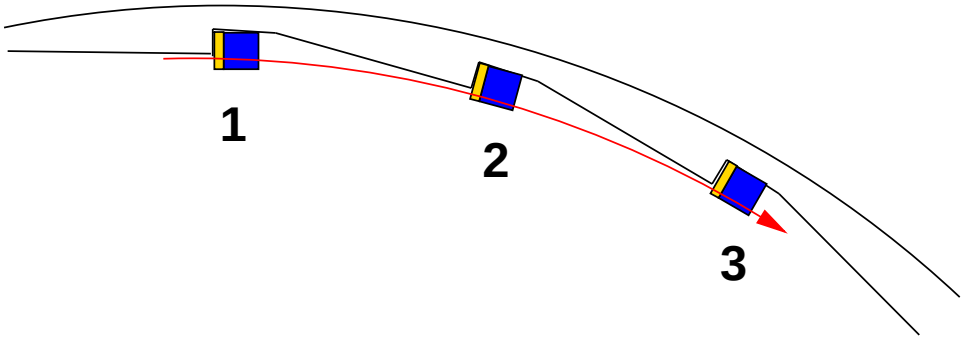



Figure 7.7: The detected lost muon signal, with muon passing through three detector stations, leaving a signal in the FSDs (gold) but only minimum-ionizing energy in the calorimeters (blue).

real two-fold coincidence together with an uncorrelated particle in the third detector. A similar construction is used to correct for accidental energy deposited in the first calorimeter, and also for the deadtime of the discriminators.

The other significant background to the measurement arises from losses of protons, which are estimated to account for 10 percent of the beam just prior to the injection point. Because they do not decay, their relative abundance increases with time, becoming essentially 100 percent by the end of the fill. They do not deplete the population of muons, so they should not be included in $f_\Lambda(t)$, and the time dependence of their loss is different than that of the muons. Unlike muons, which almost never initiate a shower in a calorimeter, protons sometimes do. Consequently, there is a measurable category of "visible" proton events, which are events that meet the criteria to be counted in the muon loss spectrum but which also have an energy greater than 1 GeV in the third calorimeter. The time dependence of these events is measured; it is normalized using the ratio of "invisible" to "visible" protons that are dumped as the quadrupoles are turned off at the end of the fill.

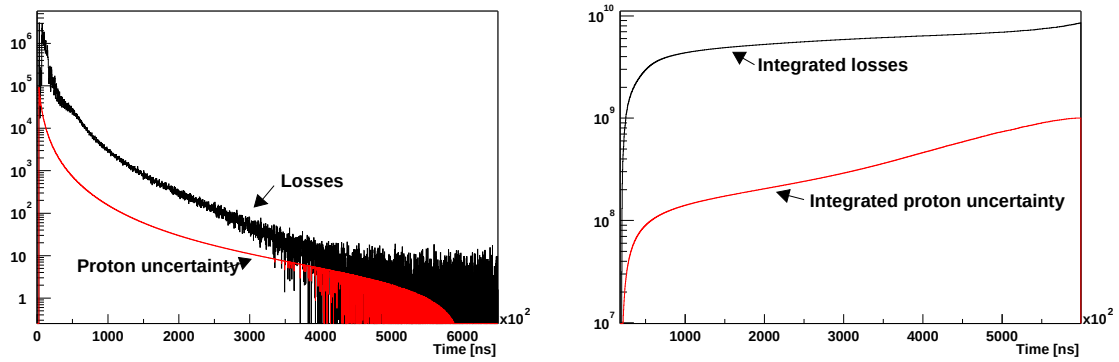
The time spectrum of the triple FSD coincidences is proportional to the absolute muon loss rate:

$$f_\Lambda^{abs}(t) = N_\Lambda f_\Lambda^{FSD}(t) = N(t) f_\Lambda(t) \ . \tag{7.25}$$

The FSD triple coincidence counter subtends only a small part of the solid angle into which muons can be lost. Its efficiency is small, probably less than 1 percent, but essentially unknown; it depends strongly on the precise nature of the mechanism that causes the losses. Consequently, the normalization parameter $N_\Lambda$ must be included as a floating term in the fit. We make a first-order approximation at this point and, temporarily neglecting losses, use $N(t) \propto e^{-t/\tau}$ to determine $f_\Lambda(t)$ via

$$f_\Lambda(t) = N_\Lambda e^{t/\tau} f_\Lambda^{FSD}(t) \ . \tag{7.26}$$

Because the ordinary fit to the data is so sensitive to backgrounds, it is allowed to adjust the muon loss shape within its error band. An error band $\delta f_\Lambda^{FSD}(t)$ is determined from the uncertainty in the proton




(a) Before time integration

(b) After time integration

Figure 7.8: Muon loss function $f_\Lambda(t)$ together with its uncertainty $\delta f_\Lambda(t)$.

correction. The loss shape is actually allowed to vary by a multiple of this error band:

$$f_\Lambda(t) = N_\Lambda e^{t/\tau}[f_\Lambda^{FSD}(t) + N_{\delta\Lambda}\delta f_\Lambda^{FSD}(t)] \; . \tag{7.27}$$

The final form of the muon loss factor is therefore

$$\Lambda(t) = \exp[-N_\Lambda/\tau \int_{t_o}^{t} e^{t/\tau}(f_\Lambda^{FSD}(t') + N_{\delta\Lambda}\delta f_\Lambda^{FSD}(t'))dt'] \tag{7.28}$$

The muon loss functions are shown in Figure 7.8. The overall normalization factors $N_\Lambda$ and $N_{\delta\Lambda}$ were determined once from a global $\chi^2$ minimization and were thereafter fixed at the values $N_\Lambda = 1.62 \times 10^{-7}$ (arbitrary units) and $N_{\delta\Lambda} = 2.53$ (dimensionless). This level corresponds to a relative loss rate of order $10^{-3}$ per muon lifetime.

## 7.5 Data selection

The data are divided into numbered runs. They are of varying length, but a typical run represents about 45 minutes of data collection. Some of these runs are not suitable for inclusion in the data set to be analyzed for $\omega_a$; the runs may have been taken as part of a special study under unusual conditions, or there may have been operational problems that invalidate the data. A list of acceptable runs was constructed. First, the experimental logbooks were reviewed to exclude runs that are known to contain artifacts such as LED test pulses or shortened electric quadrupole firings. Runs taken during intervals where the fixed NMR probe data was not available were also removed at this point. Finally, a set of reference runs was selected, and each of the other runs was statistically compared against the reference set. To do this, an unbinned Kolmogorov-Smirnov test [108] was applied to the $(E,t)$ spectra.

Some runs that contain laser calibration pulses are included in the data set, because the laser fires only

| Issue | Number of positron events |
|---|---|
| Total included (not cut) | $3.73 \times 10^9$ |
| Laser fills | $9.1 \times 10^7$ |
| Data processing faults (est.) | $6.1 \times 10^7$ |
| Quadrupole trace not recorded | $4.8 \times 10^7$ |
| Number of $t_0$ events $\neq 1$ | $2.5 \times 10^7$ |
| $t_0$ trace not recorded | $5.0 \times 10^6$ |
| Possible quadrupole spark | $1.6 \times 10^6$ |

Table 7.1: Number of events with $E > 2.0$ GeV and $t > 50$ $\mu$s lost due to various cuts (other than run selection) compared with the number of events included.

in every other fill. The subset of the fills that do contain laser pulses are identified and excluded using the digitized signal from the laser reference photodiode. The AGS extraction system occasionally fails in a way that leaves a bunch circulating in the AGS on a destabilized orbit so that a number of protons leak down the primary beamline every turn. This background is removed by excluding fills where no pulse is seen in the $t_0$ counter; the intensity of the extraneous particles is not sufficient to trigger it. The electric quadrupoles were pulsed for varying lengths of time, and they occasionally spark, though less frequently than once in $10^5$ fills. The digitized traces from the monitoring probes connected to them were checked to ensure that the trace extends to at least 700 $\mu$s after the data acquisition start signal, or about 640 $\mu$s after injection. Finally, a small amount of data is excluded solely because of data processing problems. Some statistics on the number of positrons lost by each of these cuts are given in Table 7.1.

Two of the 24 detector stations had hardware faults that force them to be excluded from the analysis. The WFD for station 2 had a problem that led to correlations between consecutive samples. Consequently, the computed average pulse shape is quite distorted. The pulse fitting procedure does not work properly with this pulse shape. The energy distribution in station 20 is very different from that of the other detectors because the traceback chambers are located directly upstream from it. These chambers and the intervening air gaps constitute a substantial amount of matter in which decay positrons deposit energy before reaching the calorimeter. More importantly, station 20 has a large background, uniform in time and quite visible at late times, that is not present in other detectors. The source of this background is not understood. This station is therefore excluded from the final result as well.

## 7.6 Fitting range

Although corrections are applied for muon losses, overlapping pulses, and energy scale changes, these corrections are determined only approximately. For example, the shape of the muon loss spectrum varies somewhat with position in the storage ring, implying that there may be several distinct muon loss mechanisms. Not all locations in the storage ring are instrumented with FSDs, so the total muon loss function cannot be completely determined. No attempt is made to account for three-particle pileup, and fluctuations in the pulse shape can lead to effects that are not accounted for in the pulse fitting procedure, so the construction of

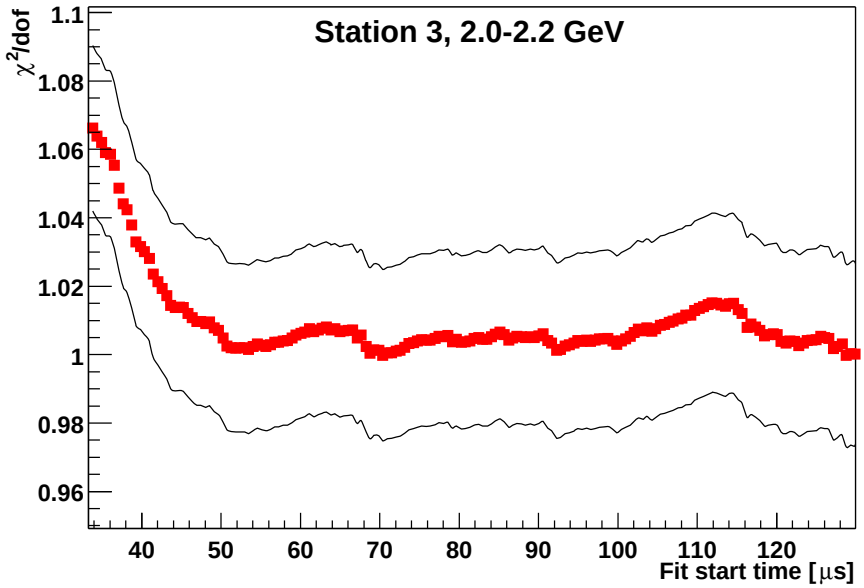



Figure 7.9: Trend of $\chi^2$/dof versus fit start time for a single energy bin; the algorithm chooses an earliest allowed start time of 51.29 $\mu$s for this bin. The outer envelope represents the error contour determined from $\sqrt{2/N_{dof}}$.

overlapping pulses is incomplete. Energy scale changes are constructed by detector station, but some photomultipliers may require different corrections than others within the same station. Also, the determination of the average energy as a function of time is itself biased by overlapping pulses. These effects have a very small impact on the fitted value of $\omega_a$, and systematic uncertainties will be computed and attributed to each of them. However, they have a large impact on the $\chi^2$/dof that is used to judge the quality of the fit.

It is possible to account for these effects by introducing empirically determined "residual slow effect" terms into the fitting function. This procedure was demonstrated to work in the context of other analyses of this data set. However, it is extremely cumbersome to construct these functions for the energy-binned analysis method. The impact of each of the contributing effects tends to vary by detector station and by energy, so a different function would be needed for each of the 198 fits. Consequently, it was determined to choose start times for each fit so that only time bins that can be described with physically motivated terms are included. The criterion that is used to determine these times is the stability of the $\chi^2$/dof versus fit start time. Figure 7.9 shows a rather typical example of the trend that the $\chi^2$/dof follows. The point at which it becomes statistically constant is identified and chosen for the final fits.

The standard deviation of the (non-reduced) $\chi^2$ distribution with $N_{dof}$ degrees of freedom is $\sqrt{2N_{dof}}$. When comparing $\chi^2$ values from different fit start times, many of the degrees of freedom are common to the two times; these should not be included when determining the standard deviation of the difference between the two values. The correlated error between the $\chi^2$/dof at time $t_1$ (with $N_{dof1}$ degrees of freedom) and the

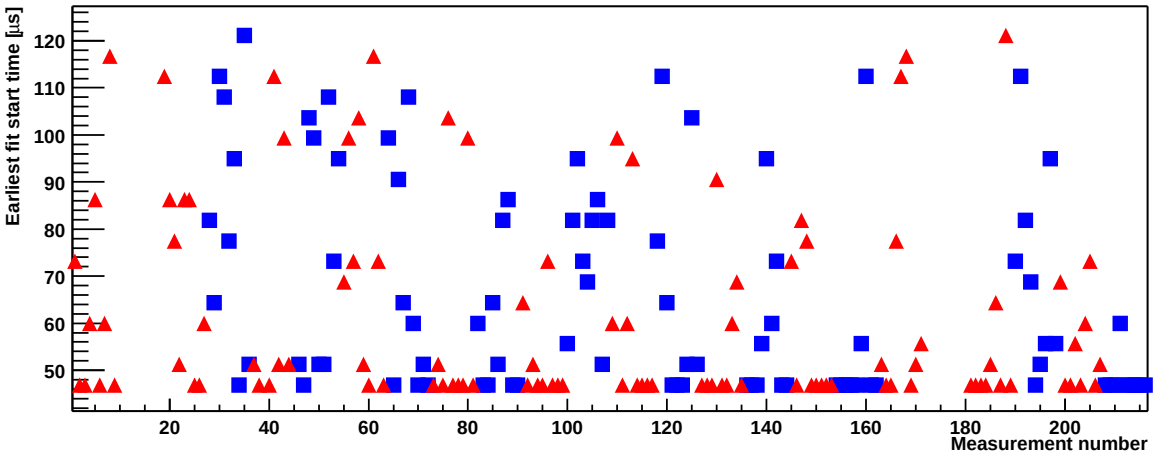



Figure 7.10: Earliest start times used for each energy bin within each detector. The nine leftmost points refer to station 1 and are arranged in order of increasing energy. The next nine points are absent, corresponding to the nine energy bins of station 2; other stations follow similarly. The marker color and style alternate by station as well.

same quantity at time $t_2$ (with $N_{dof2}$ degrees of freedom) is then [109]

$$\frac{\sqrt{2(N_{dof1} - N_{dof2})}}{N_{dof2}} \,. \tag{7.29}$$

The fit start time that was chosen was the first for which no more than 15.9 percent of the subsequent points fall above this correlated error band and no more than 15.9 percent fall below it. By applying this criterion, the start times shown in Figure 7.10 were obtained.

The $\chi^2$ minimization procedure requires each bin in the fitting range to have enough counts that its statistical distribution is approximated well by a Gaussian. In order to ensure that this requirement is met, the last time bin to be included in each fit is selected to be the first bin with fewer than twenty counts.

## 7.7 Fit results

Figures 7.11 and 7.12 show the fit parameters as a function of energy. For the parameters that are not related to CBO, an ordinary weighted average is used to combine the 22 included detectors. The CBO parameters require more care. CBO amplitudes are small quantities that are always positive. When they are averaged in this way, the result is biased towards larger values since the error bars on the input points are not actually symmetric. Consequently, these parameters are rotated by $2\pi/(\text{station} - 6)$ radians to align them geometrically with station 6, and the sine and cosine parts are then averaged separately. The amplitude and phase are computed from the combined sine and cosine parts.

All of the parameters behave as expected, except that the fitted lifetime $\tau$ varies in a statistically sig-

nificant way as a function of energy. Figure 7.13 shows that this discrepancy is much larger for a data set without gain corrections applied or with exaggerated gain corrections. The gain corrections are only believed to be valid at the level of 30 to 40 percent; this figure demonstrates that the variation of $\tau$ versus energy may be explained entirely by this level of inaccuracy in the gain corrections. In particular, this graph must not be read as an indication of a physically different lifetime as a function of positron energy; the fitted lifetime is constant within its own systematic uncertainty.

The fit parameters are shown as a function of fit start time in Figures 7.15 and 7.16. Most of the parameters vary dramatically by energy, so it is necessary to move the values onto the same scale in order to create a reasonable weighted average. Also, in order to avoid artificial inflation of the CBO-related amplitudes from the asymmetric error bars, the sine and cosine components are shown separately. The parameters fall into three categories:

- $R$, $\tau$, $\chi^2$/dof: these parameters are in principle energy-independent, so they are not transformed.
- $N$, $A$, $A_{CBO1C}$, $A_{CBO1S}$, $A_{CBO2C}$, $A_{CBO2S}$, $A_{CBO3C}$, $A_{CBO3S}$: these parameters are amplitudes, and they vary substantially with energy. The values in each energy bin are rescaled to pass through 1 at 49.1 $\mu$s. Consequently, the stability plot indicates relative changes.
- $\phi_a$: this parameter is a phase, and it varies with energy. The values in each energy bin are shifted additively to pass through 0 at 49.1 $\mu$s. The stability plot shows phase changes in units of radians.

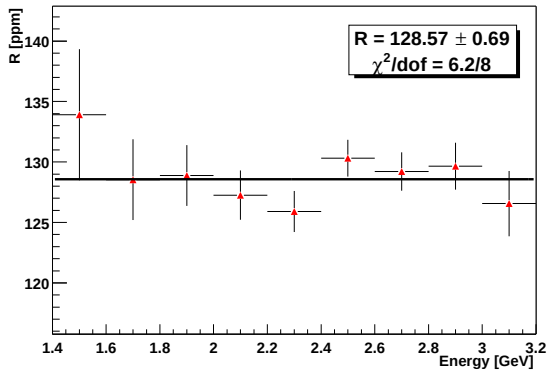
The data from a station and energy bin are only included at times after the earliest allowed start time. However, so that the statistical error bar increases monotonically with time, the value from the earliest start time is included in all previous points. Consequently, these are actually "pseudo-start time" scans. The correlated error band that is shown is computed from the formula [110, 111]

$$\sigma_{AB} = \sqrt{\sigma_A^2 - \sigma_B^2}\,, \tag{7.30}$$

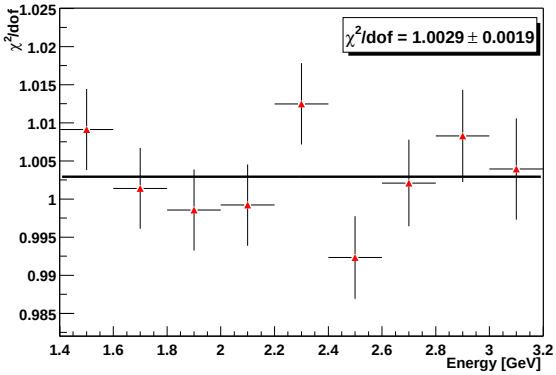
which applies when the set $A$ of included data is a subset of a larger data set $B$, as is the case when comparing a later fit start time to an earlier one. This is a one standard deviation uncertainty, so approximately two-thirds of the later points should fall within the correlated error band. Most parameters do not stray too far from this relation. Indeed, Figure **??** shows the distribution of differences in $R$ between adjacent start time points, normalized by the correlated error of Equation 7.30. Indeed, Figure 7.17 shows the distribution of point-to-point differences between fitted $R$ values at consecutive fit start times, normalized to the correlated error from Equation 7.30. The distribution is centered at 0, suggesting that there is not a statistically significant trend versus fit start time. However, it is approximately 10 percent too wide, and it is not quite Gaussian. This feature is typically the signature of a small amount of "phase pulling," an oscillation in the fitted frequency that is caused a background that is not included in the fitting function.

