Precession frequency analysis for Run-2 and Run-3 of the Muon $g-2$ experiment

Hannah Binney

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington
2022

Reading Committee:
David Hertzog, Chair
Jason Detwiler
Gray Rybka

Program Authorized to Offer Degree:
Physics

© Copyright 2022

Hannah Binney

University of Washington

Abstract

Precession frequency analysis for Run-2 and Run-3 of the Muon $g-2$ experiment

Hannah Binney

Chair of the Supervisory Committee:

Professor David Hertzog
Physics

The Fermilab Muon $g-2$ experiment aims to measure the anomalous magnetic momentum of the muon, a_μ, to 140 part per billion precision. This measurement is compared to the Standard Model (SM) value for a_μ as a precision test of the SM. The result of the first run of data taking was published in April 2021 and verified the results of the previous experiment at Brookhaven National Laboratory. The average measured value of a_μ now stands in tension with the official Standard Model calculation to 4.2σ. The anomalous magnetic moment a_μ is measured by taking the ratio of two frequencies: the anomalous precession frequency ω_a, which is the difference between the spin and cyclotron precession frequencies of a muon in a magnetic storage ring, and $\tilde{\omega}_p$, which is the muon-weighted storage ring magnetic field as measured by the precession frequency of protons in nuclear magnetic resonance probes. This thesis presents a precession frequency analysis of the second and third data taking runs, which have a combined statistical uncertainty of 200 ppb, more than twice as precise as the Run-1 measurement. A summary of the data preparation for fitting will be described, including an update to the algorithm used to reconstruct decay positron information from the detectors. Then, the fitting procedure used to extract ω_a will be discussed. Finally, the systematic uncertainties associated with the precession frequency analysis will be analyzed.

Acknowledgements

Writing this thesis has been a long journey. I would not have accomplished it without the knowledge, encouragement, commiseration, and support of friends, family, and coworkers over the past six and a half years.

First, I'd like to thank my advisor David Hertzog, whose belief in me, enthusiasm for physics, and wisdom have have been unwavering. I am so glad I decided to work on Muon g-2. I'd also like to thank all the past and current UW group members, who taught me most of the things I know about the experiment: Aaron Fienberg, Jason Hempstead, Jarek Kaspar, Kim Siang Khaw, Rachel Osofsky, Patrick Schwendimann, Christine Claessens, Erik Swanson, and Peter Kammel. In particular, I would like to thank Brynn MacCoy and Josh Labounty, who have been with me almost every step of this journey and who have been the best office mates a grad student could ask for. I would additionally like to thank $g-2$ collaborators James Mott, Sean Foster, and Kevin Labe for their valuable discussions on this analysis and for answering all my Slack messages. Thank you to my committee, Martin Savage, Anna Goussiou, Gray Rybka, Jason Detwiler, and Michael McCarthy.

I'd further like to thank my Seattle friends, who listened to me complain, who suffered through homework with me, and who have brought me joy and friendship along this journey: Sam Kowash, Lauren Pittis, Paul Malinowski, Sebastian Alvis, and Amanda Manaster. Thank you for the potlucks, trivia, and Renaissance faires. I'd also like to thank Anna Burkatovsky, Nicole Hodgkinson, and Helaina Peck, whom I've known for half my life or longer, and on whom I continue to rely to this day. I'd like to thank Sean, who has been a bright spot in the darkest of times, and who always makes me coffee.

Finally, I'd like to thank Mom, Dad, and Sarah, who have been my closest allies and biggest supporters my entire life.

DEDICATION

To Sarah and Sean

Contents

List of Figures

List of Tables

Glossary

BSM Beyond the Standard Model.

CBO Coherent betatron oscillation, the largest beam oscillation measured in the calorimeters, arising from horizontal oscillations that are aliased at the cyclotron frequency.

clock ticks Time between digitizer samples, equivalent to about 1.25 nanoseconds. Abbreviated as c.t.

Detector acceptance Likelihood that a particle (e.g. positron) is measured in a detector, depending on the particle's characteristics (e.g. decay position in the storage region).

ESQ Electrostatic quadrupole.

FFT Fast Fourier transform. Used to identify frequencies in the residual of the precession frequency fit function.

Recon East Reconstruction algorithm that uses a "global" fitting approach, where a 3×3 square of channels is fit around the channel that exceeds an energy threshold. Recon East also uses spatial separation in its clustering algorithm.

Recon West Reconstruction algorithm that uses channel-by-channel fitting thresholds and time partitioning for clustering. This analysis is based on the Recon West algorithm.

SM The Standard Model of particle physics.

TDR Technical Design Report.

Chapter 1

Introduction

The Fermilab Muon $g-2$ experiment is an effort on the part of 191 collaborators across 35 institutions in 7 countries to measure a single value: the anomalous magnetic moment of the muon a_μ. Further, hundreds of theorists have contributed to the Standard Model (SM) prediction for a_μ and its possible beyond the Standard Model (BSM) implications. A long history of experiments and calculations have made possible the extremely high precision comparison performed today. The Brookhaven National Lab (BNL) Muon $g-2$ experiment published its final result in 2006 [1], and a tension has persisted between theory and experiment since then, motivating the creation of a new higher-precision experiment at Fermilab National Laboratory (FNAL). At the time of the publication of the Run-1 FNAL Muon $g-2$ experiment, the tension stood at 3.7σ. Following this publication [2–5], upon combination with the BNL result, the official Standard Model calculation [6] now stands in tension with the experimental average to 4.2σ. Figure 1.1 shows the SM, BNL, and FNAL results. This tension is a tantalizing hint of physics beyond the Standard Model (BSM) and is of great interest to the particle physics community, with many hundreds of citations of the Brookhaven and Fermilab publications. However, it still remains below the 5σ threshold for a particle physics discovery, giving strong motivation for both improvement of the SM calculation and analysis of data beyond Run-1.

This thesis will primarily focus on the contributions that this author has made to the analysis of Run-2 and Run-3 data taken in 2019-2020, with other relevant topics summarized. The two components of the a_μ measurement, the anomalous precession frequency of the muon and the magnetic field measurement, are equally important, but this thesis focuses on the former. Chapter 1 will introduce magnetic dipole

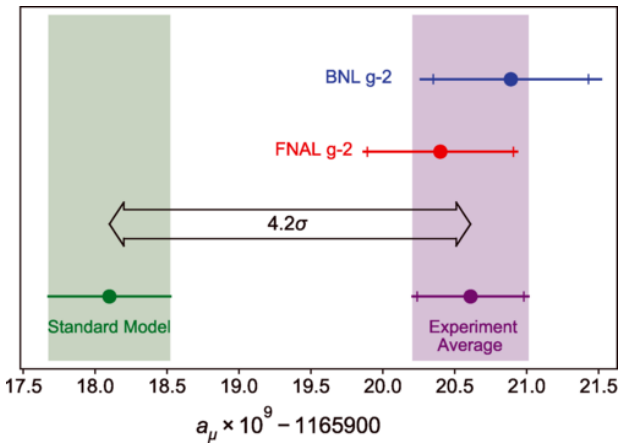


Figure 1.1: Comparison of the Standard Model value for a_μ from the 2020 Theory Initiative White Paper [6] to the Brookhaven measurement (in blue) [1], the 2021 FNAL Run-1 measurement (in red) [2], and the experimental average (in purple). The tension between theory and experiment is now 4.2σ. The SM value does not include the recent lattice QCD results discussed in subsection 1.4.3. Reproduced from [2].

moments and will summarize the Standard Model (SM) calculation used to determine the theory value of a_μ. Chapter 2 will discuss the experimental principle of the Fermilab $g-2$ experiment and the hardware systems used to store the muons and measure their decay into positrons. Chapter 3 will summarize the full analysis procedure and discuss the main sources of uncertainty in Run-1 and how they may be mitigated in future runs. Chapter 4 will discuss the procedures used to reconstruct physically meaningful data from raw detector hits, including an update to a reconstruction algorithm implemented by this author. Chapter 5 will discuss the details of the fitting procedure used to extract the anomalous precession frequency of the muon, the main focus of this thesis. Finally, Chapter 6 will discuss systematic uncertainties on the measurement of the anomalous precession frequency ω_a. At the time of this writing, the values of ω_a and a_μ remain blinded.

1.1 Background and motivation

The anomalous magnetic moment of the muon a_μ can be both calculated and measured very precisely, making it a good test of the Standard Model. Since the measurement of a_μ at BNL, which published its final result in 2006 [1], a discrepancy has persisted between the SM and experimental values. The Muon $g-2$ experiment at Brookhaven National Lab (BNL) measured $a_\mu \equiv \frac{g-2}{2}$ to be

$$a_\mu^{\mathrm{BNL}} = 116\,592\,080(54)(33) \times 10^{-11}, \tag{1.1}$$

where the statistical and systematic uncertainties are given, respectively.

The Standard Model (SM) calculation for Muon $g-2$ has progressed over decades to reach its current high precision. The 2020 Muon $g-2$ Theory Initiative White Paper [6] established the official theoretical prediction for a_μ^{SM} to 370 ppb precision:

$$a_\mu^{\text{SM}} = 116\,591\,810(43) \times 10^{-11}. \tag{1.2}$$

Notably, this value does not incorporate some recent calculations using lattice quantum chromodynamics (QCD), which are discussed further in subsection 1.4.3. Despite many improvements made since 2006 to improve the rigor of the SM calculations, the central value has remained approximately the same, resulting in a tension of 3.7σ between a_μ^{SM} and a_μ^{BNL}.