The final value of $R$ resulting from this analysis is

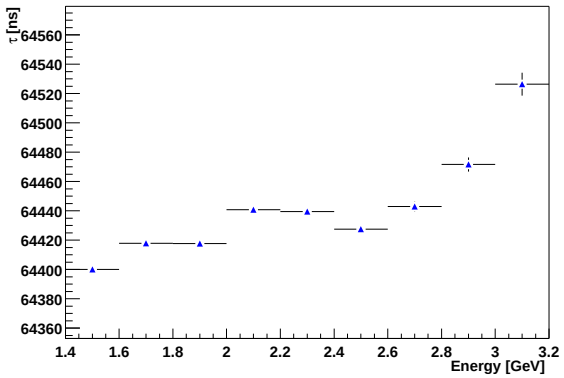
$$R = 128.71 \pm 0.69\ (\text{stat.})\ \text{ppm}. \tag{7.31}$$



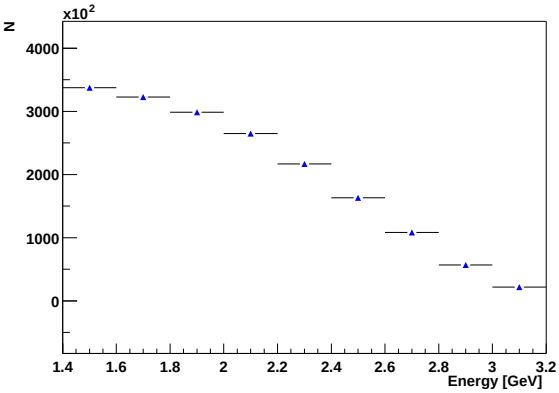

(a) $R$



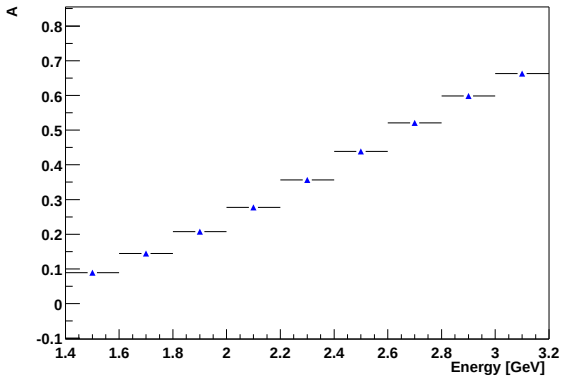

(b) $\chi^2$/dof



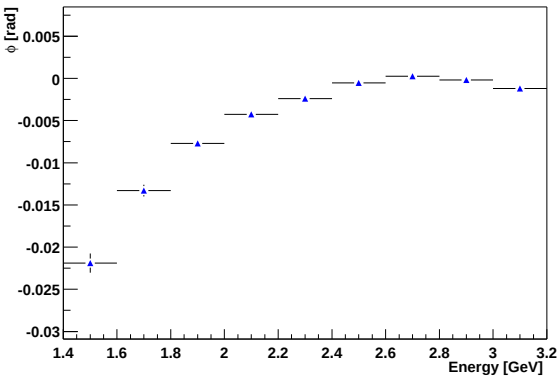

(c) $\tau$




(d) $N$

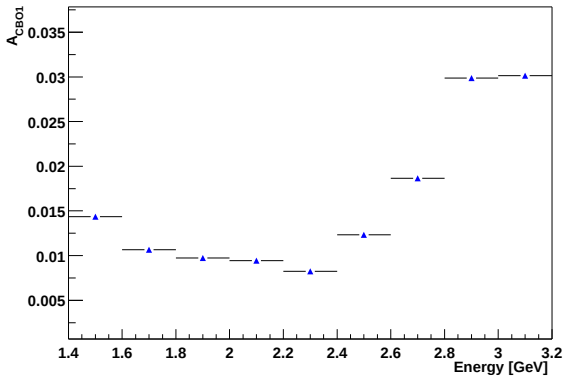



(e) $A$

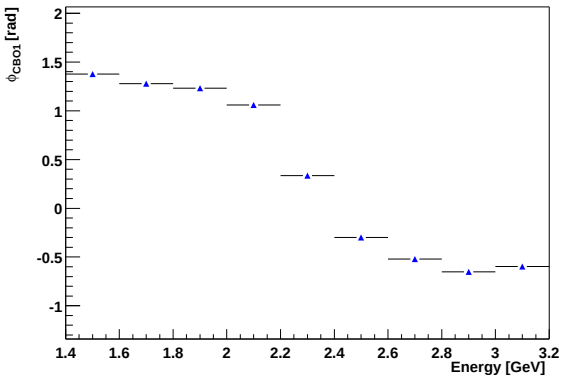



(f) $\phi_a$

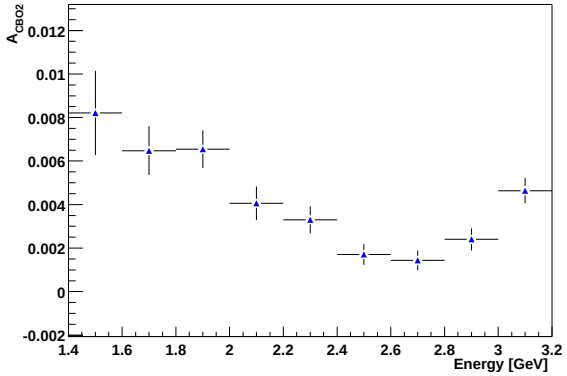
Figure 7.11: Non-CBO parameters as a function of energy after averaging over detectors. All behave as expected except for the lifetime $\tau$, which is discussed further in the text.



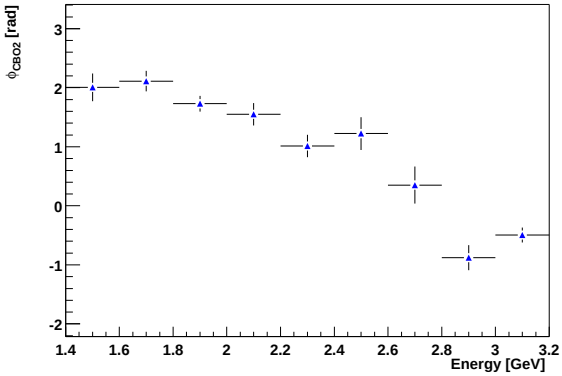

(a) $A_{CBO1}$



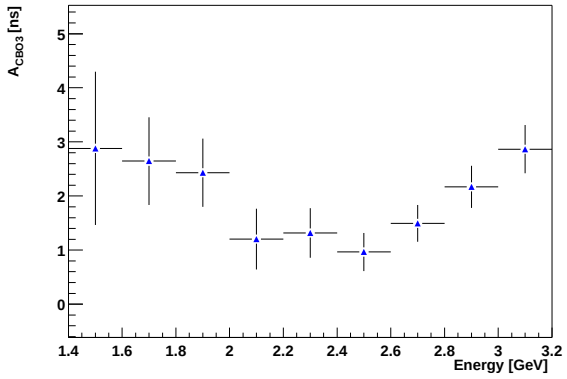

(b) $\phi_{CBO1}$



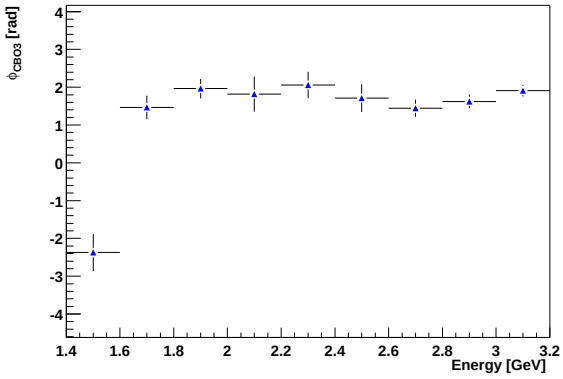

(c) $A_{CBO2}$




(d) $\phi_{CBO2}$




(e) $A_{CBO3}$




(f) $\phi_{CBO3}$

Figure 7.12: CBO parameters as a function of energy. Stations have been combined with a vector sum after geometric phase alignment. The energy dependence is described by the model explained in Section 5.4.

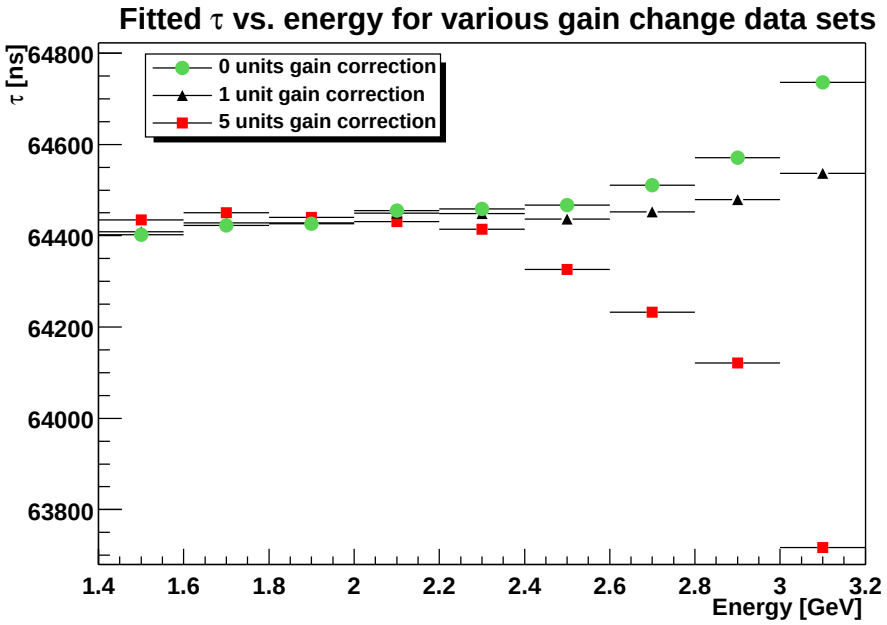
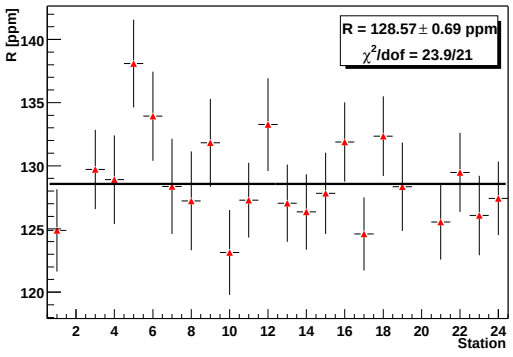
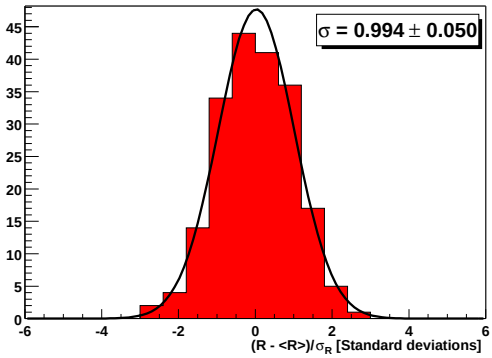



Figure 7.13: Fitted muon lifetime as a function of energy for data sets with different gain corrections applied. Uncertainty in the gain corrections at the level of 30 percent can account for the fact that the lifetime for the data set with one unit of gain correction is not flat versus energy.

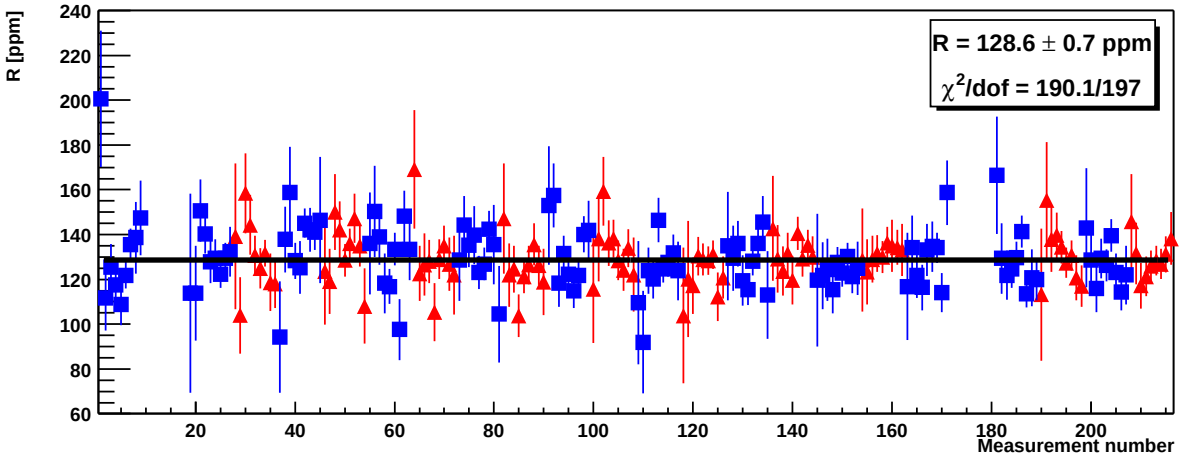
This value is averaged over five data sets built with different seeds for the fast rotation randomization process, whereas most of the systematic checks to be described in the next chapter were performed with only a single random seed. In Chapter 10, this value will be compared against other analyses of the same data set, and it will be combined with the average magnetic field to compute $a_\mu$.




(a) By station

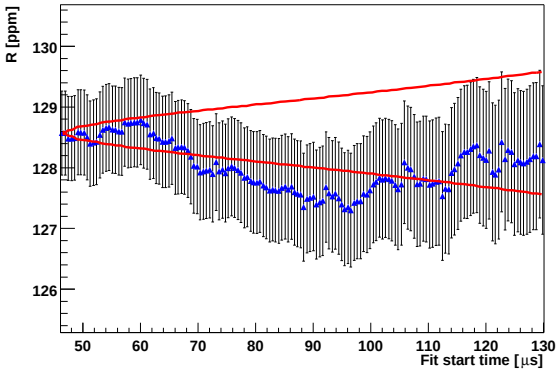



(b) Distribution

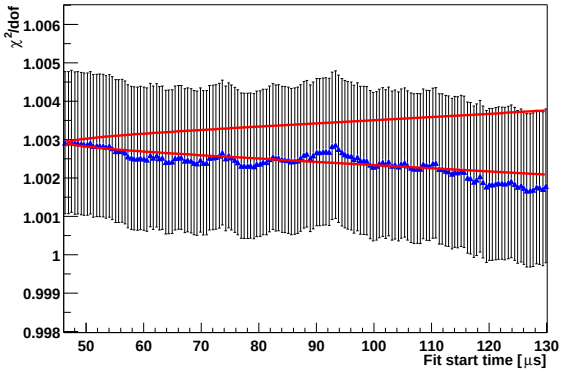



(c) All results together

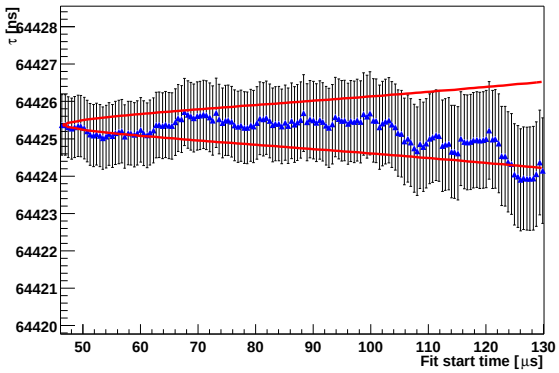
Figure 7.14: Results for $R$ (a) by station, (b) as a distribution, and (c) with all 198 results shown together, demonstrating the $R$ is consistent in all cases.



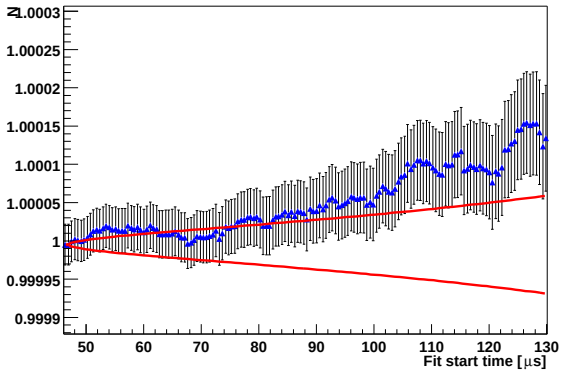

(a) $R$



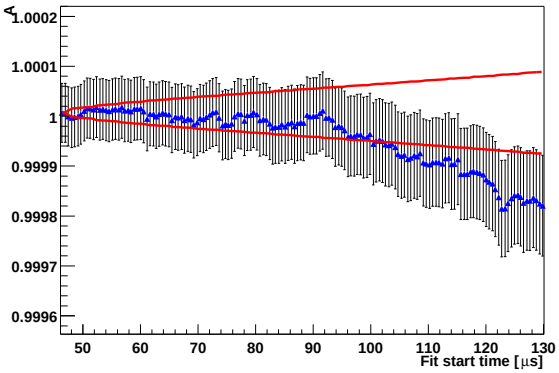

(b) $\chi^2$/dof



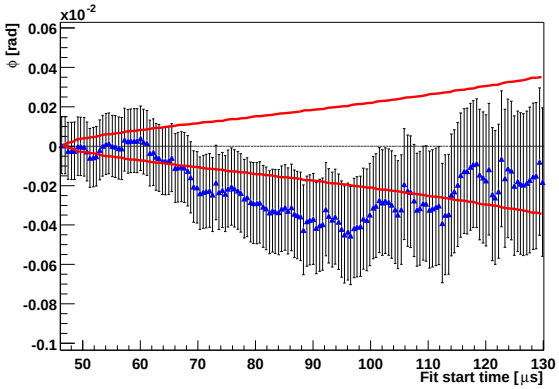

(c) τ




(d) $N$

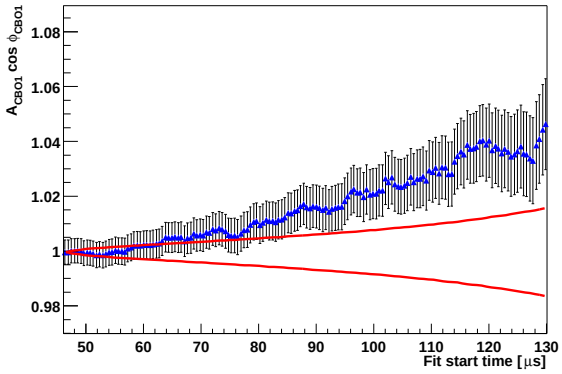



(e) $A$

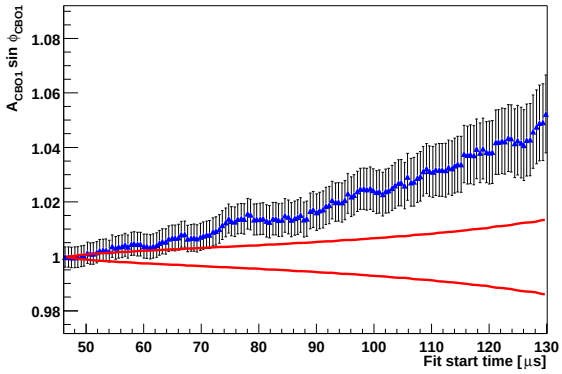



(f) $\phi$

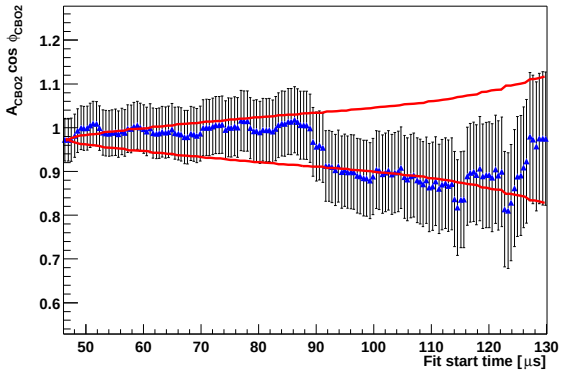
Figure 7.15: Fit start time scans for non-CBO related fit parameters.