This tension motivated the creation of the FNAL Muon $g-2$ experiment to improve the precision of the measurement, ideally either resolving or confirming the BNL discrepancy. At FNAL, a higher intensity muon beam could be generated, allowing the experiment to achieve significantly higher precision [7]. The same experimental procedure used at BNL is used at FNAL, involving storing muons in a large storage magnet and observing their decay into positrons. Additionally, the storage ring magnet was transported from BNL to FNAL for reuse in the FNAL experiment. However, many upgrades to the detector and ring systems systems, field uniformity, simulations, and data acquisition hardware were made for the new FNAL experiment [7, 8]. For example, although the storage magnet is reused from BNL, the field has been fully re-shimmed, and it has been re-calibrated using the MRI magnet present at Argonne National Laboratory. The calorimeter detectors were completely redesigned, and a sophisticated laser system has been implemented to maintain gain and timing stability. The magnetic kicker and electrostatic quadrupole systems were also rebuilt. Therefore, although some types of systematic uncertainties are the same for both experiments, the two measurements can be treated as independent and averaged together.

The FNAL experiment started data taking in 2018 and published its first results in April 2021, releasing 4 papers simultaneously: a summary letter [2] and papers on the three major components of the analysis [3–5]. The first result of the FNAL experiment was

$$a_\mu^{\text{FNAL}} = 116\,592\,040(54) \times 10^{-11}, \tag{1.3}$$

a 460 ppb measurement. This measurement confirmed the BNL result, as seen in Figure 1.1.

The averaged experimental value a_μ^{exp} then becomes

$$a_\mu^{\text{exp}} = 116\,592\,061(41) \times 10^{-11}, \tag{1.4}$$

a 350 ppb measurement. The tension between a_μ^{SM} and a_μ^{exp} is now 4.2σ.

The ultimate uncertainty goal of the FNAL experiment is 140 ppb, as described in the technical design report (TDR) for the experiment [7]. This thesis concerns the Run-2 and Run-3 data, which will be the highest precision measurement so far, with approximately 200 ppb statistical precision.

1.2 Classical magnetic dipoles

Classically, electric current distributions generate magnetic dipole moments (MDMs). In the simplest case where the charge and mass densities are the same, the MDM is [9]

$$\vec{\mu} = \frac{q}{2m}\vec{L}, \tag{1.5}$$

where q is the charge, m the mass, and $\vec{L}$ the total angular momentum. For example, a current loop or a rotating charged sphere will have an MDM in classical mechanics. When placed in an external magnetic field B, the potential energy U on $\vec{\mu}$ will be

$$U = -\vec{\mu} \cdot \vec{B}, \tag{1.6}$$

expressing the tendency of the MDM to align with the magnetic field. The MDM experiences a torque

$$\vec{\tau} = \vec{\mu} \times \vec{B}. \tag{1.7}$$

This torque causes the precession in the magnetic field at the frequency

$$\omega = \frac{qB}{2m}. \tag{1.8}$$

The frequency ω is thus proportional to the magnetic field strength B.

1.3 Magnetic dipoles in quantum mechanics

Early experiments showed that Equation 1.5 did not hold for quantum mechanical systems. Specifically, the electron had an additional factor of approximately 2. In order to account for the proportionality factor that arises in QM, the equation can be rewritten as

$$\vec{\mu} = g\frac{q}{2m}\vec{s}, \tag{1.9}$$

where g is the dimensionless "g-factor" and $\vec{s}$ is the spin of the particle. This means that the precession frequency in Equation 1.8 becomes

$$\omega = g\frac{qB}{2m}. \tag{1.10}$$

Therefore, by measuring ω and B, the g-factor can be determined.

The g-factor cannot be predicted without accounting for both quantum mechanics and special relativity. The Dirac equation [10] provided the solution for the experimental observation that the g-factor of the electron $g_e \approx 2$. Following the notation in [11], the covariant form of the Dirac equation in the presence of an external electromagnetic field for a particle with charge -1 is written as

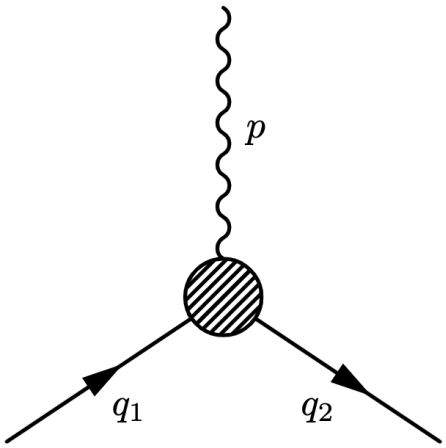
$$(i\not{\partial} - e\not{A} - m)\psi = 0, \tag{1.11}$$

which means that

$$\big((i\partial_\mu - eA_\mu)^2 - \frac{e}{2}F_{\mu\nu}\sigma^{\mu\nu} - m^2\big)\psi = 0, \tag{1.12}$$

where $\sigma^{\mu\nu} \equiv \frac{i}{2}[\gamma^\mu, \gamma^\nu]$ and γ are the Dirac matrices. Compared to the spin-0 Dirac equation, this equation contains an extra term $\frac{e}{2}F_{\mu\nu}\sigma^{\mu\nu}$, which represents the interaction of the magnetic dipole with the electromagnetic field. In the non-relativistic limit, this produces the Hamiltonian

$$H = \frac{\vec{p}^2}{2m} + V(r) + \frac{e}{2m}\vec{B}\cdot(\vec{L} + g\vec{s}), \tag{1.13}$$



[h]

Figure 1.2: Feynman diagram representation of the type of interaction between a fermion and photon (external magnetic field) that contributes to the anomalous magnetic moment a. The shaded region represents any loop.

where the spin $\vec{s} = \frac{1}{2}\vec{\sigma}$. By comparison of Equation 1.9 to Equation 1.13, the prediction of $g = 2$ for fundamental spin 1/2 particles is evident.

Later experiments, such as those relating to the hyperfine-structure of hydrogen [12, 13], determined that g is not exactly equal to 2 for the electron. This difference arises from higher-level loop corrections to the tree-level interaction of the particle with the electromagnetic field. The anomalous magnetic moment a is defined as

$$a \equiv \frac{g-2}{2}, \tag{1.14}$$

half the difference from the tree-level value of 2. Contributions to a come from graphs of the type in Figure 1.2, which represents a fermion interaction with a photon (external magnetic field). The shaded region represents any loop.

The matrix element of this general interaction can be derived as [11]:

$$i\mathcal{M}^{\mu} = (-ie)\bar{u}(q_2)\left[F_1\left(\frac{p^2}{m^2}\right)\gamma^{\mu} + \frac{i\sigma^{\mu\nu}}{2m}p_{\nu}F_2\left(\frac{p^2}{m^2}\right)\right]u(q_1), \tag{1.15}$$

where F_1 and F_2 are two independent form factors. The tree-level diagram will have $F_1 = 1$, $F_2 = 0$. Only the second term containing F_2 is spin-dependent, so the anomalous magnetic moment only depends on F_2. Additionally, since a is evaluated at low p, the anomalous magnetic moment is only related to $F_2(0)$.

Quantum field theory textbooks such as [11] derive the relation

$$g = 2 + 2F_2(0), \tag{1.16}$$

meaning that the anomalous magnetic moment

$$a = F_2(0). \tag{1.17}$$

1.4 Standard Model prediction for the anomalous magnetic moment

The anomalous magnetic moment of a lepton, $F_2(0)$, can be calculated in the Standard Model. Any diagram of the form shown in Figure 1.2 can contribute. The largest contribution comes from the first-order quantum electrodynamics (QED) loop, shown in Figure 1.3 (second from the left), also known as the Schwinger term, since it was first calculated by Julian Schwinger [14]. This can be derived as

$$a_{\text{Schwinger}} = \frac{\alpha}{2\pi} \approx 0.00116, \tag{1.18}$$

where α is the fine-structure constant. This contribution is the same for all leptons. Higher-order loop contributions involving W, Z, or Higgs particles are suppressed by at least a factor of $\frac{\alpha}{2\pi}\frac{m_l^2}{m_W^2}$ [15], where m_l is the mass of the lepton in question. Therefore these contributions are $m_\mu^2/m_e^2 \approx 43,000$ times larger in the muon than the electron. Similarly, if new physics is present at a certain mass scale Λ, it is generally assumed that the contribution to a from the new physics is proportional to the lepton mass squared and inversely proportional to Λ squared [16]:

$$|\Delta a^{\text{new physics}}| \propto \frac{m_l^2}{\Lambda^2}, \tag{1.19}$$

Therefore, although a_e is one of the most precisely measured quantities in physics, a_μ is still significantly more sensitive to potential new physics, justifying the choice to measure a_μ in this experiment. The following sections summarize calculations of contributions to a_μ in the Standard Model, which is generally

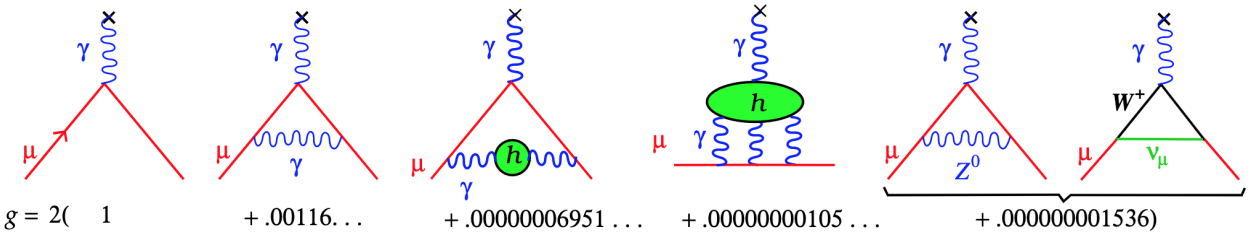


Figure 1.3: Representative Feynman diagrams for four contributions to a_μ^{SM}. From left to right: tree level diagram, lowest-order contribution to a_μ^{QED} (Schwinger term), $a_\mu^{\text{had,HVP}}$, and $a_\mu^{\text{had,LbL}}$, and a_μ^{EW}. The relative contributions to g from each category are shown on the bottom. Reproduced from L. Roberts.