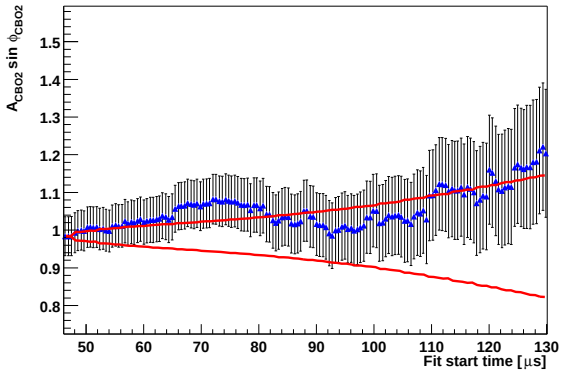

(a) $A_{CBO1C}$



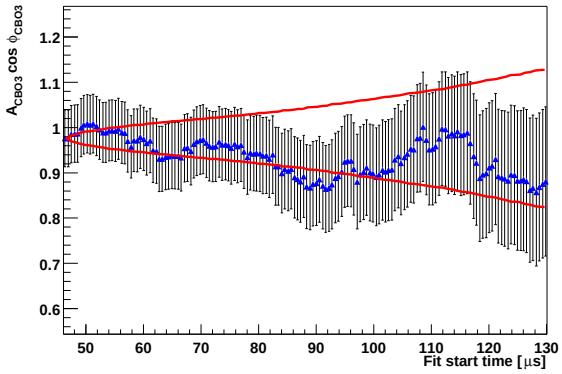

(b) $A_{CBO1S}$



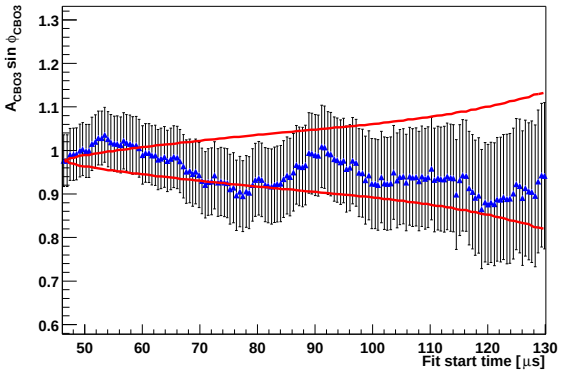

(c) $A_{CBO2C}$




(d) $A_{CBO2S}$




(e) $A_{CBO3C}$




(f) $A_{CBO3S}$

Figure 7.16: Fit start time scans for CBO related fit parameters.

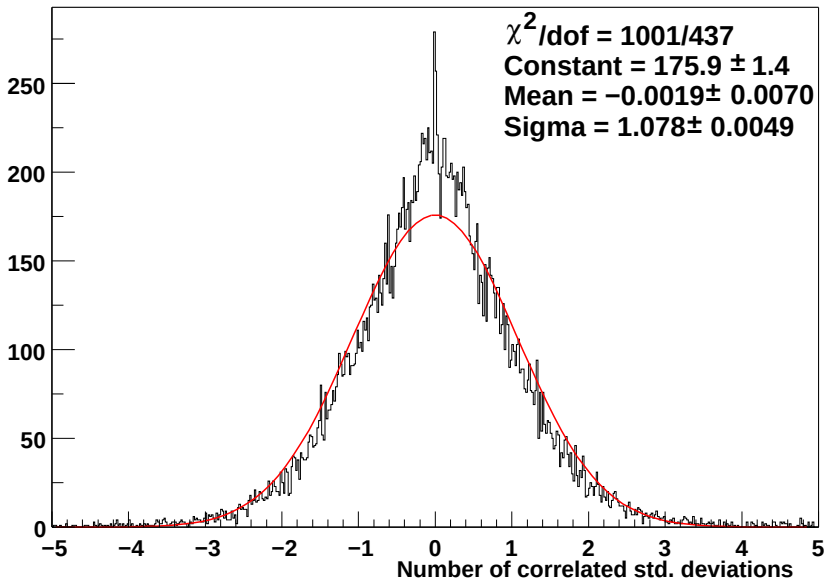



Figure 7.17: Distribution of point-to-point differences between *R* values at successive fit start times in each energy bin, normalized by the correlated uncertainty.

# Chapter 8

# Systematic uncertainties in $\omega_a$

This chapter describes the systematic uncertainties introduced in the determination of $\omega_a$. The largest of these effects, which are summarized in Table 8.1, arise from coherent betatron oscillations and overlapping pulses. The general technique used to evaluate most of the systematic uncertainties begins by defining parameters that characterize the scale of the effect. Next, the sensitivity of the fitted value of $R$ to the scale parameter is determined for fit start times used in the final result. Finally, an estimate of the scale parameter is made.

## 8.1 CBO "half-ring" effects

The functional form used to fit the data (Equation 7.1) contains factors $C_1(t)$, $C_2(t)$, and $C_3(t)$ that account for three different effects of the CBO: its modulations of the overall rate $N$, the $(g-2)$ asymmetry $A$, and the $(g-2)$ phase $\phi_a$. The first effect, an overall rate modulation, is clearly present in a Fourier transform of the time spectrum, and it is impossible to obtain an acceptable fit without accounting for it. The other effects are more subtle; indeed, the data were initially fit without $C_2(t)$ and $C_3(t)$. The resulting dependence of $R$ on the detector station number is shown in Figure 8.1. Using this simplified fitting function, the value of $R$ is not consistent around the ring; it appears to vary sinusoidally with an amplitude of about 3 ppm. A systematic bias of some sort is demonstrated even more clearly in Figure 8.2, which shows the trend versus fit start time for the average $R$ value for two halves of the ring, consisting of stations 1–12 and 13–24. The apparent discrepancy between the two halves of the ring increases and decreases as the phase of the sinusoidal curve versus detector number in Figure 8.1 changes with fit start time.

The frequency spectra of the residuals following fits with various functions are shown in Figure 8.3. To make these graphs, a fit is performed, the differences between the data points and the fit function are computed, and a Fourier transform is applied to the resulting deviations. The simplest function does not account for the CBO at all; a large peak appears at the frequency $\omega_{CBO}/2\pi = 465$ kHz, together with sidebands at $(\omega_{CBO} \pm \omega_a)/2\pi = 236$ kHz, 694 kHz. A somewhat more complete function includes the overall rate modulation $C_1(t)$ but sets $C_2(t) = C_3(t) = 0$. It successfully removes the primary CBO peak from the Fourier transform of the residuals, and it reduces the amplitude of the sidebands. However, the lower-frequency sideband is still clearly present, generated by the effects for which $C_2(t)$ and $C_3(t)$ account.

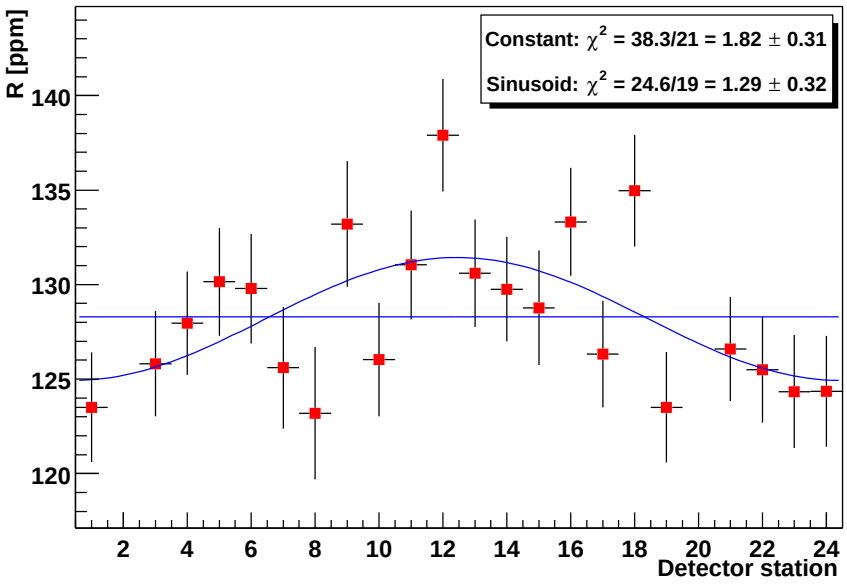



Figure 8.1: Fitted value of $R$ as a function of detector station number with a simplified function where $C_2(t) = C_3(t) = 0$. A conventional threshold fit was used in this case, with a start time of 49.1 $\mu$s.

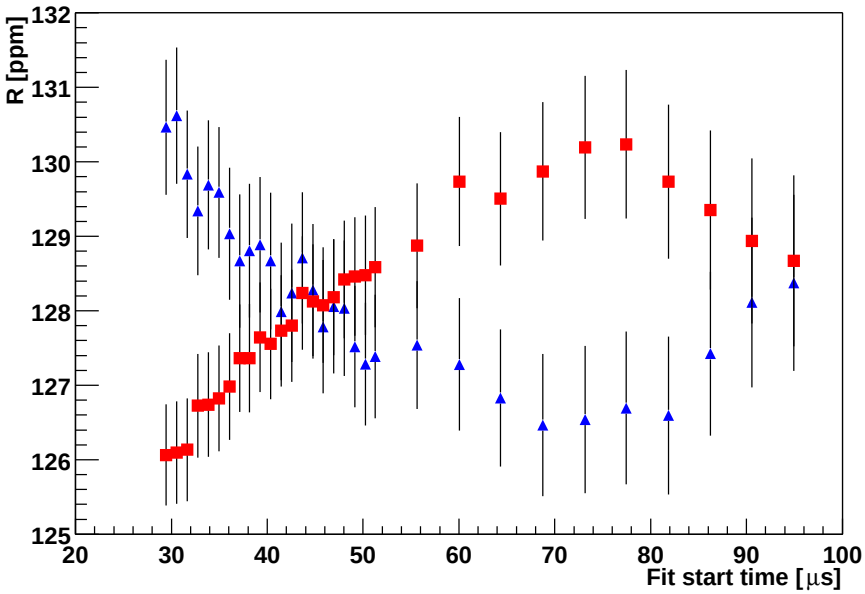



Figure 8.2: Average fitted value of $R$ for stations 1–12 (blue triangles) and 13–24 (red squares) as a function of fit start time with a simplified function where $C_2(t) = C_3(t) = 0$.

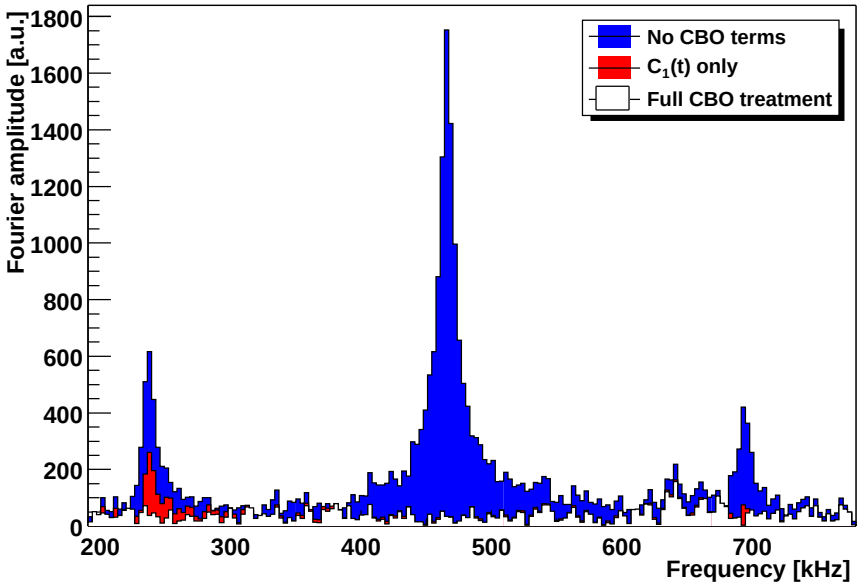



Figure 8.3: Fourier transform of the residuals of fits to the spectra of stations 13-24 over the energy range from 2.0 to 3.2 GeV.

Because $\omega_a/2\pi = 229.1$ kHz overlaps with this lower-frequency sideband, there is a systematic pulling of $R$ by the CBO.

By adding the $C_2(t)$ and $C_3(t)$ terms to the fitting function, the sidebands are removed from the frequency spectrum of the residuals [112, 113]. Even better, $R$ becomes consistent versus station number, and the two halves of the ring behave more consistently, as shown in Figure 8.4. However, a price is paid for the inclusion of these parameters in the fitting function. Each of the parameters related to the "half-ring effect" ($A_{CBO2C}$, $A_{CBO2S}$, $A_{CBO3C}$, and $A_{CBO3S}$) is correlated to $R$ with an off-diagonal term in the fitting covariance matrix of approximately 15 percent. The fitted statistical uncertainty on $R$ increases by 11 percent as a result of these correlations, as though a systematic error of 0.30 ppm had been folded with it in quadrature. Thus, in this analysis, the systematic uncertainty from the half-ring effect is absorbed as an increased statistical error because it is included in the fitting function.

Other analyses of this data set took the alternative approach of neglecting the half-ring effect. They rely on the fact that the systematic bias cancels out when the results from all stations are combined, because all manifestations of the CBO vary in phase through $2\pi$ around the ring. The cancellation is not exact, of course; two detectors are omitted from the average entirely, and the energy distributions of the other detectors are not perfectly matched to each other. Various methods lead to a suppression factor of about an order of magnitude, yielding estimated systematic uncertainties near 0.3 ppm. In the context of the energy-binned analysis described in this thesis, however, it was not possible to obtain as strong a limit on the cancellation factor, partially because a different fit start time was chosen for each bin. The estimated cancellation factor was approximately 3, leading to a 1 ppm systematic uncertainty. Consequently, there is no reasonable choice

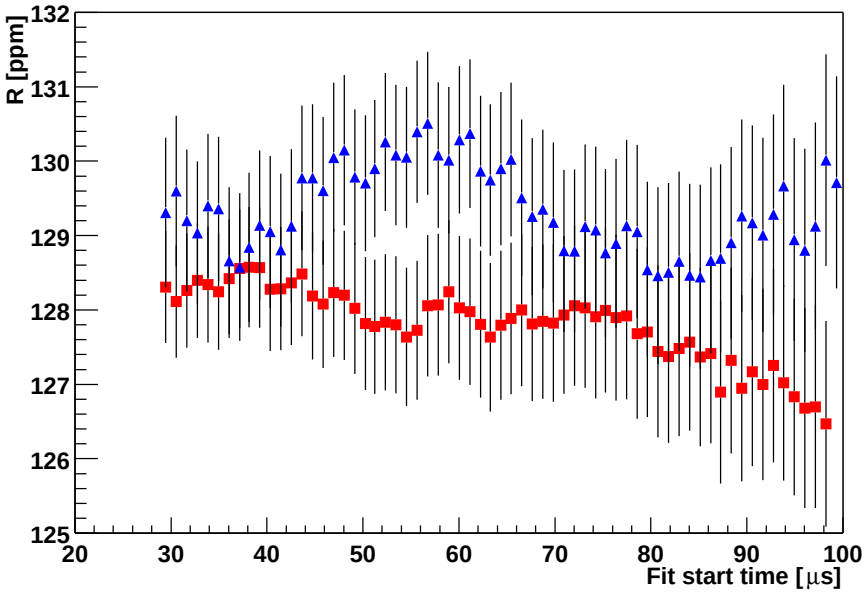



Figure 8.4: Average fitted value of $R$ for stations 1–12 (blue triangles) and 13–24 (red squares) as a function of fit start time for the full function used for the final result.

in this analysis other than to fit for all three of the CBO effects.

## 8.2 Gain stability

Gain changes as a function of time in the detector or the pulse fitter may be measured from the average energy over a defined range versus time after subtracting pileup:

$$N_{PUS}(E,t) = N(E,t) - N_P(E,t) \tag{8.1}$$

$$\bar{E}(t) = \frac{\int_{E_{low}}^{E_{high}} E N_{PUS}(E,t)\,dE}{\int_{E_{low}}^{E_{high}} N_{PUS}(E,t)\,dE}\,. \tag{8.2}$$

The average energy is strongly modulated at the $(g-2)$ frequency $\omega_a$; to see other effects, bins of width $2\pi/\omega_a$ must be used. The lower limit of integration is chosen to be $E_{low} = 2$ GeV; two different upper thresholds $E_{high} = 3.2$ GeV and $E_{high} = 6.2$ GeV are used. The different upper energy cuts provide some measure of the systematic error from incomplete pileup subtraction in $\bar{E}(t)$. The 3.2 GeV threshold excludes the portion of the energy distribution that is due only to pileup, so it is taken as the better estimate of the true gain stability, and the other curve is used as a systematic check.

The relationship between $\bar{E}(t)$ and the gain $G(t)$ is rather subtle. If the energy distribution were uniform

then $\bar{E}(t)$, within a range of observed energies, would not be related at all to $G(t)$; the average energy would always be $\frac{1}{2}(E_{low}+E_{high})$. The sensitivity to gain changes arises primarily from the shifting of events across the boundaries set by the upper and lower thresholds. Defining

$$G(t) = 1 + g(t) \ , \tag{8.3}$$

it may be shown [114] that the corresponding average energy change is

$$\bar{E}(t) = \bar{E_0}[1+\varepsilon g(t)] \ , \text{ where} \tag{8.4}$$

$$\varepsilon = 1 - \frac{E_{high}^2 N_{PUS}(E_{high}) - E_{low}^2 N_{PUS}(E_{low})}{\int_{E_{low}}^{E_{high}} E N_{PUS}(E) dE} - \frac{E_{high} N_{PUS}(E_{high}) - E_{low} N_{PUS}(E_{low})}{\int_{E_{low}}^{E_{high}} N_{PUS}(E) dE} \ . \tag{8.5}$$

For the energy thresholds that were employed, $\varepsilon \approx \frac{1}{2}$, so changes in the average energy must be multiplied by a factor of approximately 2 to determine the corresponding gain changes. In practice, these $\varepsilon$ factors were computed separately for each detector's energy distribution.

The measured gain changes were then parameterized with a function consisting of the sum of two exponentials:

$$G(t) = 1 + A_{G1} e^{-t/\tau_{G1}} + A_{G2} e^{-t/\tau_{G2}} \ . \tag{8.6}$$

These fits are superimposed on the data in Figures 8.6 through 8.9. With few exceptions, the gains are stable at the level of 0.2 percent or better after 50 $\mu$s. Also, there is generally good qualitative agreement between the points determined from the two upper energy cuts after this time. At earlier times, the effects of residual pileup are quite evident.