Contribution	Value $\times 10^{11}$	Uncertainty contribution [ppb]
QED	116 584 718.931(104) [17, 18]	0.9
EW	153.6(1.0) [19–23]	8.5
HVP	6845(40) [24–30]	343
HLbL	92(18)[31–45]	154
Total SM Value	116 591 810(43)	368

Table 1.1: Standard Model contributions to a_μ from the 2020 Theory Initiative summary paper [6]. Although the QED contribution is the largest in magnitude, its value is known very precisely. The hadronic contributions are the largest sources of theoretical uncertainty.

expressed by [6]

$$a_\mu^{\text{SM}} = a_\mu^{\text{QED}} + a_\mu^{\text{EW}} + a_\mu^{\text{had}}. \tag{1.20}$$

Here, a_μ^{QED}, a_μ^{EW}, and a_μ^{had} refer to the quantum electrodynamics (QED), electroweak (EW), and hadronic contributions, respectively. The hadronic contribution a_μ^{had} is further expressed as

$$a_\mu^{\text{had}} = a_\mu^{\text{had,HVP}} + a_\mu^{\text{had,LbL}}, \tag{1.21}$$

where $a_\mu^{\text{had,HVP}}$ is the hadronic vacuum polarization (HVP) and $a_\mu^{\text{had,LbL}}$ is the hadronic light-by-light (HLbL). Representative Feynman diagrams for each of these 4 components are shown in Figure 1.3, where the lowest-order contributions to a_μ^{QED} (Schwinger term), $a_\mu^{\text{had,HVP}}$, $a_\mu^{\text{had,LbL}}$, and a_μ^{EW} are shown, respectively.

Each contribution has an associated uncertainty, as given in Table 1.1. The QED term is by far the largest in magnitude, but it is known very precisely. The predominant uncertainties arise from the hadronic terms, since they cannot be calculated perturbatively. Currently, the total SM uncertainty for a_μ is 368 ppb [6].

Recent theory updates from lattice quantum chromodynamics (QCD) calculations are not included in

this table, nor in the comparison between the theoretical and experimental values. They are discussed in more detail in subsection 1.4.3.

1.4.1 Quantum electrodynamics contribution

The QED contribution is calculated from Feynman diagrams with photons and leptons only. Since a_μ is dimensionless, a_μ^{QED} can be written in terms of the ratios of lepton masses [6]:

$$a_\mu^{\text{QED}} = A_1 + A_2(m_\mu/m_e) + A_2(m_\mu/m_\tau) + A_3(m_\mu/m_e, m_\mu/m_\tau), \tag{1.22}$$

where m_μ, m_e, and m_τ are the muon, electron, and tau masses, respectively. The first constant A_1 is then mass-independent and independent of lepton type. Each A_i (where $i = 1, 2, 3$) in the previous equation can be expanded in terms of the fine-structure constant $\alpha = 1/137.035...$:

$$A_i = \left(\frac{\alpha}{\pi}\right) A_i^{(2)} + \left(\frac{\alpha}{\pi}\right)^2 A_i^{(4)} + \left(\frac{\alpha}{\pi}\right)^3 A_i^{(6)} + ... \tag{1.23}$$

The Schwinger term, $A_1^{(2)}$, is 1/2 [14]. The constants A_i^{2n} have been calculated up to 8th order by several groups and up to 10th order by one [17].

Summing the QED contributions gives [17, 18]

$$a_\mu^{QED}(\alpha(\text{Cs})) = 116\,584\,718.931(104) \times 10^{-11} \tag{1.24}$$

or

$$a_\mu^{QED}(\alpha(a_e)) = 116\,584\,718.842(106) \times 10^{-11}, \tag{1.25}$$

depending on whether cesium atom experiments or the anomalous magnetic moment of the electron (a_e) is used to determine the fine-structure constant. The primary sources of uncertainty come from estimates of 12th order QED and the fine-structure constant measurements. However, the uncertainty of < 1 ppb is irrelevantly small compared to the other contributions to a_μ^{SM}.

1.4.2 Electroweak contribution

Feynman diagrams with W, Z, or Higgs bosons are contained in a_μ^{QED}. These diagrams are suppressed by the large boson masses compared to the muon mass. The one loop contribution is proportional to [6]

$$a_\mu^{\text{EW}(1)} \propto \frac{\alpha}{4\pi s_W^2}\frac{m_\mu^2}{M_W^2}, \tag{1.26}$$

where m_μ is the muon mass and M_W the W boson mass; the factor of $\frac{m_\mu^2}{M_W^2}$ means that this contribution is small. The final result for the EW contribution is [19–23]

$$a_\mu^{\text{EW}} = 153.6(1.0) \times 10^{-11}. \tag{1.27}$$

The 1×10^{-11} arises from uncertainties in the hadronic contributions to a_μ^{EW}. This uncertainty is significantly smaller than the uncertainty on the hadronic contributions.

1.4.3 Hadronic contributions

The hadronic contributions include diagrams that contain hadrons and do not contain weak bosons. Because of quark confinement, and because the strong coupling constant is large at low energies, the hadronic contribution cannot be calculated perturbatively in the low-energy region, resulting in higher theoretical uncertainties. Until recently, the hadronic contributions could not be calculated to high precision from first principles (i.e., using lattice QCD), so data-driven dispersive calculations were required. The value of a_μ^{SM} published in the 2020 theory White Paper [6] incorporates only these dispersive calculations. In 2021, the first lattice HVP result competitive with dispersive uncertainties was published [47]. Figure 1.4 shows the hadronic light-by-light (HLbL) and hadronic vacuum polarization (HVP) calculations from different groups on the left and right, respectively. Red squares show data-driven results [28, 29, 31, 46]. Blue circles indicate lattice results that were considered for the official SM value in the 2020 White Paper [6] [36, 47–52]. Open circles indicate lattice results that were not included in the 2020 White Paper [47, 53]. The shaded gray region indicates the official SM value, whereas the shaded orange region indicates the experimental average from [1, 2].

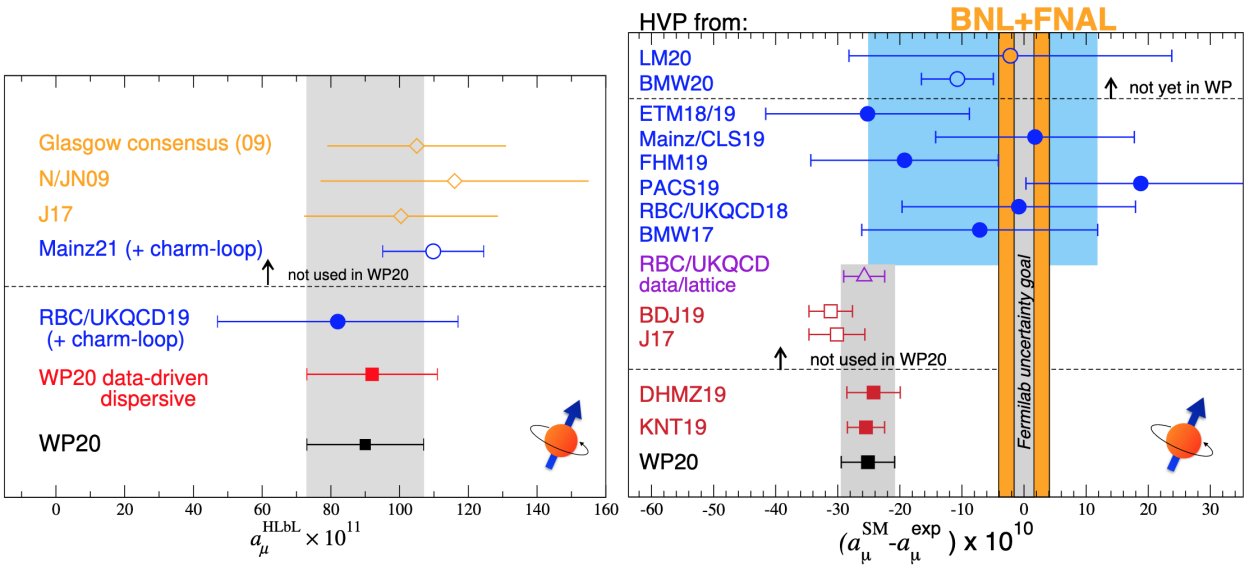


Figure 1.4: Comparison of various data-driven and lattice QCD calculations for the hadronic light-by-light (left) and the hadronic vacuum polarization contribution (right). Red squares show data-driven results [28, 29, 31, 46]. Blue circles indicate lattice results that were considered for the official SM value in the 2020 White Paper [6, 36, 47–52]. Open circles indicate lattice results that were not included in the 2020 White Paper [47, 53]. The shaded gray region indicates the official SM value, whereas the shaded orange region indicates the experimental average from [1, 2]. Reproduced from [54].