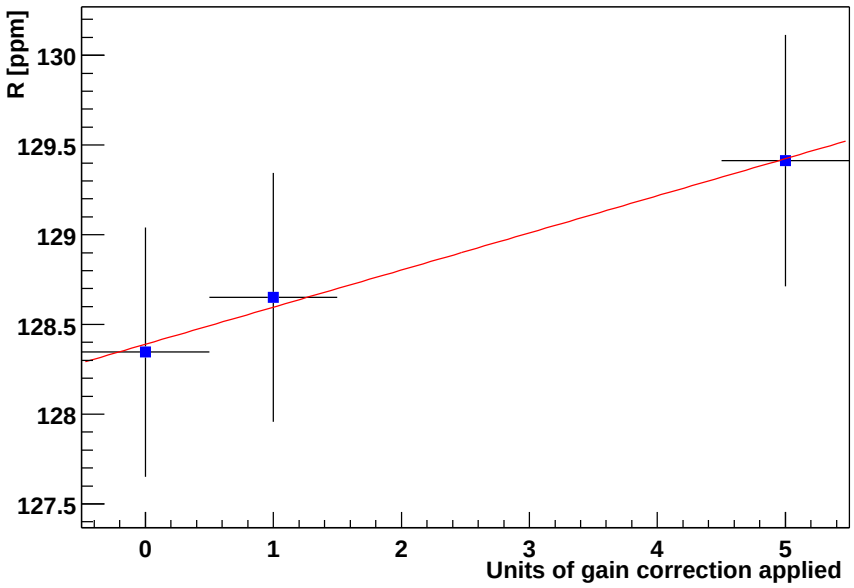
The sum-of-exponentials parameterizations were used to correct the energies of the pulses as they were filled into the histograms that were fit to determine the final result. To determine the systematic error from the time-dependent gain, fits were performed on three slightly different data sets:

- one without gain corrections applied,
- one with gain corrections applied, and
- one with five times the gain corrections applied.

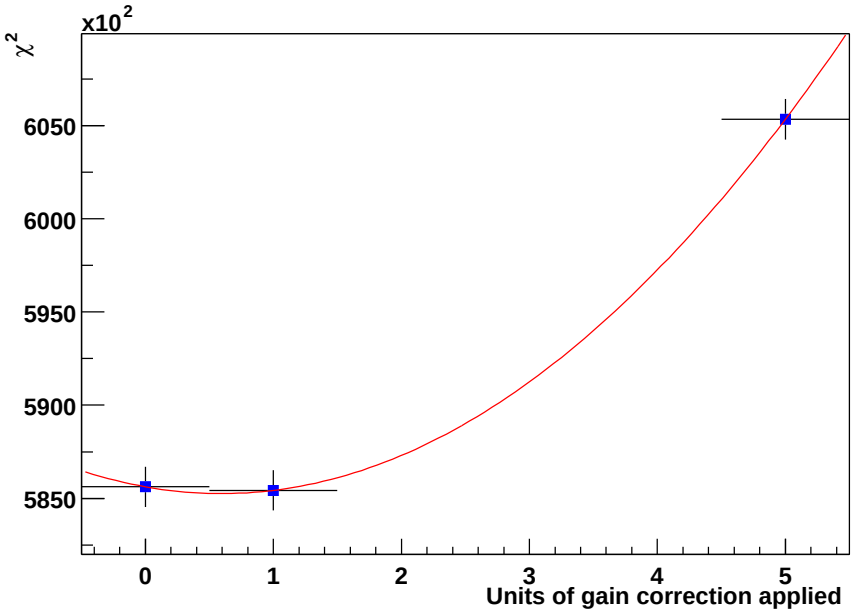
The slope of $R$ relative to the number of units of applied gain change was determined from Figure 8.5(a) to be 0.21 ppm per unit of computed gain change. This result is valid only for the set of start times chosen for the final result.

The next step is to determine a level of confidence in the corrections that were applied. This level of confidence is estimated from two different sources. First, the difference between $G(t)$ with upper energy cuts at 3.2 GeV and 6.2 GeV may be quantified by taking

$$\frac{G_{3.2}(t) - G_{6.2}(t)}{G_{3.2}(t) - 1} \tag{8.7}$$




(a) $R$




(b) $\chi^2$

Figure 8.5: $R$ and $\chi^2$ for fits with different amounts of gain correction applied. The $\chi^2$ is summed over all 198 energy bins, without normalization by the number of degrees of freedom. The slope of $R$ is 0.21 ppm per unit of gain change, and the optimal point in the $\chi^2$ sweep is at 0.59 units of gain correction.

at each point from 50 $\mu$s to 200 $\mu$s. An appropriately weighted average over the time points is taken for each station. After averaging over all stations, the resulting average discrepancy between the two energy cuts is 34 percent. Second, this value is approximately confirmed by the position of the $\chi^2$ minimum in Figure 8.5(b); the optimal point is at 0.59 units of gain correction. This value deviates from 1 by 41 percent, which will be used as the level of confidence. Therefore, the systematic uncertainty assigned for gain changes is $0.41 \cdot 0.21 \text{ ppm} = 0.09 \text{ ppm}$.

## 8.3 Overlapping pulses

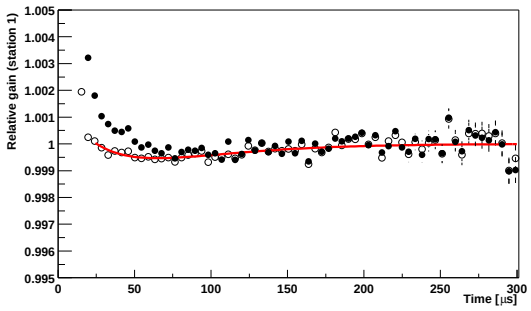
Four separate systematic uncertainty components are assigned to account for pileup. The first is a statistical issue; it derives from the statistical uncertainties in the pileup fits. The second component accounts for a possible global underestimation or overestimation of pileup. The third component relates to a possible energy dependence of the pileup construction efficiency. The final pileup-related systematic uncertainty accounts for the unseen pileup below the pulse fitter's threshold.

Overlapping pulses are accounted for in the fitting procedure by first fitting the pileup spectrum and then inserting a parameterization of the pileup into the ordinary fits. The fits to the pileup spectrum also determine the statistical uncertainties of the pileup parameters $N_P$, $A_{P1}$, $\phi_{P1}$, $A_{P2}$, and $\phi_{P2}$. Under the assumption that the fitted values of the parameters are uncorrelated, the statistical part of the pileup systematic error is evaluated by performing many fits to the ordinary spectrum in which each pileup parameter is modified by a random number. The random perturbations are chosen from a Gaussian distribution whose width is the statistical uncertainty of the parameter. The RMS of the resulting distribution of $R$ values is then computed; it is 0.12 ppm. This value is taken as the statistical part of the pileup error.

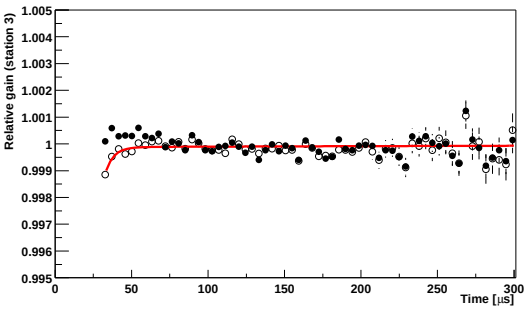
The systematic error from a global pileup construction inefficiency was evaluated by performing fits with the entire pileup parameterization $N_P(E,t)$ multiplied by various scale factors $M_P$ for all stations and energy bins. The slope of a line fit through the points shown in Figure 8.10(a) was determined; it is 4.93 ppm per unit of constructed pileup. This is potentially an extremely dangerous systematic bias, and care must be taken in estimating the level of efficiency of the pileup construction procedure.

This efficiency has been established as 3.4 percent based on three different sources of information, leading to a systematic error estimate for this part of 0.17 ppm. First, the difference between the energy distributions at early and late times gives an estimate of how well the procedure works at very high energies. This method provides useful information above approximately 3.0 GeV; at lower energies, the differences between the spectra are dominated by other effects. The early-time data is taken from 49.11 $\mu$s to 66.57 $\mu$s (4 precession cycles) and the late-time data from 300 $\mu$s to 439.69 $\mu$s (32 precession cycles). The differences for a typical station are shown in Figure 8.11; the two data sets correspond to the situations before and after pileup is subtracted. The residual pileup, defined by the ratio of the integrals of the absolute values of these two curves, appears as a function of station number in Figure 8.12. The average inefficiency determined with this method is 3.4 percent.

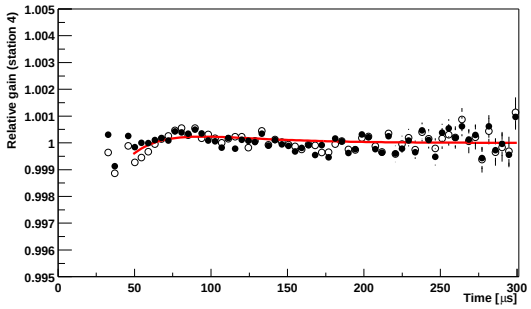
Another piece of evidence for this level of pileup construction efficiency comes from the position of the optimal value of $\chi^2$ as the pileup multiplier $M_P$ is varied, as shown in Figure 8.10(b). This procedure



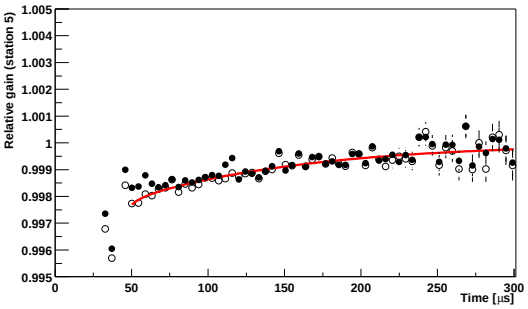

(a) Station 1



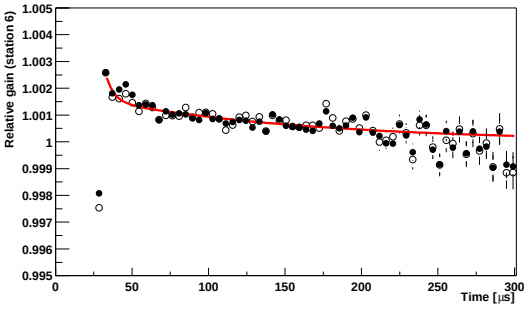

(b) Station 3




(c) Station 4

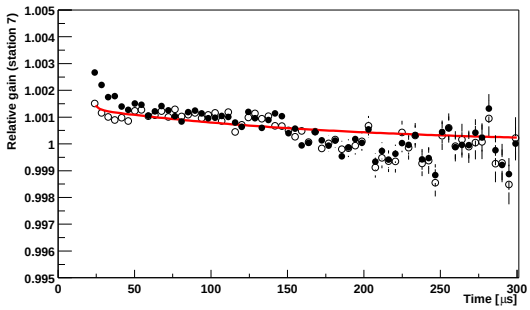



(d) Station 5

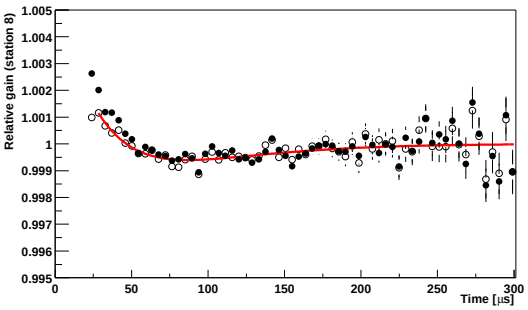



(e) Station 6

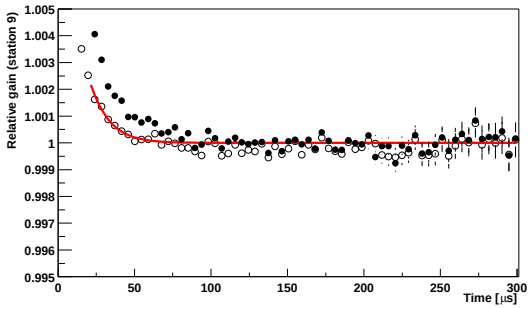
Figure 8.6: Gain changes $G(t)$ after pileup subtraction; the open circles are integrated from 2.0 to 3.2 GeV while the closed circles are integrated from 2.0 to 6.2 GeV. The conversion factor from $\bar{E}$ to gain is already applied. Stations 1–6.



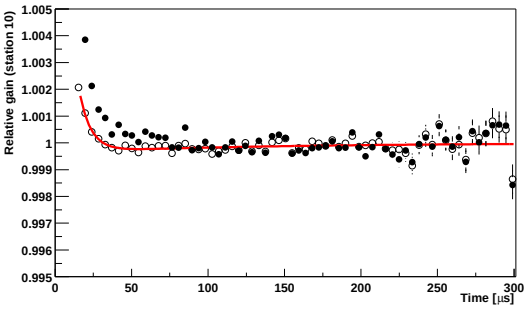

(a) Station 7



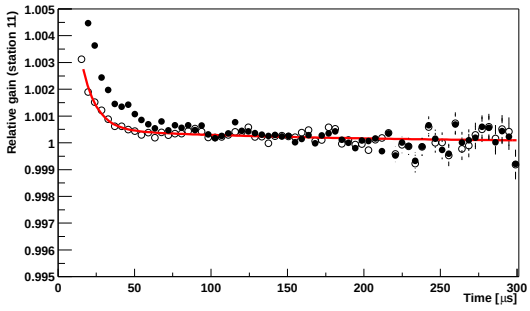

(b) Station 8



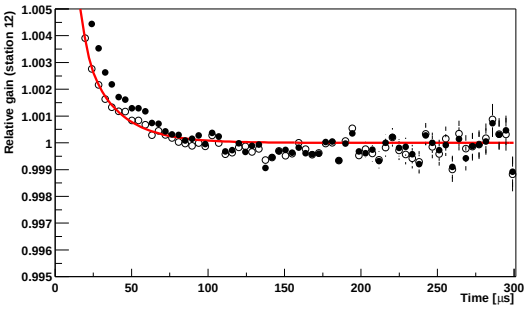

(c) Station 9




(d) Station 10

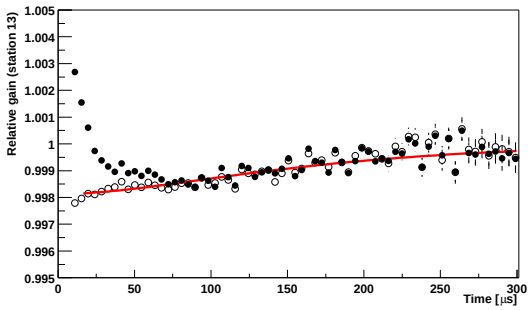



(e) Station 11

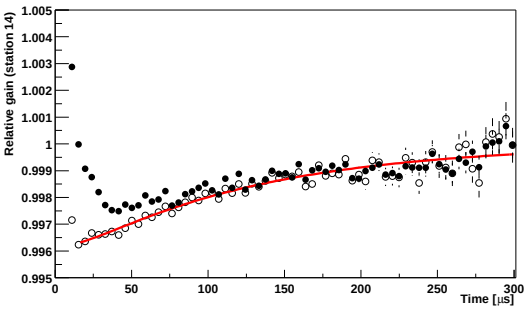



(f) Station 12

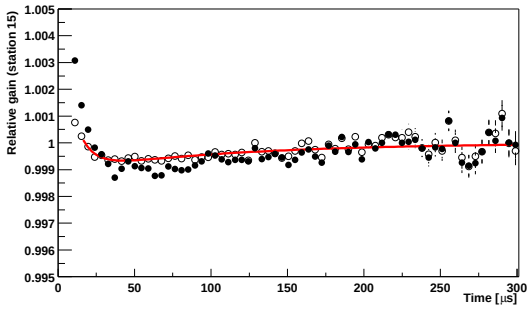
Figure 8.7: Gain changes $G(t)$ after pileup subtraction; the open circles are integrated from 2.0 to 3.2 GeV while the closed circles are integrated from 2.0 to 6.2 GeV. The conversion factor from $\bar{E}$ to gain is already applied. Stations 7–12.



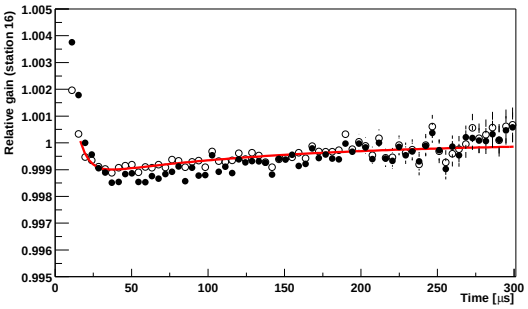

(a) Station 13



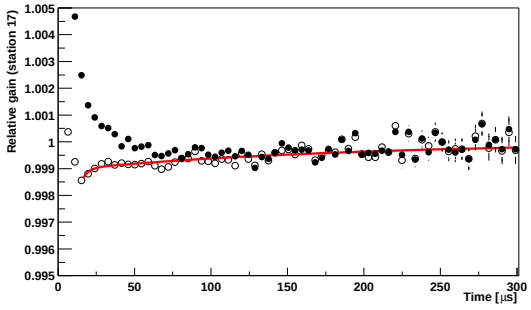

(b) Station 14



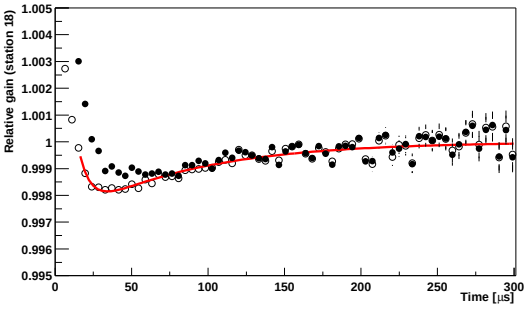

(c) Station 15




(d) Station 16

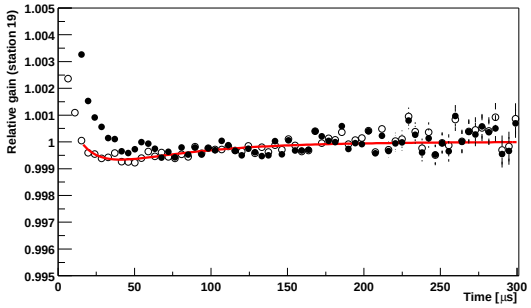



(e) Station 17

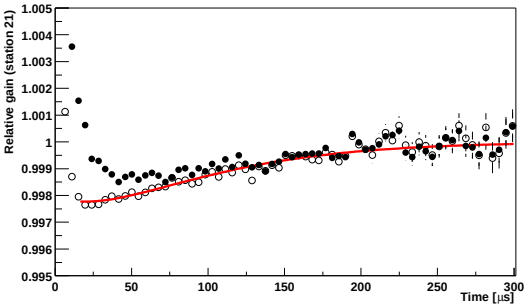



(f) Station 18

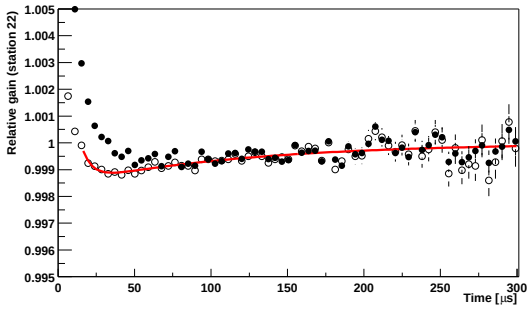
Figure 8.8: Gain changes $G(t)$ after pileup subtraction; the open circles are integrated from 2.0 to 3.2 GeV while the closed circles are integrated from 2.0 to 6.2 GeV. The conversion factor from $\bar{E}$ to gain is already applied. Stations 13–18.



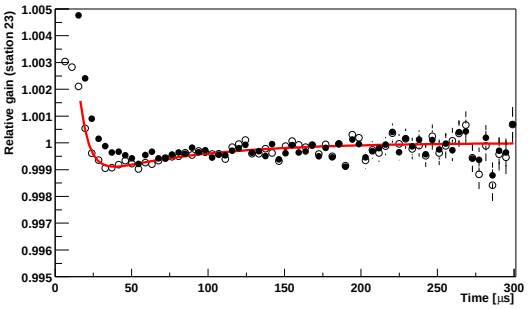

(a) Station 19



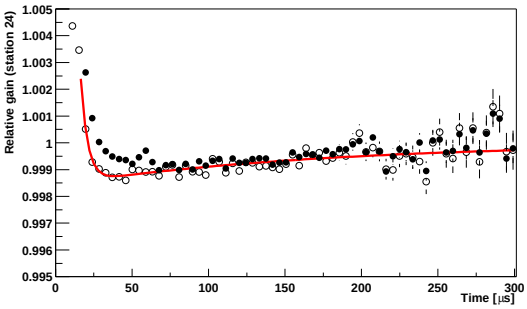

(b) Station 21




(c) Station 22

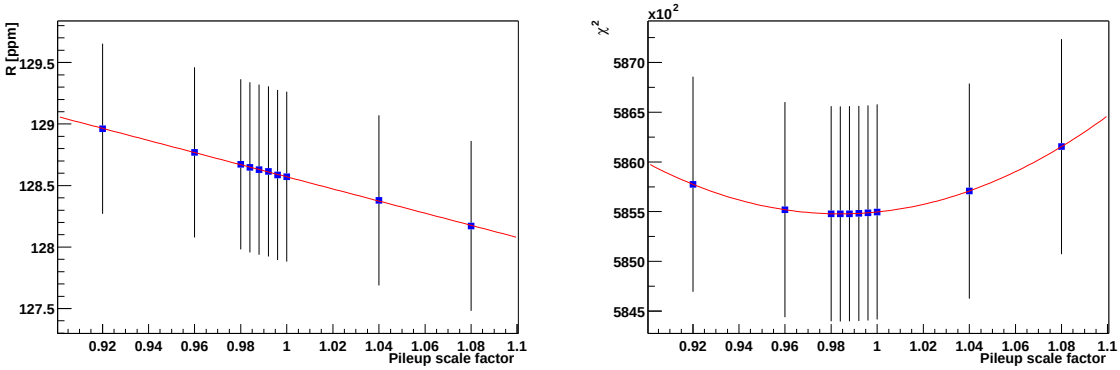



(d) Station 23




(e) Station 24

Figure 8.9: Gain changes $G(t)$ after pileup subtraction; the open circles are integrated from 2.0 to 3.2 GeV while the closed circles are integrated from 2.0 to 6.2 GeV. The conversion factor from $\bar{E}$ to gain is already applied. Stations 19-24.




(a) $R$ (b) $\chi^2$

Figure 8.10: $R$ and $\chi^2$ for fits with varying pileup scale factors applied.

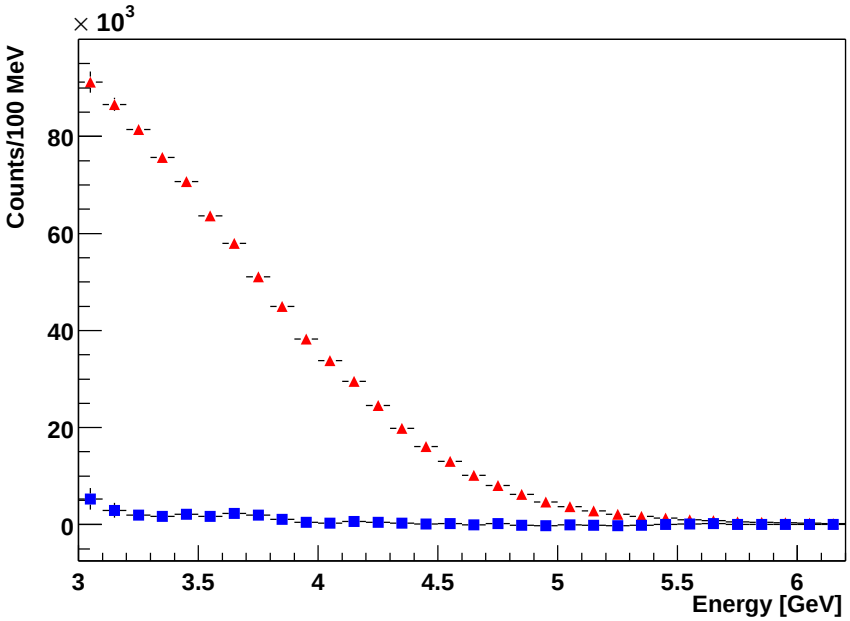



Figure 8.11: Difference between energy spectra from early and late times before (red triangles) and after (blue squares) pileup subtraction. The data are from station 1; only the high-energy tail is shown.

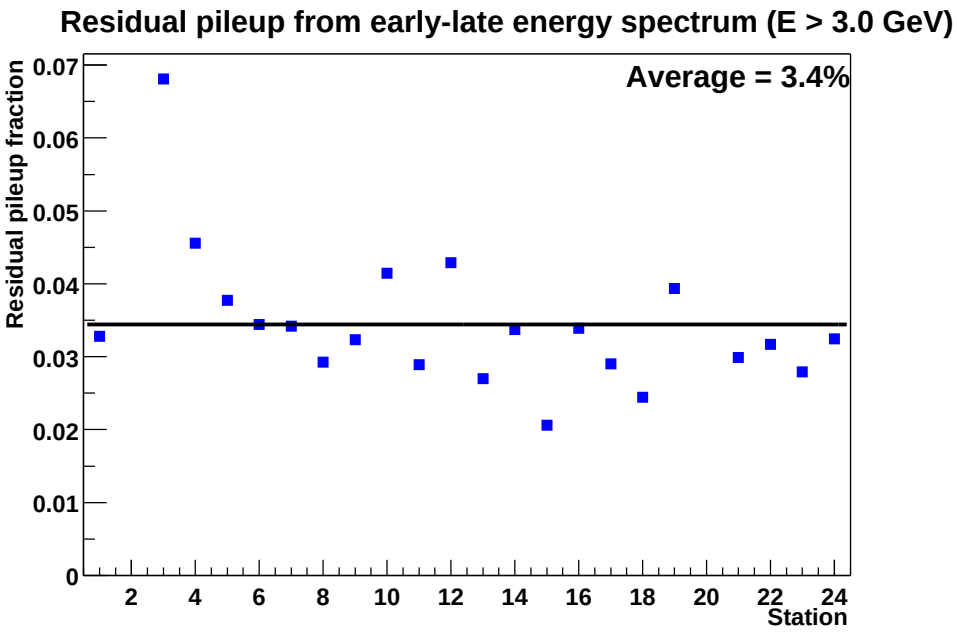



Figure 8.12: Residual pileup fractions for each station for energies above 3 GeV based on difference between the energy spectra at early and late times.

determines an optimal pileup multiplier $M_P = 0.984 \pm 0.004$, which differs from 1 by only 1.6 percent. Finally, the difference between the constructed gain corrections $G(t)$ with upper energy cuts at 3.2 GeV and 6.2 GeV can be used to obtain an additional confirmation. These corrections were recomputed for various values of the artificial pileup multiplier $M_P$. For each value of this parameter, the squared distance between $G_{3.2}(t)$ and $G_{6.2}(t)$ was integrated over the time range from 40 $\mu$s to 100 $\mu$s. These distances exhibit parabolic minima as a function of $M_P$. The positions of these minima are shown for each station in Figure 8.13; the largest discrepancy from 1 is 3.5 percent, and the average is 2 percent. Consequently, a global pileup construction inefficiency of 3.4 percent should be a safe assumption.

The third component of the pileup systematic uncertainty accounts for a possible energy dependence of the pileup subtraction efficiency. This sort of energy dependence would be a possible cause of an incorrectly constructed pileup phase [115]. No particular concrete mechanism has been identified that would cause such an effect, so it is difficult to characterize its scale. However, it seems unlikely that the change in inefficiency over the range from 1.4 GeV to 3.2 GeV would be any greater than twice the global inefficiency estimate of 3.4 percent. Consequently, three functions representing linear changes in the inefficiency over this energy range were constructed; they are drawn in Figure 8.14:

$$\text{(A)}\ M_P(E) = 1 \tag{8.8}$$

$$\text{(B)}\ M_P(E) = 1 + 0.034\,\frac{E - 2.3\ \text{GeV}}{0.9\ \text{GeV}} \tag{8.9}$$

$$\text{(C)}\ M_P(E) = 1 - 0.034\,\frac{E - 2.3\ \text{GeV}}{0.9\ \text{GeV}} \tag{8.10}$$

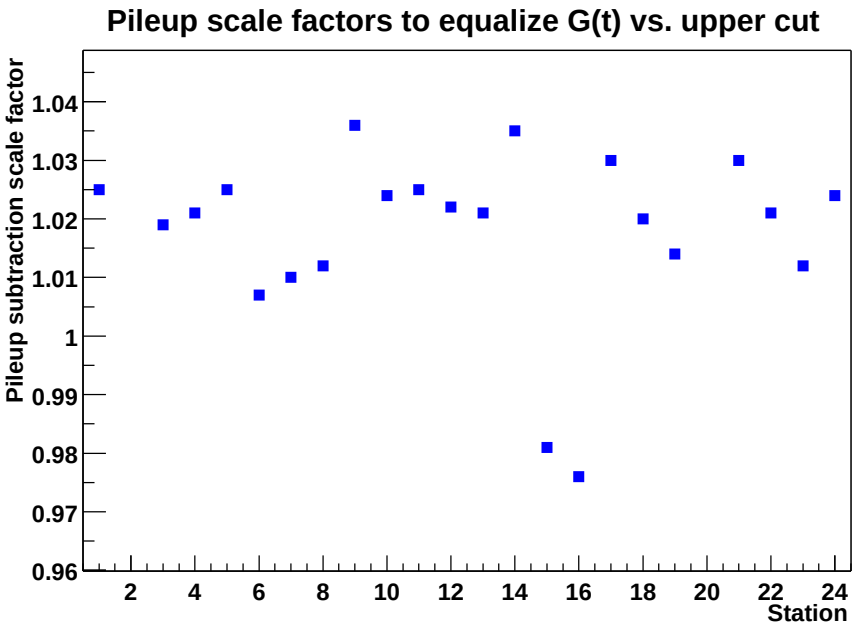



Figure 8.13: Assumed pileup construction inefficiency giving minimum distance between gain corrections $G(t)$ computed with 3.2 GeV and 6.2 GeV upper energy thresholds.

It is found that the differences between the resulting fitted $R$ values is quite small; $R_B - R_A = -0.08$ ppm and $R_C - R_A = 0.07$ ppm. Consequently, a value of 0.08 ppm is assigned for this part of the pileup systematic uncertainty.

The final component relates to the unseen pileup, which consists of pulses below the software threshold of 275 MeV whose only effect is a degraded energy resolution at early times. The difficulty in estimating this uncertainty arises from a lack of precise knowledge of the number of these pulses and their energy, asymmetry, and phase distribution. Consequently, this analysis proceeds in three steps: first, a simulation is used to generate an unseen pileup $(E,t)$ spectrum with an arbitrary normalization. Next, the effect of this distribution on the fitted $(g-2)$ asymmetry is used to set an upper bound on the normalization. Finally, the effect of this unseen pileup on $R$ is determined.

The unseen pileup simulation first chooses an ordinary event from the pileup-subtracted $(E,t)$ distribution for a typical station, weighted by an additional factor of

$$N_{unseen}(t) = e^{-t/\tau}[1 + A_{unseen}\cos(\omega_a t + \phi_{unseen})] \; . \tag{8.11}$$

The parameters $A_{unseen}$ and $\phi_{unseen}$ are set very conservatively to $A_{unseen} = 0.08$ and $\phi_{unseen} = 100$ mrad. The simulation then chooses the energy of the unseen pulse from an estimate of the energy distribution below 200 MeV. A simulated WFD digitization interval containing the two pulses is created and processed with the pulse fitting algorithm to determine the effect of the unseen pulse. The resulting energy and time distributions are shown in Figure 8.15.

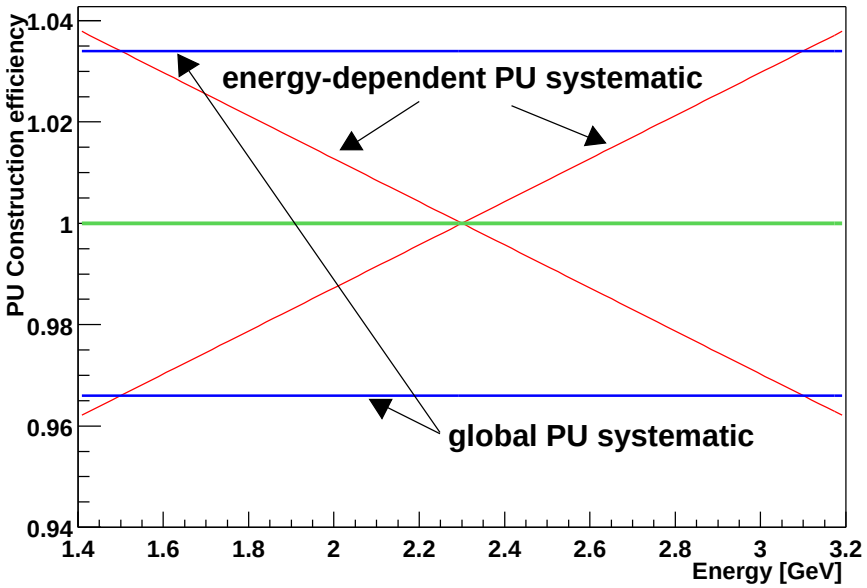
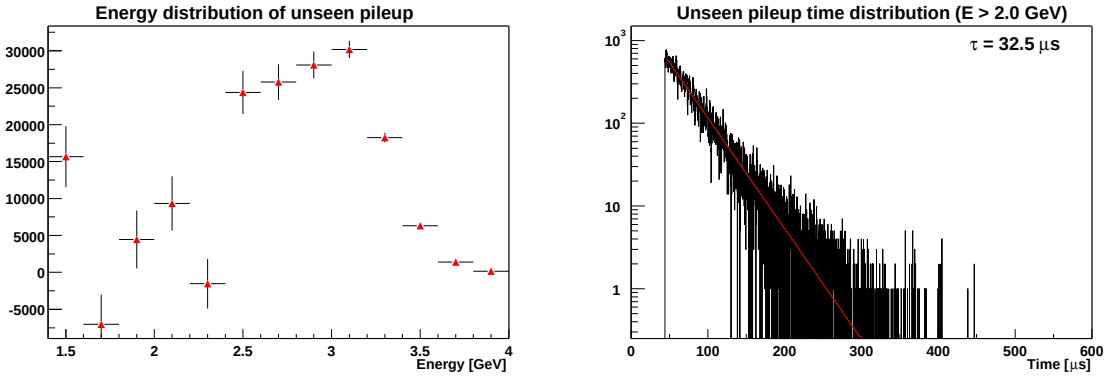



Figure 8.14: Functions chosen to estimate systematic uncertainty from a hypothetical energy dependence in the efficiency of the pileup construction procedure.




(a) Energy distribution

(b) Time distribution

Figure 8.15: Energy and time distributions of simulated unseen pileup. The normalization of these plots corresponds to 10 “units” of unseen pileup per station.

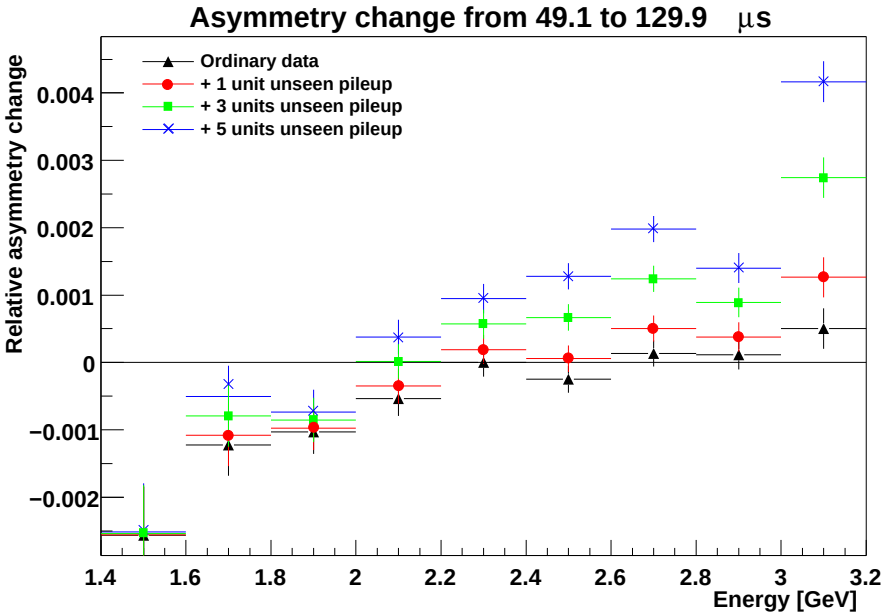
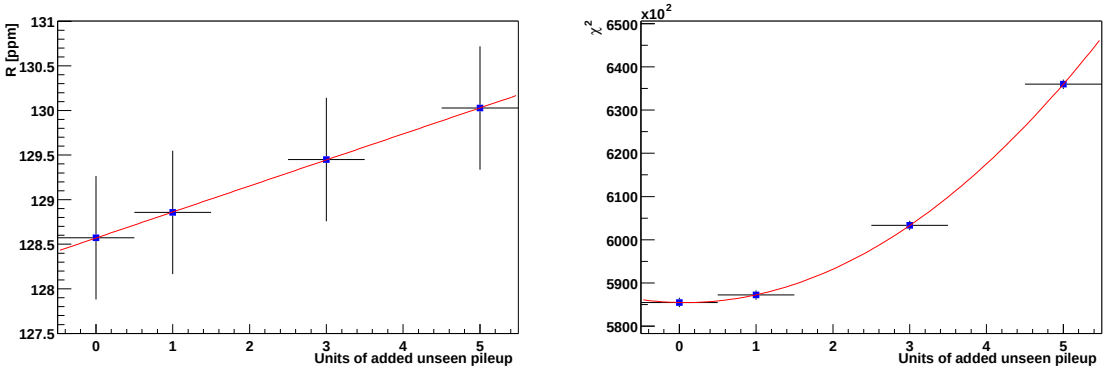



Figure 8.16: Change in the fitted asymmetry between fit start times of 49.1 $\mu$s to 129.9 $\mu$s, for the unperturbed data set and with various amounts of added unseen pileup.

The second step in this analysis is to observe the effect of the unseen pileup on the stability of the $(g-2)$ asymmetry. The simulated unseen pileup distribution is added to each station's histogram, multiplied by factors of 0, 1, 3, and 5. Fits are then performed at start times of 49.1 $\mu$s and 129.9 $\mu$s. The difference between the asymmetries at these two points, averaged over stations, is computed as a function of energy for each unseen pileup multiplier. The asymmetry for each station is rescaled to be 1 at 49.1 $\mu$s before averaging. These asymmetry shifts appear in Figure 8.16. The unperturbed spectrum has an average asymmetry that is consistent from early to late times in each energy band above 2.2 GeV. For energies less than 2.2 GeV, the asymmetry in the unperturbed spectrum decreases as a function of fit start time. One mechanism that might account for this decrease is the slow loss of protons from the storage ring. Their energy distribution cuts off near 1.8 GeV, their $(g-2)$ asymmetry is zero, and they become an increasingly larger fraction of the data at later times. Finally, it is clear that the asymmetry consistency statistic is much more sensitive to unseen pileup at very high energies than at lower energies. Consequently, the best limit on the effects of unseen pileup can be obtained by looking at these energies.