Hadronic vacuum polarization contribution

The leading-order HVP contribution is shown in the third diagram in Figure 1.3, where the green circle represents a hadronic blob. The HVP contribution is the largest source of theoretical uncertainty in a_μ^{SM}. To calculate HVP using the data-driven approach, HVP integrals can be written in terms of the cross section of a virtual photon decaying into hadrons, which can be determined from data [6]. At leading order, the dispersion integral is

$$a_\mu^{\text{HVP, LO}} = \frac{\alpha^2}{3\pi^2} \int_{M_\pi^2}^{\infty} \frac{K(s)}{s} R(s) ds, \tag{1.28}$$

where $K(s)$ is the kernel function and $R(s)$ is the hadronic R-ratio

$$R(s) = \frac{\sigma^0(e^+e^- \to \text{hadrons}(+\gamma))}{(4\pi\alpha^2)/3s}, \tag{1.29}$$

where $\sigma^0(e^+e^- \to \text{hadrons}(+\gamma))$ is the "bare" (excluding vacuum polarization effects) cross section of e^+e^- annihilation into hadrons and the denominator is the pointlike muon-pair cross section at the center

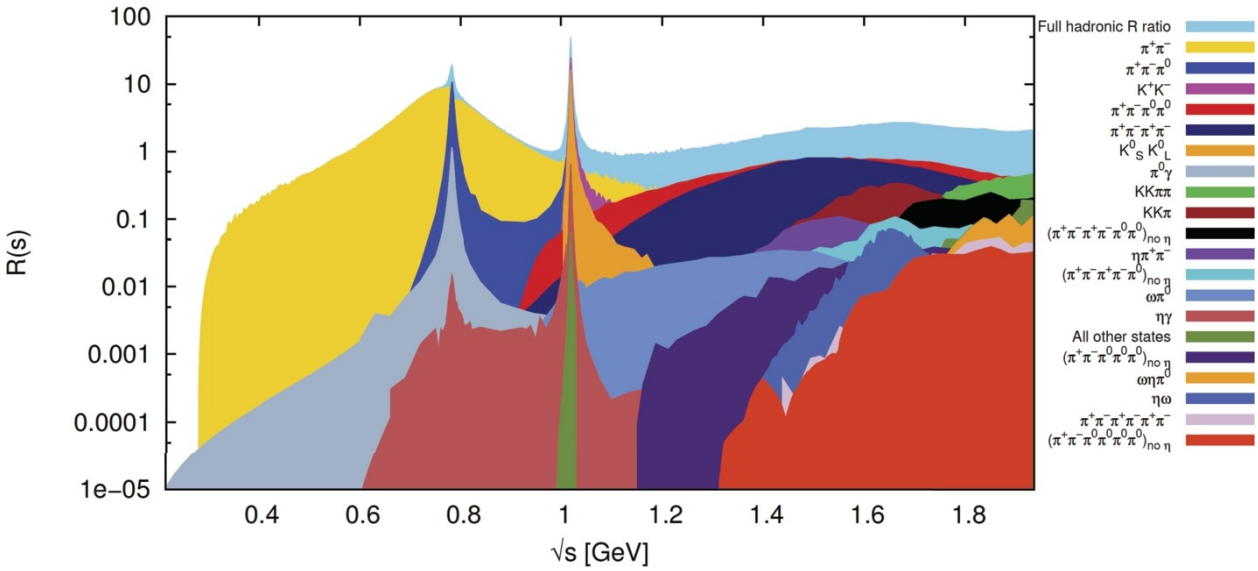


Figure 1.5: Contributions to the total hadronic cross section, expressed as the R-ratio $R(s)$. The light blue shows the total R-ratio. Each final state is shown as a new layer in decreasing order of contribution to $R(s)$. Reproduced from [25].

of mass energy $\sqrt{s}$ [6, 15]. Because the kernel function $K(s) \sim 1/s$, the integral contributes more at lower energy [15]. $R(s)$ as measured in data is shown in Figure 1.5. The total contribution is given in light blue, and the final states are included as new layers in decreasing size of the contribution to $a_\mu^{\text{HVP, LO}}$ [25]. One major challenge in the data-driven calculation is tension in the underlying data, specifically between BaBar and KLOE data for the 2π channel, which accounts for more than 70% of the total HVP. New data may resolve this discrepancy; for the moment, it is accounted for in the uncertainty calculation [6].

The current official value of the HVP contribution to a_μ is [24–30]

$$a_\mu^{\text{had, HVP}} = 6845(40) \times 10^{-11}, \tag{1.30}$$

meaning that HVP is the largest contribution to the total SM calculation uncertainty.

It is possible to calculate the HVP contribution using ab-initio lattice QCD calculations. In the past, the uncertainties on the lattice calculations were significantly larger than the uncertainties on the data-driven calculations, and therefore the White Paper value for $a_\mu^{\text{had, HVP}}$ was based on the data-driven results. The BMW20 result for the HVP, shown in Figure 1.4 in the right figure, was the first lattice QCD result with uncertainties competitive with the data-driven approach [47]. If only this result is considered, the tension

with the experimental value is only 1.5σ, as compared to 4.2σ when using the White Paper average. It is therefore essential to resolve the discrepancy between the data-driven and lattice calculations for HVP. This discrepancy is 2.1σ [54]; however, it is larger (3.7σ) in the intermediate Euclidean time window, which can be calculated more precisely in lattice calculations than other windows [47, 54]. This issue is under intense investigation, with further lattice results and additional comparisons in intermediate windows to come.

Hadronic light-by-light contribution

The hadronic light-by-light (HLbL) contribution refers to graphs like the third from the right in Figure 1.3. Similarly to HVP, it cannot be calculated perturbatively [6]. In fact, it is actually more complicated than HVP because a four point function enters HLbL instead of a two point function. Traditionally, the HLbL contribution has been considered the most difficult to evaluate. Unlike HVP, HLbL cannot be easily expressed as a single integral like Equation 1.28. However, it is suppressed by an additional factor of the fine-structure constant relative to HVP, so the calculation does not need to be as precise.

In the past, model-dependent calculations were performed to estimate HLbL. Significant improvements have been made in the past decades so that the dispersive calculation is now more model-independent; specifically, the contributions from one- and two-particle intermediate states can now be written in a model-independent way [6, 55, 56].

HLbL may also be calculated on the lattice [44, 57, 58]. Currently, the data-driven result and all lattice results are in good agreement, as shown in Figure 1.4 (left). The official White Paper value for the HLbL contribution is [31–45]

$$a_\mu^{\text{had, LbL}} = 92(18). \tag{1.31}$$

1.5 Beyond the Standard Model implications

If the difference between a_μ^{exp} and a_μ^{SM} holds, and the difference reaches the 5σ threshold, physics beyond the Standard Model (BSM) would be required to explain the discrepancy $\Delta a_\mu = a_\mu^{\text{exp}} - a_\mu^{\text{SM}}$. Notably, Δa_μ is larger than twice the EW contribution to a_μ, constraining the possibilities for BSM physics. Additionally, many constraints have been placed on possible explanations for Δa_μ from other experiments. Reference [59] gives an extremely detailed and up-to-date overview of possible models in light of the FNAL $g-2$

result and other recent data. Many BSM particles also effect the muon mass, as suggested by Equation 1.19, and in order to not exceed the measured muon mass, the mass scale of the new physics is expected to be approximately [59]

$$M_{\text{BSM}} \lesssim \mathcal{O}\,(2.1)\text{TeV}. \tag{1.32}$$

Much of this parameter space has been ruled out by measurements at the Large Hadron Collider (LHC) and other experiments. This has caused tension with models that originally seemed like the most obvious explanation, such as many supersymmetry (SUSY) models [16, 59]. Dark matter experiments also provide additional constraints on possible models; if a dark matter candidate that can explain Δa_μ is overabundant compared to the dark matter density in the universe then it is strongly excluded [59]. Direct dark matter searches may also exclude new physics candidates.

As an example of BSM physics, a dark photon, called γ', could arise from an additional U(1) gauge symmetry. The dark photon would be a massive gauge boson and would mix with the standard SM photon γ. However, the two dimensional parameter space for the mass of γ' and the kinetic mixing parameter ϵ that could account for Δa_μ has now been fully ruled out. Other excluded models include dark Z, many simple two Higgs doublet (2HDM) models, and many leptoquark models [59]. Three-field models, which have two fields of the same spin that mix together and a third field with a different spin, are the least constrained by existing data. This includes both 2S1F (two scalar one fermion) and 2F1S (two fermions one scalar) models, although the required coupling constants are large. A small amount of parameter space for the 2HDM and leptoquark models remain. Some MSSM (Minimal Supersymmetric Standard Model) options are also viable, although these scenarios may not fully explain the dark matter abundance. Overall, the measurement of a_μ creates strong constraints on BSM physics and has ruled out many models, necessitating theorists to think outside of the box in order to explain the discrepancy. Extremely strong interest remains in the possibility that the Muon $g-2$ discrepancy is evidence of new physics.

Chapter 2

Fermilab Muon $g-2$ experiment

2.1 Measurement principle

The principle of the Muon $g-2$ experiment is to store polarized muons in a magnetic storage ring and measure their decays into positrons as a function of time. Because of the parity violating nature of the weak interaction, the positrons contain information about the spin direction of the muon at the time of decay, allowing for the muon's g-factor to be measured. In a uniform dipole magnetic field, the muons undergo both spin precession and cyclotron motion, with frequencies ω_s and ω_c, respectively. If the field is perfectly uniform and there is no electric field, and if the muons are traveling perpendicular to the magnetic field, the muon spin will precess according to:

$$\vec{\omega}_s = -\frac{ge\vec{B}}{2m} - (1-\gamma)\frac{e\vec{B}}{m\gamma}. \tag{2.1}$$

The muons will travel around the magnetic storage ring at the frequency:

$$\vec{\omega}_c = -\frac{e\vec{B}}{m\gamma}, \tag{2.2}$$

where γ is the Lorentz factor. The second term in Equation 2.1 is the relativistic contribution from Thomas precession. The anomalous precession frequency is defined as

$$\omega_a \equiv \omega_s - \omega_c, \tag{2.3}$$

the difference between the spin and cyclotron precession frequencies [7]. Plugging in Equation 2.1 and Equation 2.2, gives:

$$\vec{\omega}_a = -\left(\frac{g-2}{2}\right)\frac{e\vec{B}}{m} = a_\mu \frac{e\vec{B}}{m}, \tag{2.4}$$

where the anomalous magnetic moment of the muon a_μ is defined in Equation 1.14. The anomalous precession frequency $\vec{\omega}_a$ therefore depends only on the *anomalous* magnetic moment (the difference of g from 2), and not on g itself. Without the SM loop corrections discussed in the previous section, $g = 2$, and $\vec{\omega}_a$ would be 0; the spin and momentum vectors of the stored muons would not precess with respect to each other. However, since $g - 2 > 0$, $\vec{\omega}_a \neq 0$. By measuring $\vec{\omega}_a$ and $\vec{B}$, a_μ can be measured.

The magnetic storage ring provides a horizontal restoring force on the muons, but there is no restoring force in the vertical direction. In order to vertically focus the muon beam, an electric quadrupole field is required. A simulation of the quadrupole field in the storage region is shown in Figure 2.1 (left). Also, muons will have vertical oscillations and therefore do not always travel perpendicular to the magnetic field. This means that the assumptions made for Equation 2.4 are not fully correct. The full equation for ω_a including all terms (assuming that the muon's electric dipole moment is zero) is

$$\omega_a = \frac{e}{m}\left[a_\mu \vec{B} - a_\mu\left(\frac{\gamma}{\gamma+1}\right)\left(\vec{\beta}\cdot\vec{B}\right)\vec{\beta} - \left(a_\mu - \frac{1}{\gamma^2-1}\right)\frac{\vec{\beta}\times\vec{E}}{c}\right], \tag{2.5}$$

where γ is the Lorentz factor and $\vec{\beta} = \vec{v}/c$. The first term is the simple form shown in Equation 2.4. The second term in general is nonzero if muons move parallel to the magnetic field, $\vec{\beta}\cdot\vec{B} \neq 0$. The small component of the momentum parallel to the magnetic field necessitates the "pitch" correction, discussed in subsection 3.2.4. The third term gives the contribution from the motional magnetic field experienced by a relativistic particle. This term is nonzero because the electric field used to focus the muons means that $\vec{E} \neq 0$. However, by inspection, if $\gamma = \sqrt{1 + 1/a_\mu} \approx 29.3$, this third term will cancel. The momentum

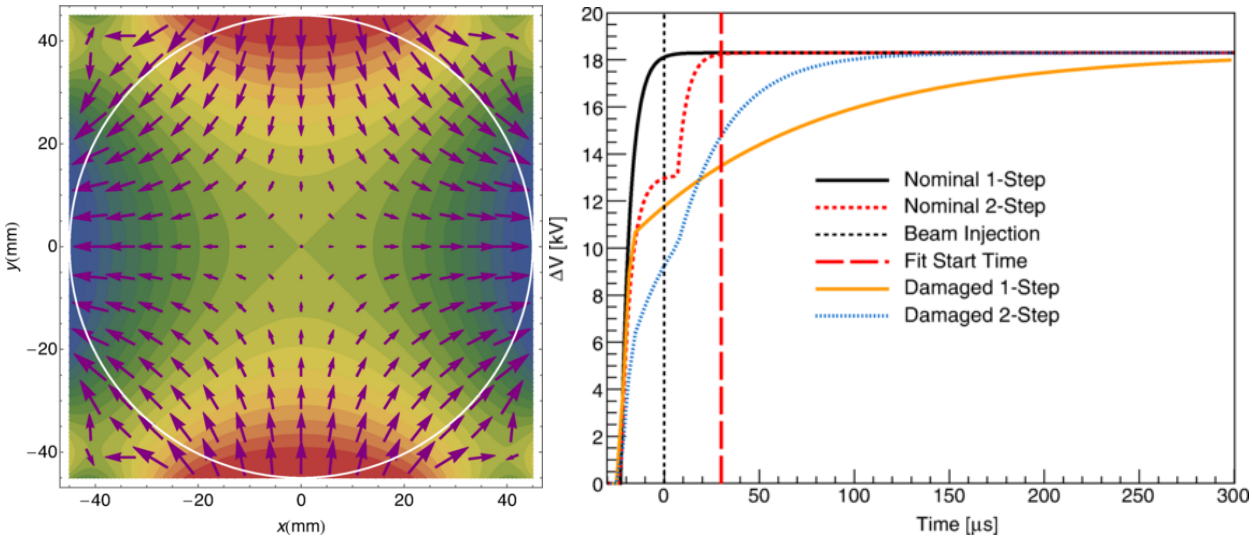


Figure 2.1: Left: simulated electrostatic quadrupole field with respect to the storage region shown as a white circle. Right: ESQ voltage as a function of time in fill for the damaged resistors in Run-1 and the nominal resistors. The fit start time is indicated with a vertical dashed line. Reproduced from [4].

that cancels the third term, $p_0 \approx 3.094$ GeV, is the so-called "magic momentum" chosen for the experiment. Operating at this momentum allows for the experiment to be performed. However, because of the finite momentum distribution of the beam, not all muons are exactly at the magic momentum, so an "electric field" correction, discussed in subsection 3.2.3, is required due to this term.

The second component of the measurement is the magnetic field B. In order to measure B, the Larmor precession frequency of the free proton, ω_p is measured. It is related to B by

$$B = \frac{\hbar\omega_p}{2\mu_p}, \tag{2.6}$$

where μ_p is the proton's magnetic dipole moment (MDM). Further, the electron charge e, can be expressed as

$$e = \frac{4m_e\mu_e}{\hbar g_e}, \tag{2.7}$$

where m_e, μ_e, and g_e are the mass, MDM, and g-factor of the electron, respectively. Substituting Equation 2.6 and Equation 2.7 into Equation 2.4 gives

$$a_\mu = \frac{\omega_a}{\tilde{\omega}_p}\frac{g_e}{2}\frac{\mu_p}{\mu_e}\frac{m_\mu}{m_e}. \tag{2.8}$$

Quantity	Source	Uncertainty [ppb]
g_e	Quantum electron cyclotron	0.00026 [60]
μ_p/μ_e	Hydrogen spectroscopy	3.0 [61]
m_μ/m_e	Muonium hyperfine splitting	22 [62]
$\omega_a/\tilde{\omega}_p$	Fermilab Muon $g-2$	140 (goal) [7]

Table 2.1: Contributions to a_μ including experimental source and uncertainty. All uncertainties are small compared to the uncertainty of the Fermilab Muon $g-2$ experiment.

Note that $\tilde{\omega}_p$, which replaces ω_p, is a weighted average of of ω_p with the the muon spatial distribution within the storage region. The equation for a_μ is now in terms of two frequencies, ω_a and $\tilde{\omega}_p$, and other quantities that are known to high precision from other experiments. The g-factor of the electron g_e is known from one-electron quantum cyclotron experiments [60]; the ratio of the proton to electron MDMs μ_p/μ_e from hydrogen spectroscopy [61]; and the ratio of the mass of the muon to the mass of the electron m_μ/m_e from muonium hyperfine splitting [62], as summarized in Table 2.1. Each of these quantities is known to 22 ppb or better, compared to the 140 ppb uncertainty goal of the Muon $g-2$ experiment. Measurement of ω_a and $\tilde{\omega}_p$ forms the foundation of the experimental procedure.

2.1.1 Measuring the precession frequency

The measurement of ω_a is reliant on the parity violating nature of the muon's decay into positrons, the predominant decay mode of the muon [7]. The positive muon decays into a positron and two neutrinos through the weak interaction, as shown in Figure 2.2 (left). In Figure 2.2 (right), a schematic of the decay producing the highest energy positron in the muon's rest frame is shown. The ν_e is left-handed (momentum and spin anti-aligned) and the $\bar{\nu}_\mu$ is right-handed (momentum and spin aligned). If the muon's spin is pointing up, the highest energy decay positron will be emitted back-to-back with the two neutrinos, meaning that in order to conserve angular momentum, the positron will be emitted in the direction of the muon spin. As a result, the decay is "self-analyzing," meaning that the decay positrons carry information about the spin state of the original muon. Muons also have a relatively long muon lifetime of $\approx 2.2\ \mu s$, allowing them to be produced in a beamline, and stored in a storage ring. These facts allow a_μ to be measured by observing the decay of muons into positrons.