The amount of unseen pileup in the data was estimated from this data by computing in each bin $\Delta A/\Delta A_{unseen}$, which is the observed change in the asymmetry divided by the change that would be caused by one unit of unseen pileup. The weighted average over the five highest energy bins of this quantity, which represents the number of units of unseen pileup in the data, is $0.29 \pm 0.26$ units. Each unit corresponds to $1.5 \times 10^8$ generated events with times after 45 $\mu$s.

The fitted $R$ value is also plotted as a function of the number of units of added unseen pileup, as illustrated in Figure 8.17. The resulting slope is 0.293 ppm per unit of unseen pileup. This slope is multiplied by 0.29




(a) $R$ (b) $\chi^2$

Figure 8.17: $R$ and $\chi^2$ for fits with varying amounts of added unseen pileup.

to obtain the systematic error estimate of 0.09 ppm.

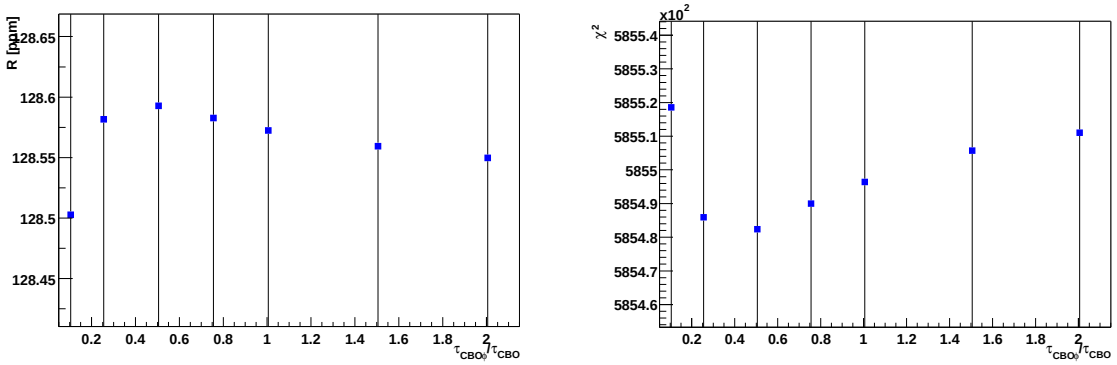
The four components of the systematic uncertainty due to pileup are, finally, combined in quadrature to give a total value of 0.24 ppm.

## 8.4 Muon losses

The systematic uncertainty resulting from muon losses is divided into three parts. First, the normalization factor $N_\Lambda$ is only known to limited precision. Second, contamination of the spectrum by lost protons can change the shape of the losses as measured in the FSDs. Both of these uncertainties mean that the best possible fit is not achieved, though the effect on $\omega_a$ is small. On the other hand, if some loss mechanism is correlated to the muon spin direction, the average $(g-2)$ phase is directly pulled, and $\omega_a$ is affected.

The usual technique was applied to evaluate the sensitivity of $R$ and $\chi^2$ to changes in the fixed parameters $N_\Lambda$ and $N_{\delta\Lambda}$. Unfortunately, there is little external evidence to be used to establish a level of confidence in these loss parameters. The value of $N_\Lambda$ that is used for the final result was determined by minimizing the global $\chi^2$ for the set of fit start times used for that result. However, varying the fit start times to other reasonable values yields optimal points of $N_\Lambda$ that vary from 0.885 to 1.002. The full width of this range is 11.7 percent; it will be used as a figure of merit. As $N_\Lambda$ varies over this range, $R$ moves by 0.05 ppm, so this value is assigned as the systematic uncertainty. On the other hand, the shape parameter $N_{\delta\Lambda}$ ought to be 0, since all relevant corrections to the muon loss shape have already been applied. The value that produces an optimal global $\chi^2$ is -4.6. This entire difference from 0 is used as a level of confidence, giving a systematic uncertainty from $N_{\delta\Lambda}$ of 0.01 ppm. These two values are added linearly to give a total muon loss systematic of 0.06 ppm, except for the part due to spin-dependent losses.

One possible way to generate a spin-dependent loss mechanism would be to couple the average spin direction at injection to the muon momentum. Because muons with different momenta orbit at different radii, the loss probability may reasonably be a function of momentum. This coupling could be produced if




(a) $R$ scan for $C_3(t)$ envelope

(b) $\chi^2$ scan for $C_3(t)$ envelope

Figure 8.18: $R$ and $\chi^2$ for fits with varying CBO oscillation envelope lifetimes.

the beamline exhibits a left-right acceptance asymmetry at the point where pions decay to yield muons. This effect was evaluated and found to be less than 0.10 ppm [116]. Folding this value in quadrature with the 0.06 ppm uncertainty found previously gives a grand total muon loss systematic uncertainty of 0.12 ppm.

## 8.5 CBO beyond half-ring effects

The CBO modulations have an envelope function that is taken to be $e^{-t/\tau_{CBO}}E_{CBO}(t)$, where $E_{CBO}(t)$ measures the deviation from a purely exponential form. Fortunately, the precise envelope function is only necessary to obtain a good $\chi^2$/dof; it does not have a large effect on the fitted value of $R$. In fact, the difference between the values extracted with the usual $E_{CBO}(t)$ and with $E_{CBO}(t) = 1$ is entirely negligible, only 0.001 ppm. Meanwhile, the fitted exponential CBO lifetime varies with energy at a level that is too large to understand fully; as shown in Figure 7.3(b), it moves by $\pm 15$ percent from very low to very high energies; this value will be taken as the uncertainty on $\tau_{CBO}$. Following the usual technique, the associated error for $R$ is 0.02 ppm.

The envelope of the CBO modulation $C_3(t)$ of the $(g-2)$ phase is derived from the same underlying effects as that of the other CBO effects, but arguments exist that suggest that its time dependence may not precisely match the others. However, it is difficult to imagine that its characteristic lifetime would differ from the others' by more than a factor of 2. This lifetime cannot be determined from the data with any precision, so it is necessary to attribute a systematic uncertainty to this envelope. Fits were performed with various values for the ratio of this lifetime to the ordinary CBO lifetime. The $\chi^2$ minimum, as shown in Figure 8.18(b), is very shallow, but fortunately $R$ does not depend heavily on this parameter. The systematic uncertainty is determined by taking half the difference between the maximum and minimum $R$ values in this graph, giving a result of 0.02 ppm.

The CBO frequency is assumed in the fitting function to be time-independent, although it is known

that it changes slightly over the fill time. These changes occur at early times as the $n$ value varies during the transition from scraping to non-scraping quadrupole voltages, and at late times because the quadrupole voltage slowly drops as charge leaks off the electrodes. The entire difference between the ordinary result and the result with a fitting function that includes reasonable estimates of these effects is 0.03 ppm; this difference will be assigned as a systematic uncertainty. Meanwhile, the fitted CBO frequency $\omega_{CBO}$ varies unexpectedly with energy, much like the lifetime. It falls in the $\pm 0.16$ kHz range from 465.79 kHz to 466.11 kHz, as shown in Figure 7.3(a). This uncertainty for $\omega_{CBO}$ is propagated into an uncertainty of 0.01 ppm for $R$. It is then combined linearly with the value from the time dependence to give a combined uncertainty of 0.04 ppm for the CBO frequency.

Putting together all of these results, the systematic uncertainty caused by the CBO, apart from the half-ring effect, is 0.05 ppm.

## 8.6 AGS background

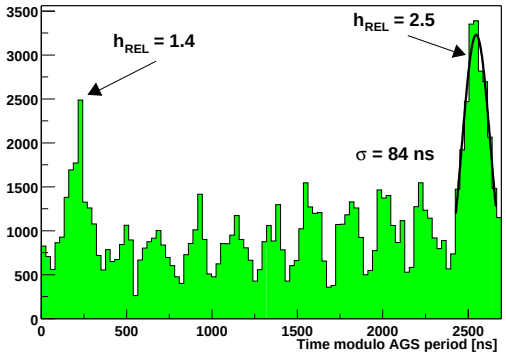
In order to measure the background level in the storage ring, the firing of the quadrupoles is inhibited for a fraction of the fills, typically one out of 25. With no focusing, the muon beam is lost after a few turns, but all other systems operate normally. The distribution of particle arrival times in these fills appears modulo the AGS cyclotron period of 2694 ns in Figure 8.19(a). Eleven peaks are visible, corresponding to eleven of the twelve bunches circulating in the AGS; the first bunch does not produce this background. The other parts of Figure 8.19 illustrate the absolute time scale of these counts as well as their energy distribution in a number of detector stations. Different detectors observe different energy distributions of background counts because the storage ring acts as a spectrometer, guiding different momenta to different positions. These distributions have been appropriately rescaled by the ratio of the number of quadrupole-on fills to the number of quadrupole-suppressed fills.

Simulated background distributions were constructed based on this data. They were added to the usual spectra at 0, 5, and 10 times their actual level, and the slope of $R$ versus the background level was determined. This slope is 0.019 ppm per unit of background. Because no correction was attempted, one unit of background is assumed, and the systematic uncertainty is 0.02 ppm.

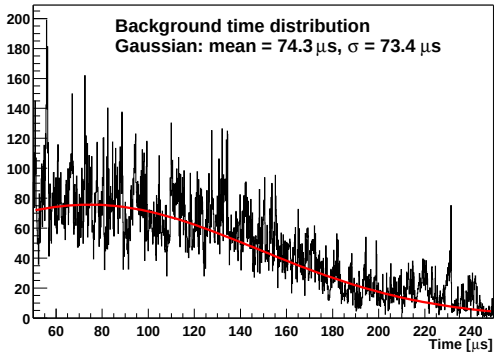
## 8.7 Cyclotron motion

Five random seeds were used for the removal of the cyclotron frequency modulation in the final result. The RMS of the resulting distribution of $R$ values, which are shown in Figure 8.20(a), is 0.1 ppm. The systematic uncertainty that is assigned to account for the randomization process is therefore $0.1 \text{ ppm}/\sqrt{5} = 0.04$ ppm.

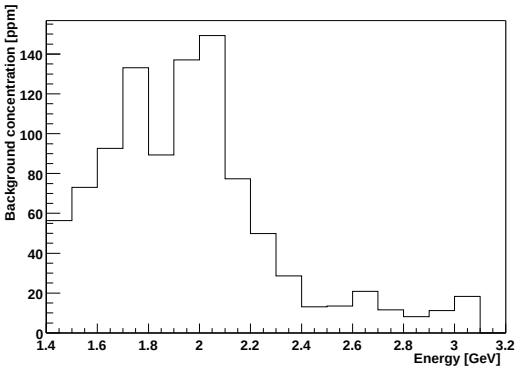
The cyclotron randomization period was varied from 149.0 ns to 149.4 ns in steps of 0.1 ns; the nominal period is 149.2 ns. For this scan, the histogram time bin width remained fixed at 150 ns. The RMS of the resulting distribution of $R$ values, shown in Figure 8.20(b), is 0.085 ppm. This is clearly an upper limit on the effect, since there are statistical variations among the various data sets as well, and no trend is apparent. Consequently, this sweep is regarded as a cross-check rather than a systematic error measurement.



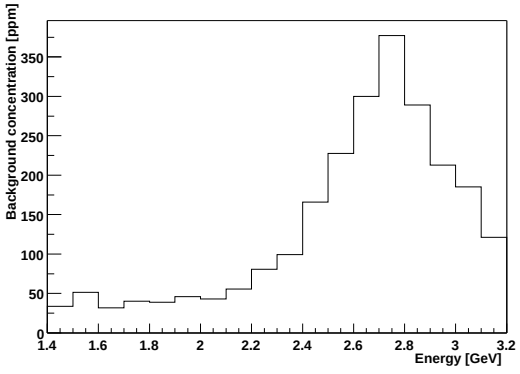

(a) Time distribution modulo the AGS cyclotron period



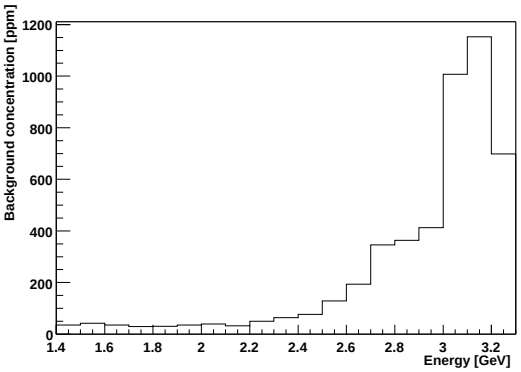

(b) Long-scale time distribution



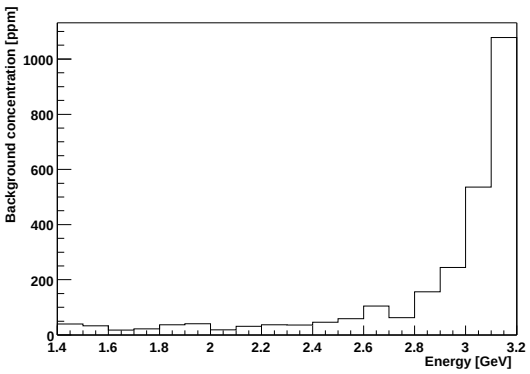

(c) Energy distribution in station 3




(d) Energy distribution in station 4

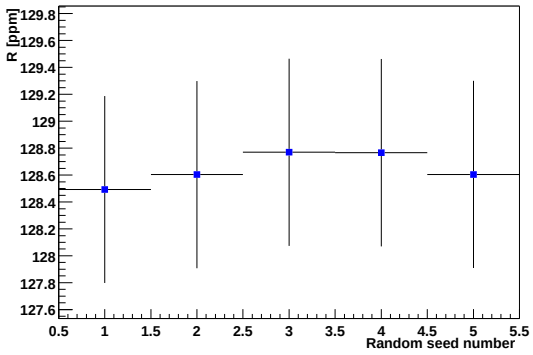



(e) Energy distribution in station 5

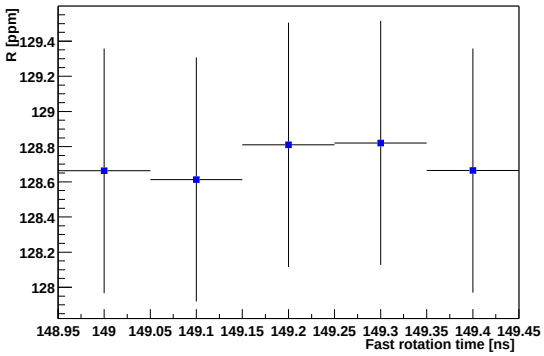



(f) Energy distribution in station 6

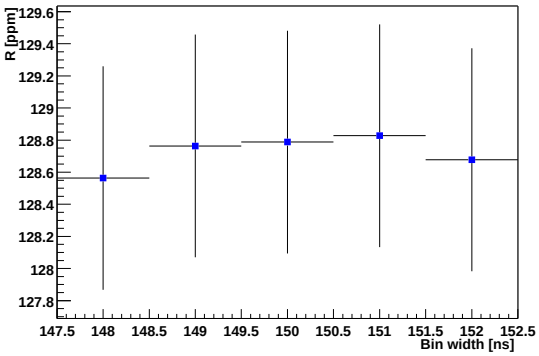
Figure 8.19: Background distributions determined from the quadrupole-inhibited fills.




(a) Random seed




(b) Fast rotation time




(c) Bin width

Figure 8.20: Values of $R$ resulting from scans over parameters related to fast rotation and time binning.

Finally, the histogram bin width was varied from 148 ns to 152 ns in steps of 1 ns. The RMS of the distribution of $R$ values, shown in Figure 8.20(c), is 0.094 ppm; it is assigned as the systematic uncertainty from the bin width. Again, this is certainly a conservative upper bound because of allowed statistical variations. However, in this case a trend in the points is possible, so it is included in the final systematic error estimate.

## 8.8 Vertical oscillations and double-frequency CBO

Particles in the storage ring perform vertical betatron motion at a frequency $f_y = 2.36$ MHz. Beating with the fast rotation frequency leads to a coherent mode–the "vertical waist"– with a frequency of $f_{VW} = f_c - 2f_y = 1.75$ MHz. Each of these effects can couple to the detector acceptance and therefore modulate the counting rate. These frequencies are well separated from $\omega_a$, and the associated time constants are very short. The

lifetime of the vertical betatron oscillations is approximately 3 $\mu$s, while that of the vertical waist is about 15 $\mu$s. Consequently, the effect on $R$ is very small. Also noticeable is the second harmonic of the radial CBO at a frequency of 932 kHz. The time constant for this effect is significantly longer, about 50 $\mu$s, but the amplitude is very small.

The frequencies and time constants were estimated from overall fits at extremely early start times. The amplitudes and phases were then determined for each energy bin for each station from fits starting at 31.6 $\mu$s. Stations 3 through 6 are still gated off at 31.6 $\mu$s; they were assigned vertical oscillation and vertical waist amplitudes and phases from the average of the other stations. For the vertical waist, the phase was adjusted properly for the ring position. The double-frequency CBO is sufficiently long-lived that its amplitude and phase can be reliably determined with a fit starting at 49.1 $\mu$s, so it is not necessary to interpolate values for stations 3-6 for this quantity.

For each of these modulations, fits were performed with the relevant amplitudes fixed at 0, 1, 5, and 10 times the values found from the very early-time fits. The slope of $R$ versus the amplitude multiplier was determined in each case. The effect of the primary vertical betatron oscillations at these start times is too small even to write down; it is less than 0.0003 ppm in all cases. The systematic shifts from the vertical waist and double CBO are larger, though still very small: 0.003 ppm and and 0.008 ppm, respectively. Consequently, a systematic uncertainty of 0.01 ppm is included to account for all of these effects together.

## 8.9 Summary

The systematic effects discussed in this chapter are summarized in Table 8.1. Combining them in quadrature, and applying the electric field and pitch corrections discussed in Section 5.5, the final value of $R$ from this analysis is

$$R = 127.95 \pm 0.69\ (\text{stat.}) \pm 0.31\ (\text{syst.})\ \text{ppm}. \tag{8.12}$$

Removing the previously secret offset, converting to $\omega_a$, and combining the statistical and systematic uncertainties in quadrature yields

$$\omega_a = 229\,074.06(0.17)\ \text{Hz}. \tag{8.13}$$

| Effect | Uncertainty [ppm] |
|---|---|
| Statistics | 0.69 |
| *Including* Half-ring effect | 0.30 |
| Overlapping pulses (pileup) | 0.24 |
| Lost muons | 0.12 |
| Gain changes | 0.09 |
| Binning | 0.09 |
| Ordinary CBO | 0.05 |
| Cyclotron randomization | 0.04 |
| Electric field and pitch corrections | 0.03 |
| AGS background | 0.01 |
| Vertical CBO/Double-frequency CBO | 0.01 |
| Total systematics | 0.31 |
| Total uncertainty | 0.76 |

Table 8.1: Summary of statistical and systematic uncertainties for the energy-binned $\omega_a$ analysis.

# Chapter 9

# Determination of $\omega_p$

This chapter describes the measurement of the average magnetic field in which the muon spin precesses. The field is mapped periodically with a trolley carrying NMR probes; between trolley runs, it is tracked by fixed NMR probes. The crux of the analysis of the $\omega_p$ data is the calibration of the various probes with respect to each other and also with respect to an absolute standard, the determination of appropriately weighted averages over temporal and spatial distributions, and the evaluation of the systematic uncertainties in the measurement.[1]

## 9.1 Trolley measurements

The field was mapped every few days during the data-taking period with a trolley containing 17 NMR probes, together with the associated readout electronics, sealed inside a pressurized enclosure. The arrangement of the probes is illustrated in Figure 9.1. With this geometry, field multipoles through the decupole can be determined. While $\omega_a$ data are obtained, the trolley is parked in a "garage" located outside the storage region but still within the vacuum system. When an appropriate moment for a field measurement arrives, often in conjunction with AGS downtime, the trolley is pulled out of the garage and driven around the ring on rails that are built into the vacuum chambers. Two cables are attached to opposite ends of the trolley so that it can be pulled in either direction. One cable is a coaxial conductor that provides power and reads out the data from the probes; the other is made of plastic so that it may be left in the vicinity of the high-voltage pulsed kicker magnet. Figure 9.2 shows the azimuthal dependent of the field around the ring as measured by the trolley; variations over a range of $\pm 100$ ppm are observed. The azimuthal position of the trolley is determined by its effect on the field measured at nearby fixed probes together with the readout of a potentiometer mechanically coupled to the drum that pulls the cable. The uncertainty of the position of each measurement leads to a 0.10 ppm systematic uncertainty in the determination of the azimuthally averaged field.