With the approximation that the positron has energy $E >> m_e c^2$, the differential decay distribution in

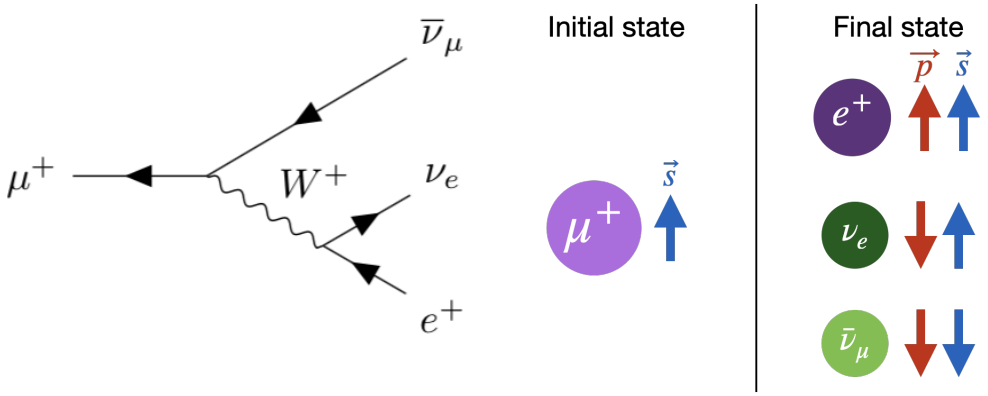


Figure 2.2: Left: Feynman diagram for μ^+ decaying into positrons and neutrinos. Right: Schematic of the muon decaying into the highest energy positron in the muon rest frame. Because of the parity violating nature of the weak interaction, the highest energy decay positrons will have momenta aligned with the spin of the original muon. The red arrows indicate the particle's momentum and the blue arrows its spin.

the muon rest frame is given by [7, 63]

$$\frac{d^2P}{dE\, d\cos(\theta)} = N_r(E)[1 + A_r(E)\cos(\theta)], \tag{2.9}$$

where θ is the angle between the muon spin and positron momentum and the subscript r indicates the rest frame of the muon. The decay asymmetry A_r indicates the degree of correlation between the muon spin and positron momentum. A_r increases with the decay positron energy, approaching 1 as the positron energy approaches its maximum possible energy E_{max}, matching the schematic described above. Figure 2.3 (left) shows the differential decay distribution as a function of $\hat{s}_{\mu^+} \cdot \hat{p}_{e^+}$, the correlation between the muon spin and the positron momentum. The figure on the right shows A_r as a function of E/E_{max}, where E is the decay positron energy. It additionally shows the number density of positrons N_r increasing as a function of E/E_{max}.

The decay distribution in the rest frame of the muons is then boosted into the lab frame. The muons also undergo exponential decay at the boosted muon lifetime $\gamma\tau_\mu$. If the angle between the spin and momentum vectors of the muon in the lab frame is α, then the positron number measured in the lab frame is

$$N(t, E) = N(E)e^{-t/\gamma\tau_\mu}[1 + A(E)\cos(\alpha)], \tag{2.10}$$

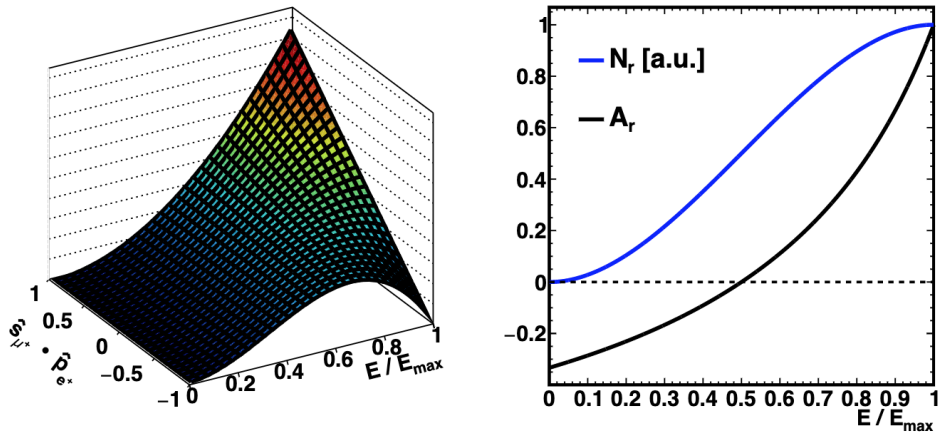


Figure 2.3: Left: differential decay as a function of $\hat{s}_{\mu^+} \cdot \hat{p}_{e^+}$, the correlation between the muon spin and the positron momentum, showing that the highest energy positrons are preferentially emitted in the direction of the muon spin. Right: asymmetry and number in the rest frame of the muon, showing that A_r approaches 1 at the maximum energy. Reproduced from [63].

where N and A are the lab frame values distinct from N_r and A_r. As the muon's spin and momentum vectors precess with respect to each other at ω_a, the number of positrons above a certain energy oscillates at ω_a. Figure 2.4 (left) shows the decay positron energy spectrum in the lab frame for three different α. When the vectors are parallel, more decay positrons above an energy threshold are measured (indicated by the shaded blue region). When the vectors are antiparallel, fewer decay positrons above an energy threshold are measured (indicated by the shaded black region).

If the precession frequency $\vec{\omega}_a$ is constant in time, the angle α between the muon spin and its momentum is $\omega_a t - \phi$, and the positron number measured in the lab frame is then

$$N(t, E) = N(E) e^{-t/\gamma\tau_\mu} [1 + A(E) \cos(\omega_a t - \phi(E))]. \tag{2.11}$$

The number of positrons above an energy threshold oscillates at ω_a, the anomalous precession frequency. By fitting Equation 2.11 to a histogram of positrons above an energy threshold, a measurement of ω_a is made. For a given histogram of positrons above an energy threshold as a function of time, the basic fit function is

$$N(t) = N_0 e^{-t/\gamma\tau_\mu} [1 + A \cos(\omega_a t - \phi_0)]. \tag{2.12}$$

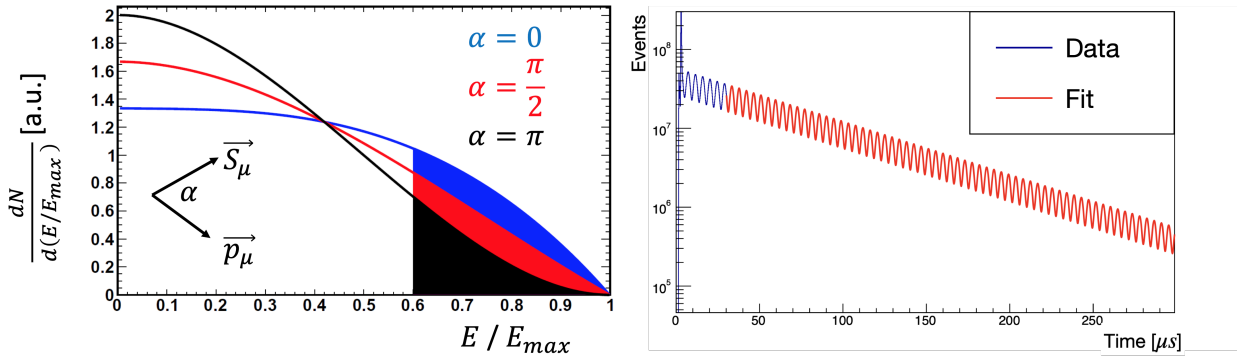


Figure 2.4: Positron energy spectrum for three different angles α between the spin and momentum vectors of the muon. When the spin and momentum are aligned ($\alpha = 0$), more positrons are measured at high energies (shaded in blue). When the spin and momentum are anti-aligned ($\alpha = \pi$), fewer positrons are measured at high energies (highlighted in black). If $\alpha = \omega_a t - \phi$, measurement of the positrons above an energy threshold allows for the anomalous precession frequency ω_a. Reproduced from [63]. Right: measured positrons above an energy threshold as a function of time after injection. A five parameter fit has been applied to the data. The visible oscillation is at the anomalous precession frequency ω_a.

A sample histogram of positrons above energy threshold vs. time is shown in Figure 2.4 (right), with the fit function shown in red. The details and complications of this fitting procedure form the primary content of this thesis.

2.1.2 Measuring the magnetic field

The determination of a_μ also relies on a high-precision measurement of the magnetic field B experienced by the muons. The measurement of B is performed using nuclear magnetic resonance (NMR). A $\pi/2$ radio frequency (RF) pulse is used to tip a proton's spin, and the resulting free induction decay (FID) signal is detected. A sample FID signal, from which the Larmor precession frequency of the proton ω_p is extracted, is shown in Figure 2.5.

The goal of the magnetic field analysis is to produce $\tilde{\omega}'_p(T_r)$, the proton spin precession frequency in a spherical water sample at 34.7 $^\circ C$, weighted by the muon position in the storage region. In order to measure the field, a trolley containing 17 nuclear magnetic resonance (NMR) probes drives around the storage ring and takes periodic measurements of the magnetic field. The field is measured in terms of moments m_i. For an azimuthal slice of the ring, the field B can be expressed by the source-free Laplace equation in polar

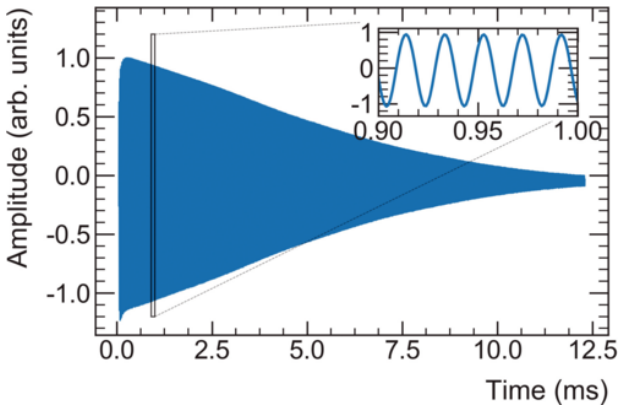


Figure 2.5: Sample free induction decay (FID) signal from a trolley probe. The inset shows the oscillation at the Larmor precession frequency of the proton, ω_p. Reproduced from [5].

coordinates (r, θ) [5]:

$$B \approx B_y = A_0 + \sum_{n=1} \left(\frac{r}{r_0}\right)^n [A_n \cos(n\theta) + B_n \sin(n\theta)], \tag{2.13}$$

where the approximation has been made that the field only has a vertical component, $|\vec{B}| = B_y$. Here, r is the radius from the center of the muon orbit and $r_0 = 4.5$ cm is a normalization to the outer edge of the storage region. The constants A_n and B_n are the normal and skew multipoles of the field, so for example, the normal dipole $m_1 = A_0$ and the normal quadrupole $m_2 = A_1 \frac{r}{r_0} \cos(\theta)$. A sample azimuthally averaged field map from the trolley is shown in Figure 2.6 (left). Variations in ppm are given with respect to the central trolley probe value. From this data, the field multipoles are extracted. The multipoles as a function of azimuth during a single trolley run are shown in Figure 2.6 (right).

The trolley can only be in the ring when no beam is present, so fixed probes above and below the storage region are used to interpolate the field between trolley runs. A sample fixed probe interpolation between trolley runs is seen in Figure 2.7, with the green and red diamonds marking the times of two trolley runs about 70 hours apart. The orange points are the uncorrected fixed probe interpolation, and the blue points are the corrected interpolation after the data has been backwards fit from the second trolley run [5].

The field measurement is written as $\omega_p(x, y, \phi)$, where ϕ is the azimuthal location around the ring. The

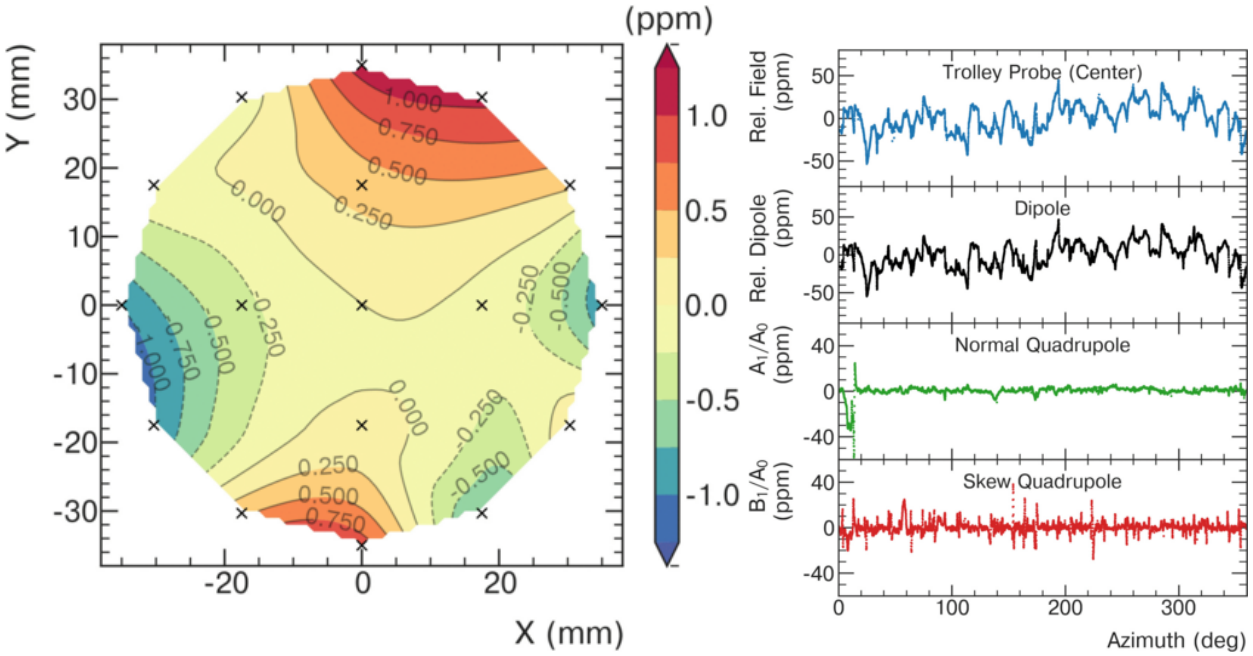


Figure 2.6: Left: azimuthally averaged field map measured by the trolley, with variations in ppm given with respect to the central trolley probe. Right: field multipoles measured during a trolley run as a function of azimuth. Reproduced from [5].

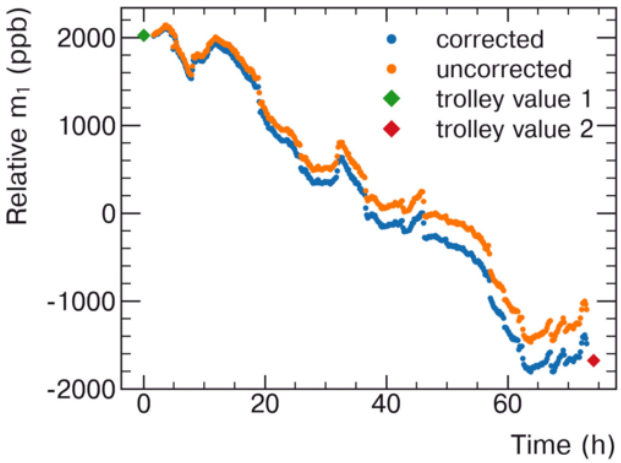


Figure 2.7: Trolley measurements are shown in green and red at two different times, and the fixed probe interpolation between the trolley runs is shown in blue and orange. The blue points show the uncorrected interpolation before they have been backwards corrected from the second trolley run, whereas the orange points have this correction applied. Reproduced from [5].

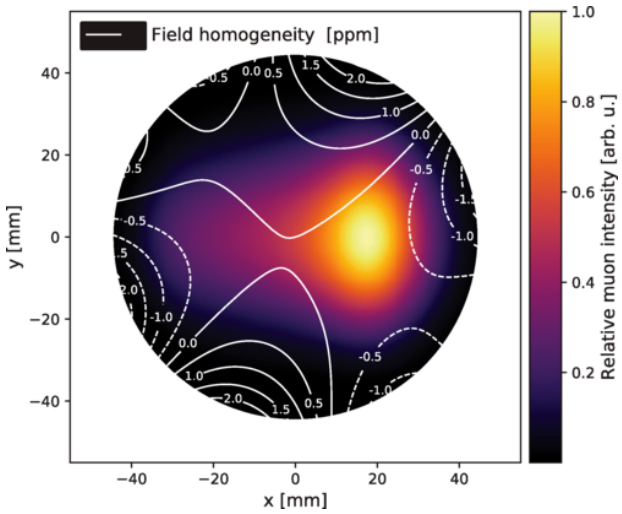


Figure 2.8: Run-1 azimuthally averaged map of ω_p in parts per million relative to the central value overlaid with the beam map measured by the trackers. The weighted average $\tilde{\omega}_p$ enters the equation for a_μ. The beam is off-center in the storage region due to an insufficient magnetic kick. Reproduced from [2].

field measurement is averaged with the position of the muons in the storage region $M(x, y, \phi)$ to give $\tilde{\omega}_p$:

$$\tilde{\omega}_p \equiv \langle \omega_p(x, y, \phi) \times M(x, y, \phi) \rangle. \tag{2.14}$$

This weighted average will enter the final equation for a_μ. This muon distribution $M(x, y, \phi)$ is measured by the tracking detectors, which reconstruct the original location of the muon decay by measuring positron tracks. An azimuthally averaged overlay of the field and beam maps used to produced $\tilde{\omega}_p$ is shown in Figure 2.8. The field homogeneity relative to the value at the center of the storage region is given in ppm, overlaying the beam intensity in arbitrary units. As previously mentioned, the beam is off center in the storage region because of the insufficient kick, meaning that the normal moments have relatively large weights.