At the heart of each of the probes is a cylindrical volume 2.5 mm in diameter and 15 mm long; it is filled with water in which a small amount of copper sulfate is dissolved. This active volume is surrounded by a copper coil that is used for both excitation and pickup; it is coupled to the magnetometer electronics [118]

[1] The measurement of the magnetic field is described in [117].

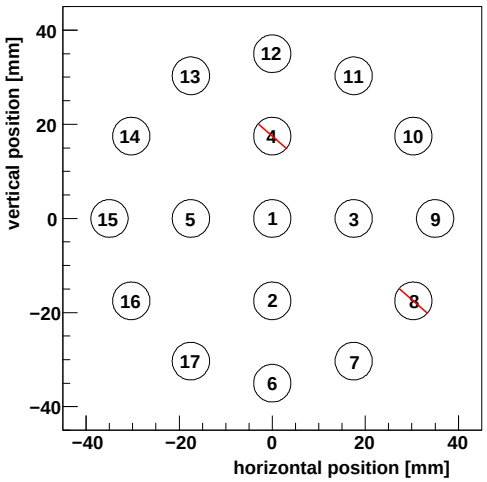



Figure 9.1: Trolley probe geometry. The two probes indicated with red slashes were excluded from the analysis because of their poor signal quality.

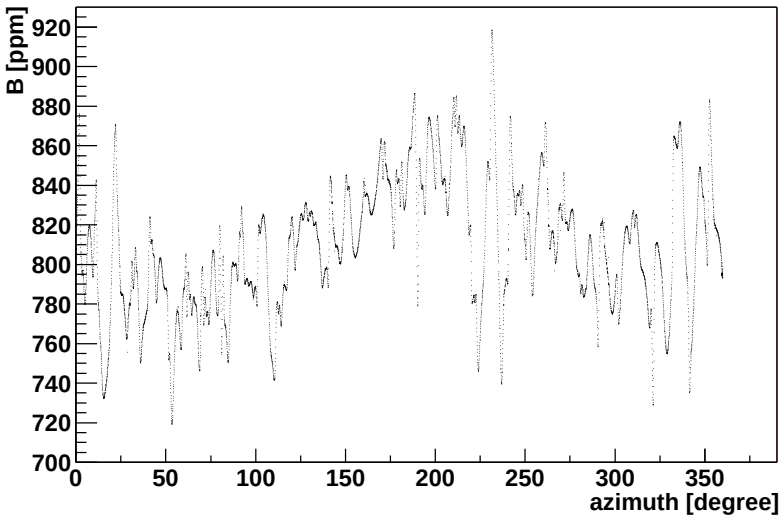



Figure 9.2: Magnetic field variation in azimuth, from the central trolley probe in a typical trolley run. (Reproduced from [117].)

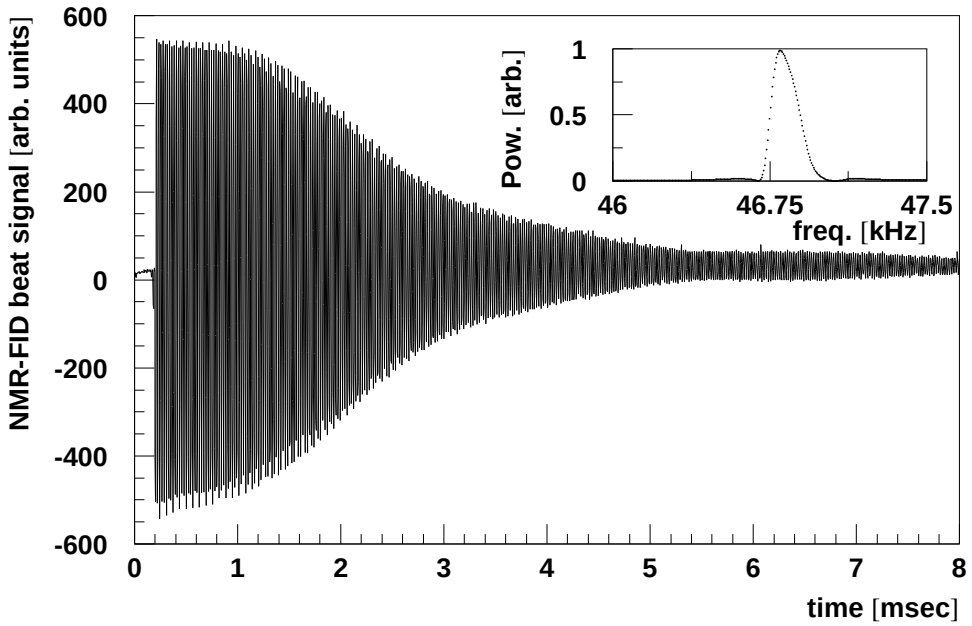



Figure 9.3: Free induction decay signal at a frequency $\omega_p - \omega_{ref}$. The inset shows the Fourier transform of the signal. (Reproduced from [118].)

through a resonant circuit with a quality factor $Q \approx 30$. For each measurement, a pulse of width 4 $\mu$s at the reference frequency $\omega_{ref} = 61.70$ MHz is amplified to a power of 10 W and transmitted to the coil. This pulse leads to an oscillating $B$ field with an amplitude of about 2 mT inside the active water sample, turning its magnetization by 90 degrees away from the direction of the main magnetic field. Following this pulse, the magnetization precesses about the direction of the main field at the angular frequency $\omega_p = \mu_p B$, gradually relaxing back to its ground state parallel to the field. The copper sulfate dissolved in the water accelerates the relaxation process, leading to characteristic times of order 10 ms. The signal picked up by the coil is amplified and mixed with the original reference frequency, leading to a beat at the frequency $\omega_{FID} = \omega_p - \omega_{ref}$, which is known as the free induction decay signal. This signal is illustrated in Figure 9.3. The NMR electronics measures the frequency $\omega_{FID}$ by counting the number of zero crossings during a fixed gate time. The same electronics are coupled to all of the probes in the trolley through a multiplexer switch, and the measuring system alternates among the probes.

The result of a typical trolley run is illustrated in Figure 9.4, where the azimuthal coordinate has been averaged away and a multipole expansion has been performed. Clearly, the field is extremely uniform; over the entire storage region, deviations are limited to less than 2.5 ppm.

Each of the trolley probes gives only a relative measurement of the field. The probes are cross-calibrated among themselves from time to time through measurements with an *x-y* plunging probe. However, the absolute scale must be established by calibrating them against a standard probe [83], a step that was performed at the beginning and repeated at the end of the data collection period. The standard probe is designed so that all perturbations to the applied field by the probe itself are small and easily determined and subtracted. The

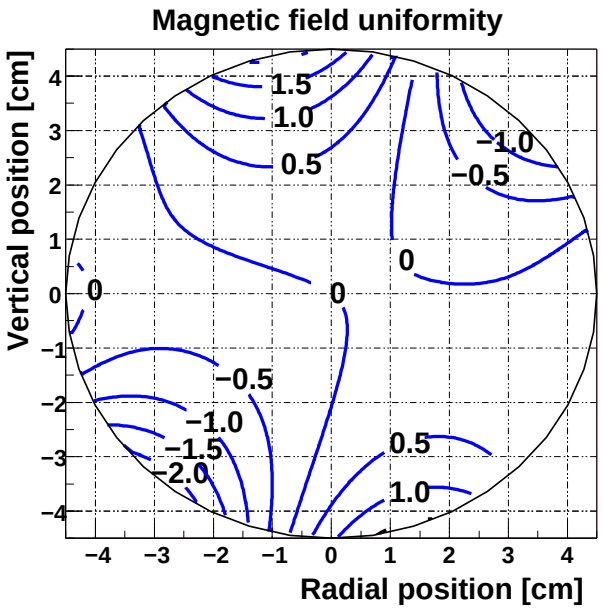



Figure 9.4: Contour plot of the magnetic field for a typical trolley run, averaged over the azimuthal coordinate. (Reproduced from [3].)

active volume consists of a spherical sample of pure water enclosed in a Pyrex ball with a diameter of 1 cm; for a spherical shape, there is no net bulk diamagnetism. Corrections are applied to account for the diamagnetic shielding of the protons by the water molecules themselves and for the susceptibilities of the materials in the probe. The absolute calibration of the standard probe is valid at the level of 0.05 ppm. The transfer of this calibration to the trolley probes is limited primarily by the ability to position them at precisely the same location as the standard probe so that they experience the same field. In order to reduce this error, the field gradient was minimized in the calibration region using current shimming coils. The remaining uncertainty was carefully evaluated by deliberately introducing perturbing fields and observing their effect on the values measured by the trolley and the standard probe. It was thereby determined that the 1 mm precision of the azimuthal positioning translates into a 0.15 ppm uncertainty in the absolute calibration.

## 9.2 Fixed probe measurements

While the $\omega_a$ data are obtained, the trolley is sequestered in its garage. During this time, drifts in the overall scale of the field are monitored by fixed probes that are secured outside the vacuum chambers. These probes are of the same design as the ones carried on the NMR trolley, and they are read out by similar electronics. The active volume in many of the fixed probes is filled with petroleum jelly rather than water in order to avoid evaporation, since many of these probes are in locations that are impossible to access while the vacuum chamber is in place. There are 360 of these probes around the ring, but many of them have been damaged, and some are in locations with large local field perturbations that render them unusable. Consequently, a subset of the probes must be selected to include in an average. Two independent analyses

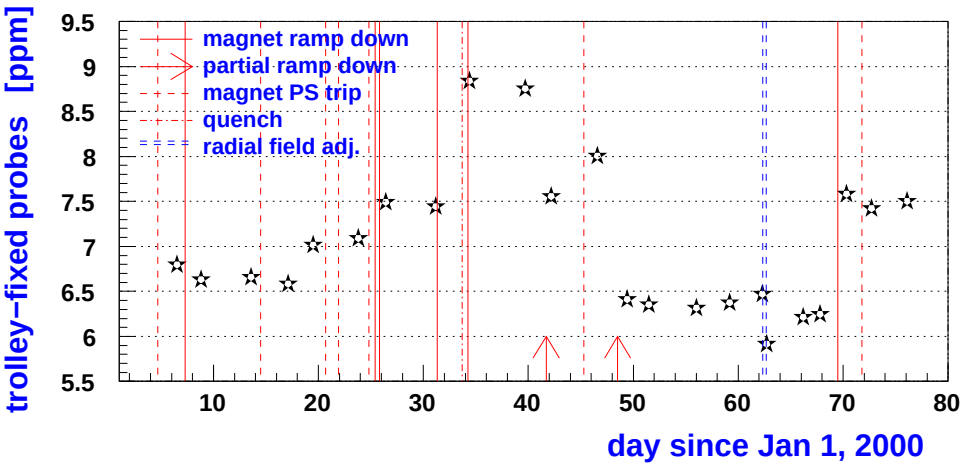



Figure 9.5: Difference between field measured by the trolley and the fixed probes as a function of time during the data collection period. (Reproduced from [119].)

of the magnetic field data were carried out; each employed approximately 120 of the probes, though the sets did not completely overlap. The effectiveness of the tracking by the fixed probes is assessed by examining the graph in Figure 9.5, which shows the difference between the field measured by the trolley and the fixed probes at the time of each trolley run. The fixed probe calibration is expected to change when the current in the magnet is cycled off and on because of hysteresis effects. These events occurred at the times indicated by vertical lines in the figure. Otherwise, the calibration is expected to remain constant between trolley runs; this figure shows that it indeed does, at the level of 0.10 ppm.

An automatic online analysis of the fixed probe data is used as part of a feedback system with the magnet power supply to stabilize the field. Ordinarily, this mechanism works well, keeping the field constant at the level of 1 ppm. However, the failure of a computer that was part of this system caused it to be disabled for a small part of the running period. The resulting variation of field as a function of run number is shown in Figure 9.6. Left to its own devices, the field drifted over a range of 50 ppm, mostly due to expansion and contraction as the yoke temperature changed. The fixed probe data permits this field drift to be measured and accounted for properly.

## 9.3 Summary

The number of positrons included for each run was conveyed by the $\omega_a$ analyzers to the $\omega_p$ team. Using this information, the positron-weighted, time-averaged magnetic field map was computed. The final step is to average the field over the muon distribution. This step was performed by inserting the field map into the beam tracking simulation program that was also used to compute the electric field and pitch corrections as described in Section 5.5. The field is quite uniform, so the dependence on the precise beam profile is small; a systematic uncertainty of 0.03 ppm is assigned to account for it.

The estimated systematic uncertainties for the $\omega_p$ measurement are summarized in Table 9.1. In addition

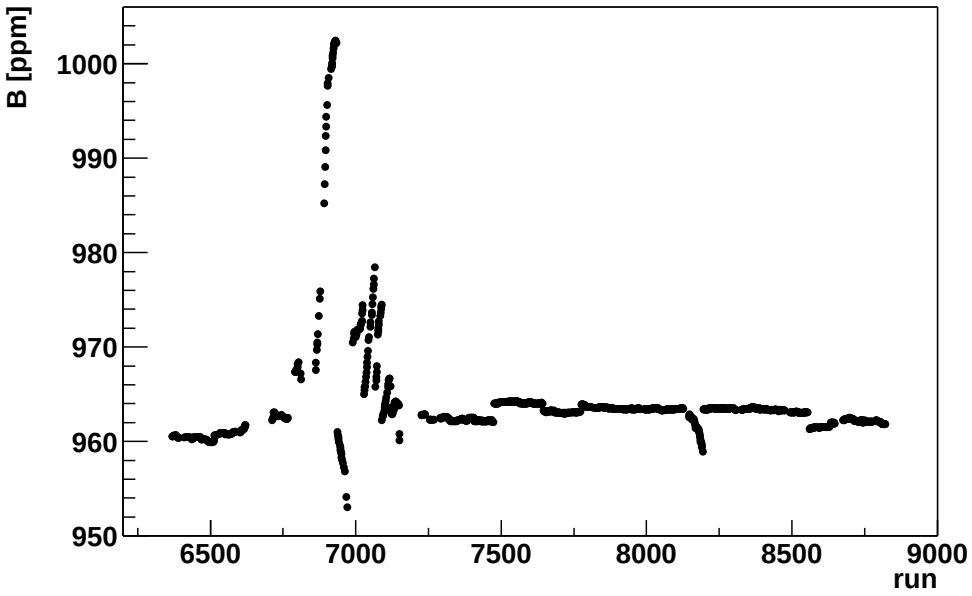



Figure 9.6: Magnetic field changes as a function of run number, illustrating the failure of the feedback system for a small part of the data set. (Reproduced from [117].)

| Effect | Uncertainty [ppm] |
|---|---|
| Absolute calibration of standard probe | 0.05 |
| Calibration of trolley against standard probe | 0.15 |
| Azimuthal position uncertainty during trolley measurement | 0.10 |
| Reproducibility of fixed probe calibration | 0.10 |
| Average over muon distribution | 0.03 |
| Others | 0.10 |
| Total systematic uncertainty | 0.24 |

Table 9.1: Summary of systematic uncertainties for $\omega_p$ measurement.

to the effects previously described in the text, there is an entry for "others." This line includes the dependence of the measured field on the trolley's temperature and power supply voltage, residual eddy currents from the kicker, and the coupling of higher multipoles to the beam, which was slightly off-center vertically. Individually, these effects are all extremely small; together, they contribute 0.10 ppm to the uncertainty in the field determination. The uncertainties listed in the table are believed to be uncorrelated, so they are combined in quadrature to obtain a total value of 0.21 ppm.

In conclusion, the appropriately weighted magnetic field in terms of the free proton spin precession frequency is[2]

$$\omega_p = 61\,791\,592.1(15.0)\ \text{Hz}. \tag{9.1}$$

[2]This value differs slightly from the one quoted in [3] because it is adjusted to match the data selection used in the energy-binned $\omega_a$ analysis rather than the average data selection of the four $\omega_a$ analyses.

# Chapter 10

# Conclusion

## 10.1 Comparison of independent $\omega_a$ analyses

Four individuals from different institutions independently extracted $\omega_a$ for this data set; the four results [120, 90, 121, 122] are listed in Table 10.1. They have each been adjusted to correct for the difference in positron-weighted $\omega_p$ values that arise from slightly different data selections; these corrections are all less than 0.1 ppm. The overlap among the different data sets included in the various analyses was computed in order to determine the statistically allowed deviations from one result to another. Among the origins of incomplete overlap are the different run and fill selections and pulse fitting implementations. However, the non-overlap between the analysis described in this thesis (indicated in the table as "D") and the others is primarily due to the lower energy threshold of 1.4 GeV and the procedure of choosing a different start time for each energy bin. With one exception, all of the analyses agree with each other within the statistical expectation alone. With no exceptions, all results agree within the estimated systematic uncertainties.

Consequently, the results were averaged to determine a final collaboration value. Each analysis was weighted equally, yielding

$$R = 128.56 \pm 0.62\ (\text{stat.}) \pm 0.31\ (\text{syst.})\ \text{ppm}. \tag{10.1}$$

The final uncertainties were determined through a covariance matrix formalism that took into account the

| Analysis | Institution | Method | Data set | $R \pm$ stat. $\pm$ syst. [ppm] |
|---|---|---|---|---|
| A | Yale | Conventional | G2OFF | $128.66 \pm 0.64 \pm 0.37$ |
| B | Brookhaven | Conventional | G2OFF | $128.34 \pm 0.64 \pm 0.36$ |
| C | Minnesota | Ratio [114, 123] | G2Too | $128.51 \pm 0.63 \pm 0.34$ |
| D | Illinois | $E$ binned | G2Too | $128.71 \pm 0.69 \pm 0.31$ |

$|R_B - R_A| = 0.32 \pm 0.14$ ppm (stat.)
$|R_C - R_A| = 0.15 \pm 0.23$ ppm (stat.)
$|R_D - R_A| = 0.05 \pm 0.47$ ppm (stat.)

Table 10.1: Comparison of the results of four independent $\omega_a$ analyses.

data overlap fractions [124]. The statistical error for the combined result is less than for any of the individual analyses because they are not completely correlated. Although the energy-binned analysis on its own has a slightly larger statistical error than the others, it is based on a very different subset of the data, so it contributes to this reduction of the final statistical uncertainty. After removing the previously secret offset and applying the electric field and pitch correction, the collective value obtained for $\omega_a$ is[1]

$$\omega_a = 229\,074.10(0.16)\ \text{Hz}\ . \tag{10.2}$$

## 10.2 Computation of $a_\mu$

In principle, $a_\mu$ may be computed from $\omega_a$ and $B$ according to Equation 1.18:

$$a_\mu = \frac{m_\mu}{e}\frac{\omega_a}{B} = \frac{m_\mu}{e}\mu_p\frac{\omega_a}{\omega_p}\ . \tag{10.3}$$

To perform this calculation, precise knowledge of three external constants ($m_\mu$, $e$, and $\mu_p$) is required. Today, these constants are known with appropriate precision to carry out the calculation directly. However, at the time the experiment was designed, they were not. Consequently, an alternative procedure inherited from the CERN storage ring experiments was employed. The following expression may be written down:

$$a_\mu = \frac{\omega_a}{\omega_s - \omega_a} = \frac{\left(\frac{e}{m_\mu}a_\mu B\right)}{\left(\frac{e}{m_\mu}\frac{g}{2}B\right) - \left(\frac{e}{m_\mu}a_\mu B\right)}\ . \tag{10.4}$$

Division of the numerator and denominator by $\omega_p$ yields

$$a_\mu = \frac{\frac{\omega_a}{\omega_p}}{\frac{\omega_s}{\omega_p} - \frac{\omega_a}{\omega_p}} = \frac{R_\omega}{\lambda - R_\omega}\ \text{with} \tag{10.5}$$

$$R_\omega = \frac{\omega_a}{\omega_p}\ \text{and} \tag{10.6}$$

$$\lambda = \frac{\omega_s}{\omega_p} = \frac{\mu_\mu}{\mu_p}\ . \tag{10.7}$$

This rearrangement leaves only one external constant $\lambda$, which is the ratio of the magnetic moments of the muon and the proton. It has been measured in an experiment at Los Alamos National Laboratory involving the hyperfine level structure of muonium [125] as

$$\lambda = 3.183\,345\,39(10)\ . \tag{10.8}$$

[1] This value differs slightly from the one given in [3]; see the footnote at the end of Chapter 9.

The same NMR system and standard probe used to determine $\omega_p$ were used in this muonium experiment, so many types of calibration errors tend to cancel out.

Following this procedure, the collaboration value for $a_\mu$ is

$$a_\mu^{exp} = 11\,659\,203.9(8.4) \times 10^{-10} \ , \tag{10.9}$$

while the value that would have been obtained from the energy-binned $\omega_a$ analysis alone is only slightly different:

$$a_\mu^{exp;E\ bin} = 11\,659\,201.9(9.3) \times 10^{-10} \ . \tag{10.10}$$

After combining $a_\mu^{exp}$ with the results from previous running periods [21, 22], a new world average is obtained:

$$a_\mu^{exp;avg} = 11\,659\,203(8) \ . \tag{10.11}$$

The indicated uncertainty is slightly larger than a naive calculation would imply because it accounts for correlated systematic uncertainties across running periods.

## 10.3 Comparison of experimental and theoretical results

The values of $a_\mu$ calculated from the standard model were described in Chapter 2:

$$a_\mu^{SM;e^+e^-} = 11\,659\,168.8(7.7) \times 10^{-10} \text{ and} \tag{10.12}$$

$$a_\mu^{SM;\tau} = 11\,659\,193.2(6.8) \times 10^{-10} \ . \tag{10.13}$$

The hadronic vacuum polarization contributions to these two values are determined from hadron production in $e^+e^-$ collisions and $\tau$ decay, respectively. The corresponding differences between the new world average experimental result $a_\mu^{exp;avg}$ and the two theoretical values are

$$a_\mu^{exp;avg} - a_\mu^{SM;e^+e^-} = 34(11) \times 10^{-10} \text{ and} \tag{10.14}$$

$$a_\mu^{exp;avg} - a_\mu^{SM;\tau} = 10(10) \times 10^{-10} \ . \tag{10.15}$$

Very different conclusions may be drawn depending on which of the standard model evaluations is used. With the $e^+e^-$ annihilation data, a discrepancy between experiment and theory of more than three standard deviations is observed. The probability that a difference this large would occur purely by chance as a statistical fluctuation is less than 0.3 percent. Consequently, this comparison could be interpreted as a clear signal of physics beyond the standard model, possibly indicating the existence of supersymmetry. However, this exciting speculation is spoiled by the $\tau$ decay data. If it is used to establish the standard model value, then experiment and theory seem to agree well, within one standard deviation. In this case, $a_\mu$ will place constraints on "new physics" rather than trumpeting its arrival.

## 10.4 Outlook

During the winter of 2001, the polarities of all of the magnetic and electric fields in the experiment were reversed. A data set was collected to permit the determination of $a_\mu$ for the negative muon. The analysis of these data is in progress; it is expected to provide a result that is statistically comparable to the one described here for the positive muon. The negative muon result will serve as a test of the *CPT* theorem, which predicts identical values for the positive and negative muon. It should also have a smaller systematic uncertainty because it was collected with the benefit of experience and history. The quadrupoles were operated at two different $n$ values, each of which yielded a CBO frequency much farther from $2\omega_a$ than the one used in 2000. Also, the average beam intensity per fill was somewhat lower, so the effects of pileup and the pion flash were both reduced.

As of this writing, however, the ball is in the standard model theorists' court. In order to disambiguate the interpretation of the result, it will be necessary to clarify which of the approaches to computing the hadronic vacuum polarization contribution is correct. Efforts to measure the hadron production cross section through radiative return from high beam energies are in progress at KLOE and BaBar [31, 32]. These measurements may prove sufficiently sensitive to effectively check the CMD-2 $e^+e^-$ data. A lattice QCD calculation of the hadronic contribution is also under way [33]. When these studies are complete, the implications of this measurement of $a_\mu$ will become clear.

# References


[1] Particle Data Group, K. Hagiwara *et al.*, Phys. Rev. **D66**, 010001 (2002).

[2] T. Kinoshita and W. B. Lindquist, Phys. Rev. Lett. **47**, 1573 (1981).

[3] Muon *g*-2 Collaboration, G. W. Bennett *et al.*, Phys. Rev. Lett. **89**, 101804 (2002), [hep-ex/0208001].

[4] M. Davier, S. Eidelman, A. Hocker and Z. Zhang, hep-ph/0208177.

[5] L3 Collaboration, M. Acciarri *et al.*, Phys. Lett. **B434**, 169 (1998).

[6] M. A. Samuel, G.-W. Li and R. Mendel, Phys. Rev. Lett. **67**, 668 (1991).

[7] S. Gasiorowicz and J. L. Rosner, Am. J. Phys. **49**, 954 (1981).

[8] P. J. Mohr and B. N. Taylor, Rev. Mod. Phys. **72**, 351 (2000).

[9] F. Halzen and A. D. Martin, *Quarks and Leptons: an introductory course in modern particle physics* (John Wiley and Sons, New York, 1984).

[10] L. Foldy, Phys. Rev. **87**, 688 (1952).

[11] J. Schwinger, Phys. Rev. **73**, 416 (1947).

[12] L. Thomas, Phil. Mag. **3**, 1 (1927).

[13] F. Farley and E. Picasso, The muon *g*-2 experiments, in *Quantum Electrodynamics*, edited by T. Kinoshita, pp. 479–559, World Scientific, Singapore, 1990.

[14] J. Jackson, *Classical Electrodynamics*, 3rd ed. (John Wiley and Sons, New York, 1999).

[15] V. Bargmann, L. Michel and V. Telegdi, Phys. Rev. Lett. **2**, 453 (1959).

[16] P. Debevec, Private communication.

[17] G. Charpak *et al.*, Nuovo Cimento Soc. Ital. Fis. **37**, 1241 (1965).

[18] J. Bailey *et al.*, Nuovo Cimento Soc. Ital. Fis. **A9**, 369 (1972).

[19] CERN Muon Storage Ring Collaboration, J. Bailey *et al.*, Phys. Lett. **B67**, 225 (1977).

[20] R. M. Carey *et al.*, Phys. Rev. Lett. **82**, 1632 (1999).

[21] Muon *g*-2 Collaboration, H. N. Brown *et al.*, Phys. Rev. **D62**, 091101 (2000), [hep-ex/0009029].

[22] Muon *g*-2 Collaboration, H. N. Brown *et al.*, Phys. Rev. Lett. **86**, 2227 (2001), [hep-ex/0102017].

[23] T. Kinoshita, Theory of the anomalous magnetic moment of the electron–numerical approach, in *Quantum Electrodynamics*, edited by T. Kinoshita, pp. 218–321, World Scientific, Singapore, 1990.

[24] T. Kinoshita and W. J. Marciano, Theory of the muon anomalous magnetic moment, in *Quantum Electrodynamics*, edited by T. Kinoshita, pp. 419–478, World Scientific, Singapore, 1990.

[25] V. W. Hughes and T. Kinoshita, Rev. Mod. Phys. **71**, S133 (1999).

[26] T. Kinoshita and M. Nio, Phys. Rev. Lett. **90**, 021803 (2003), [hep-ph/0210322].

[27] E. de Rafael, hep-ph/0208251.

[28] CMD-2 Collaboration, R. R. Akhmetshin *et al.*, Phys. Lett. **B527**, 161 (2002), [hep-ex/0112031].

[29] K. Hagiwara, A. D. Martin, D. Nomura and T. Teubner, hep-ph/0209187.

[30] ALEPH Collaboration, R. Barate *et al.*, Z. Phys. **C76**, 15 (1997).

[31] KLOE Collaboration, G. Venanzoni *et al.*, hep-ex/0210013.

[32] BaBar Collaboration, E. P. Solodov, eConf **C010430**, T03 (2001), [hep-ex/0107027].

[33] T. Blum, hep-lat/0212018.

[34] B. Krause, Phys. Lett. **B390**, 392 (1997), [hep-ph/9607259].

[35] J. Bijnens, Phys. Rept. **265**, 369 (1996), [hep-ph/9502335].

[36] E. de Rafael, Phys. Lett. **B322**, 239 (1994), [hep-ph/9311316].

[37] J. Bijnens, E. Pallante and J. Prades, Nucl. Phys. **B474**, 379 (1996), [hep-ph/9511388].

[38] M. Hayakawa, T. Kinoshita and A. I. Sanda, Phys. Rev. **D54**, 3137 (1996), [hep-ph/9601310].

[39] M. Knecht and A. Nyffeler, Phys. Rev. **D65**, 073034 (2002), [hep-ph/0111058].

[40] M. Knecht, A. Nyffeler, M. Perrottet and E. De Rafael, Phys. Rev. Lett. **88**, 071802 (2002), [hep-ph/0111059].

[41] B. Schwarzschild, Phys. Today **55**, 18 (2002).

[42] M. Hayakawa and T. Kinoshita, hep-ph/0112102.

[43] J. Bijnens, E. Pallante and J. Prades, Nucl. Phys. **B626**, 410 (2002), [hep-ph/0112255].

[44] I. Blokland, A. Czarnecki and K. Melnikov, Phys. Rev. Lett. **88**, 071803 (2002), [hep-ph/0112117].

[45] M. Ramsey-Musolf and M. B. Wise, Phys. Rev. Lett. **89**, 041601 (2002), [hep-ph/0201297].

[46] A. Czarnecki, W. J. Marciano and A. Vainshtein, hep-ph/0212229.

[47] T. V. Kukhto, E. A. Kuraev, Z. K. Silagadze and A. Schiller, Nucl. Phys. **B371**, 567 (1992).

[48] A. Czarnecki, B. Krause and W. J. Marciano, Phys. Rev. **D52**, 2619 (1995), [hep-ph/9506256].

[49] A. Czarnecki, B. Krause and W. J. Marciano, Phys. Rev. Lett. **76**, 3267 (1996), [hep-ph/9512369].

[50] M. Knecht, S. Peris, M. Perrottet and E. De Rafael, hep-ph/0205102.

[51] E. Eichten, K. D. Lane and M. E. Peskin, Phys. Rev. Lett. **50**, 811 (1983).

[52] P. Mery, S. E. Moubarik, M. Perrottet and F. M. Renard, Z. Phys. **C46**, 229 (1990).

[53] S. J. Brodsky and S. D. Drell, Phys. Rev. **D22**, 2236 (1980).

[54] K. D. Lane, hep-ph/0102131.

[55] S. P. Martin, hep-ph/9709356.

[56] A. Czarnecki and W. J. Marciano, Phys. Rev. **D64**, 013014 (2001), [hep-ph/0102122].

[57] J. L. Feng and K. T. Matchev, Phys. Rev. Lett. **86**, 3480 (2001), [hep-ph/0102146].

[58] M. Byrne, C. Kolda and J. E. Lennon, hep-ph/0208067.

[59] S. P. Martin and J. D. Wells, hep-ph/0209309.

[60] U. Chattopadhyay and P. Nath, hep-ph/0208012.

[61] J. L. Feng, K. T. Matchev and Y. Shadmi, Nucl. Phys. **B613**, 366 (2001), [hep-ph/0107182].

[62] W. Bernreuther and M. Suzuki, Rev. Mod. Phys. **63**, 313 (1991).

[63] F. Hoogeveen, Nucl. Phys. **B341**, 322 (1990).

[64] K. S. Babu, B. Dutta and R. N. Mohapatra, Phys. Rev. Lett. **85**, 5064 (2000), [hep-ph/0006329].

[65] CERN Muon Storage Ring Collaboration, J. Bailey *et al.*, J. Phys. **G4**, 345 (1978).

[66] R. McNabb, Muon *g*-2 Internal Note No. 398, 2001.

[67] Y. K. Semertzidis *et al.*, Int. J. Mod. Phys. **A16S1B**, 690 (2001).

[68] D. Brown *et al.*, Design report: BNL AGS E821: A new precision measurement of the muon *g*-2 value at the level of 0.35 ppm, 1995.

[69] L. A. Ahrens *et al.*, prepared for IEEE Particle Accelerator Conference (PAC99), New York, Mar. 29-Apr. 21, 1999.

[70] D. Kawall, Private communication.

[71] G. T. Danby *et al.*, Nucl. Instrum. Meth. **A457**, 151 (2001).

[72] A. Yamamoto *et al.*, Nucl. Instrum. Meth. **A491**, 23 (2002).

[73] F. Krienen, D. Loomba and W. Meng, Nucl. Instrum. Meth. **A283**, 5 (1989).

[74] E. Efstathiadis *et al.*, Nucl. Instrum. Meth. **A496**, 8 (2003).

[75] W. Morse, Muon *g*-2 Internal Note No. 316, 1998.

[76] W. Morse and Y. Semertzidis, Muon *g*-2 Internal Note No. 332, 1998.

[77] Y. K. Semertzidis *et al.*, Nucl. Instrum. Meth. **A**, accepted for publication.

[78] S. A. Sedykh *et al.*, Nucl. Instrum. Meth. **A455**, 346 (2000).

[79] P. Cushman and S. Hou, Nucl. Instrum. Meth. **A339**, 456 (1994).

[80] V. Druzhinin *et al.*, Muon *g*-2 Internal Note No. 315, 1998.

[81] R. Carey, Private communication.

[82] R. Prigl *et al.*, Nucl. Instrum. Meth. **A374**, 118 (1996).

[83] X. Fei, V. Hughes and R. Prigl, Nucl. Instrum. Meth. **A394**, 349 (1997).

[84] W. J. Deninger, *A measurement of the magnetic field systematic correction to the muon anomalous magnetic moment associated with muon phase space in experiment BNL E821*, Ph.D. thesis, University of Illinois at Urbana-Champaign, 1999, UMI-99-44831.

[85] W. D. Peterson, *The VMEbus Handbook*, 4th ed. (VMEbus International Trade Association, 1997).

[86] R. Carey, Muon *g*-2 Internal Note No. 301, 1998.

[87] W. Earle, E. Hazen, B. Safford and G. Varner, Muon *g*-2 Internal Note No. 207, 1994.

[88] I. Kronkvist, P. Cushman, L. Duong and C. Timmermans, (1998), prepared for Computing in High-energy Physics (CHEP98), Chicago, IL, August 31-Sept. 4, 1998.

[89] M. Nomachi *et al.*, prepared for Computing in High-energy Physics (CHEP94), San Francisco, CA, Apr. 21-27, 1994.

[90] C. S. Özben, Muon *g*-2 Internal Note No. 423, 2002.

[91] Y. Orlov, C. S. Özben and Y. K. Semertzidis, Nucl. Instrum. Meth. **A482**, 767 (2002).

[92] E. Benedict *et al.*, Beam dynamics in the muon *g*-2 storage ring, manuscript in preparation.

[93] W. Morse, Private communication.

[94] S. Granger and G. Ford, Phys. Rev. Lett. **28**, 1479 (1972).

[95] F. J. M. Farley, Phys. Lett. **B42**, 66 (1972).

[96] J. Paley, Muon *g*-2 Internal Note No. 414, 2002.

[97] F. Farley, Private communication.

[98] J. Field and G. Fiorentini, Nuovo Cimento **21A**, 297 (1974).

[99] G. Onderwater and J. Pretz, Muon *g*-2 Internal Note No. 363, 2000.

[100] F. Gray and G. Onderwater, Muon *g*-2 Internal Note No. 417, 2002.

[101] R. Brun and F. Rademakers, Nucl. Instrum. Meth. **A389**, 81 (1997).

[102] I. Logashenko, Muon *g*-2 Internal Note No. 369, 2000.

[103] I. Logashenko, Muon *g*-2 Internal Note No. 378, 2000.

[104] F. James and M. Roos, Comput. Phys. Commun. **10**, 343 (1975).

[105] M. Matsumoto and T. Nishimura, ACM Trans. Model. Comput. Simul. **8**, 3 (1998).

[106] F. Farley, C. Özben, J. Pretz and Y. Semertzidis, Muon *g*-2 Internal Note No. 377, 2001.

[107] C. Polly, Private communication.

[108] G. Fasano and A. Franceschini, Mon. Not. R. Astr. Soc. **225**, 155 (1987).

[109] S. Redin, Muon *g*-2 Internal Note No. 410, 2002.

[110] D. Kawall, Muon *g*-2 Internal Note No. 322, 1998.

[111] S. Redin, Muon *g*-2 Internal Note No. 387, 2001.

[112] Y. Semertzidis, Muon *g*-2 Internal Note No. 403, 2001.

[113] W. Morse, Muon *g*-2 Internal Note No. 404, 2002.

[114] L. H. Duong, *A precise measurement of the anomalous magnetic moment of the positive muon*, Ph.D. thesis, University of Minnesota, 2001, UMI-30-31973.

[115] Y. Semertzidis, Muon *g*-2 Internal Note No. 426, 2002.

[116] W. Morse, Muon *g*-2 Internal Note No. 346, 1999.

[117] H. Deng *et al.*, Muon *g*-2 Internal Note No. 432, 2003.

[118] A. P. Grossmann, *Magnetic field determination in a superferric storage ring for a precise measurement of the muon magnetic anomaly*, Ph.D. thesis, Ruprecht-Karls-Universität Heidelberg, 1998.

[119] R. Prigl, Private communication.

[120] M. Deile, Muon *g*-2 Internal Note No. 421, 2002.

[121] B. Bousquet, Private communication.

[122] F. Gray, Muon *g*-2 Internal Note No. 415, 2002.

[123] J. M. Kindem, *The Anomalous magnetic moment of the positive muon*, Ph.D. thesis, University of Minnesota, 1998, UMI-99-03633.

[124] O. Rind and E. Sichtermann, Muon *g*-2 Internal Note No. 380, 2001.

[125] W. Liu *et al.*, Phys. Rev. Lett. **82**, 711 (1999).

# Vita

Frederick Earl Gray, Jr. was born in Charlotte, North Carolina on August 27, 1975. In 1997, he graduated with a bachelor's degree in physics and computer science from Rice University, where he was a member of Wiess College. As a graduate student in Urbana-Champaign, in addition to performing the work described in this thesis, he participated in two experiments at the Paul Scherrer Institute in Villigen, Switzerland. One, called $\mu$Lan, will measure the lifetime of the positive muon with a precision of 1 ppm. The other, $\mu$Cap, will measure the muon capture probability in isotopically pure hydrogen gas. Following the completion of his Ph.D., he will take a position as a visiting postdoctoral scholar in the Weak Interactions Group of the Department of Physics at the University of California, Berkeley. At Berkeley, he will continue to work on the PSI experiments, and he will also join the KamLAND collaboration.