2.2 Creating a polarized muon beam

In order to measure the anomalous precession frequency, muons must enter the magnetic storage ring polarized, that is, with their spins generally aligned. The beam is created at the muon campus at Fermilab. First, 8 GeV protons are split into four bunches in the recycler. They are extracted one at a time to hit an Inconel

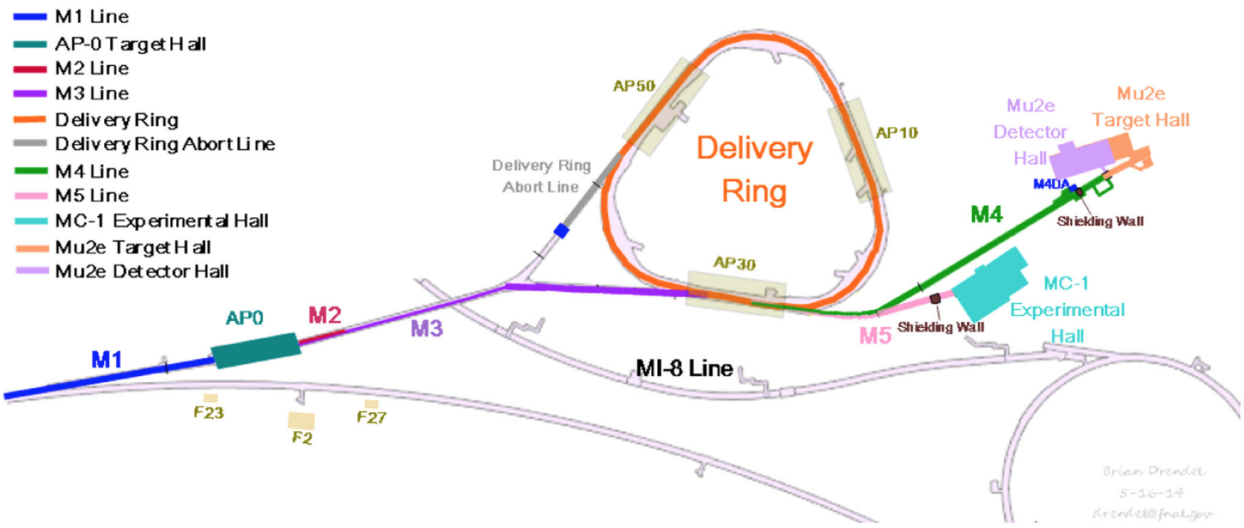


Figure 2.9: The muon campus at Fermilab. Protons hit a target at AP0, producing pions and other particles. Pions with energies around 3.1 GeV are selected using bending magnets in M2 and M3. Pions decay into muons, and the remaining protons are removed in the delivery ring. The polarized muon beam finally arrives at the Muon $g-2$ experimental hall, MC-1. Reproduced from [64].

target in AP0, shown in Figure 2.9 [7, 64], to produce pions. The secondary beam is then focused using a lithium lens, and positively charged particles with a momentum of 3.11 GeV $\pm$ 10% are selected using bending magnets.

Pions decay into muons through the weak interaction, meaning that a polarized muon beam can be easily produced. Pions decay into

$$\pi^+ \rightarrow \mu^+ + \nu_\mu, \tag{2.15}$$

which is the primary decay mode of the positive pion [7, 15]. In the rest frame of the pion, the neutrino and muon must be emitted back-to-back. Because the neutrino is always left-handed in the SM and the spin of the pion is 0, the muon must also have a left-handed helicity state. In the lab frame, this results in a correlation between the decay muon phase and its momentum. By selecting the highest-energy muons resulting from the pion decay in the M2 and M3 beam lines, a polarization of $\approx 95\%$ can be achieved [7].

The beam is then injected into the delivery ring (DR), shown in orange. Heavier protons gradually fall behind lighter muons, so the protons can be kicked out. The muons then travel down the M4 and M5 lines to the Muon $g-2$ storage ring.

The beam is injected in sets of pulses, called "fills." The time structure of the beam injection is shown

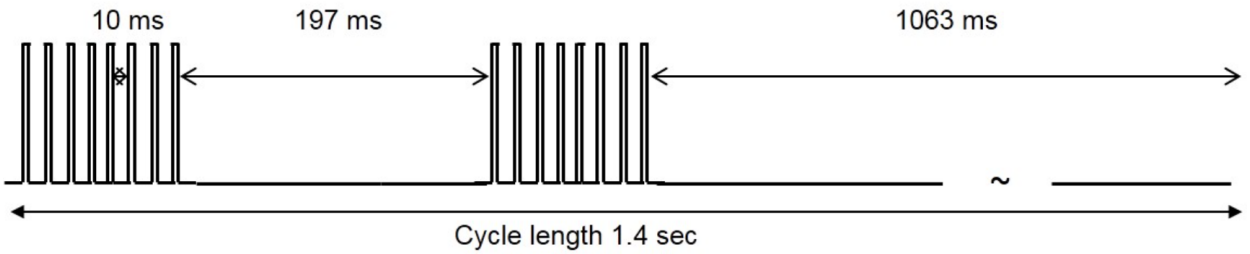


Figure 2.10: Time structure of beam pulses or "fills" arriving in the storage ring. The muons will have decayed away by the time the next pulse is injected. Reproduced from [7].

in Figure 2.10. The muons will have decayed away by the time the next muon bunch arrives. The beam additionally arrives in sets of 8 pulses, each of which may have a different shape because they are produced separately by the accelerator complex. The muons will have a large momentum spread of a few percent as they enter the storage ring, but only a small fraction of those will be stored. The stored muons will have a finite momentum distribution, and not all muons will be at exactly the magic momentum.

2.3 Magnetic storage ring

The superconducting magnetic storage ring creates an extremely spatially uniform and temporally stable magnetic field of 1.45 T in order to store the muons. The magnetic storage ring is reused from the Brookhaven Muon $g-2$ experiment after traveling to Fermilab in 2013. A full diagram of the Muon $g-2$ storage ring is shown in Figure 2.11 (left), as seen from above. The beam is injected on the top right of the figure, and the various ring and detector systems are marked. The picture on the right was taken in the experimental hall, and the beam is injected at the top of the image.

A cross section of the ring is shown in Figure 2.12. The muons will be stored in the region indicated in red. The wedge, edge shims, and iron foil laminations were used to shim the field to part per million (ppm) level uniformity. Also visible in the diagram are the fixed NMR probes, which measure the field when muons are present.

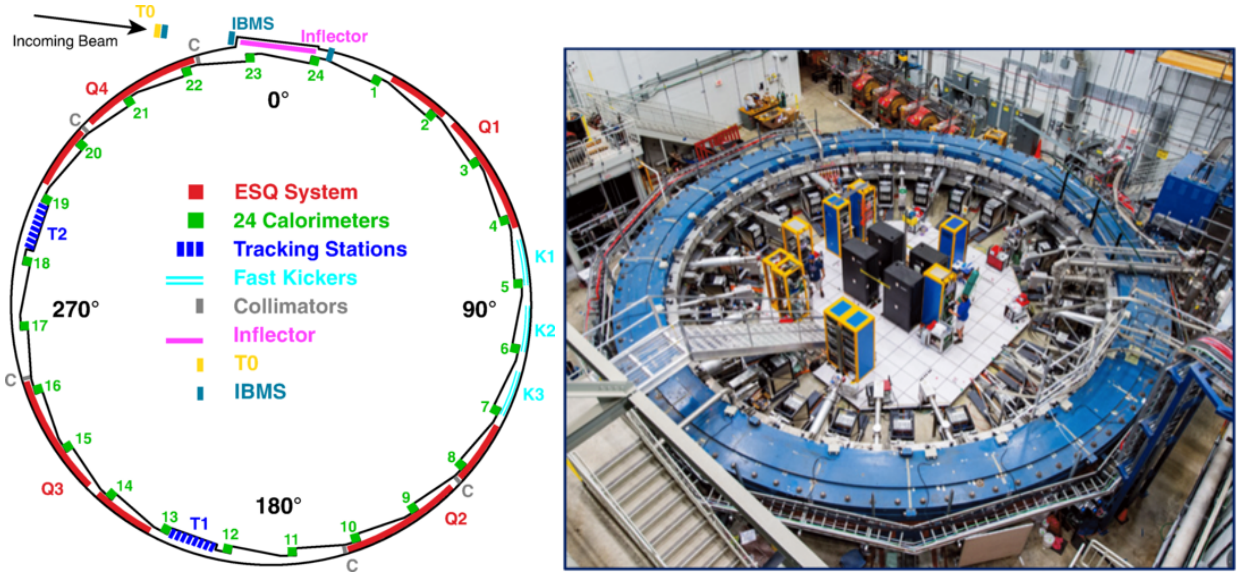


Figure 2.11: Left: diagram of the storage ring from above showing key ring and detector systems. Right: photograph of the storage ring. The muons will be injected through the beam line at the top of the image. Reproduced from [5].

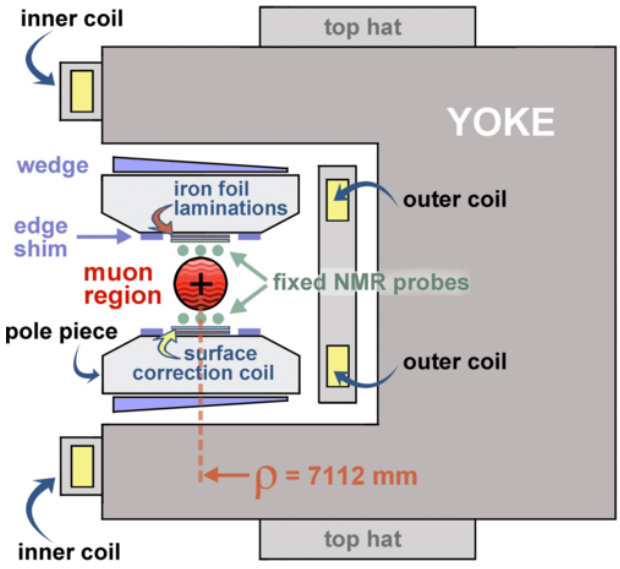


Figure 2.12: Cross section of the magnetic storage ring showing the wedges, shims, and foil laminations used to achieve part per million level field uniformity. The muons are stored in the region indicated in red. Reproduced from [5].

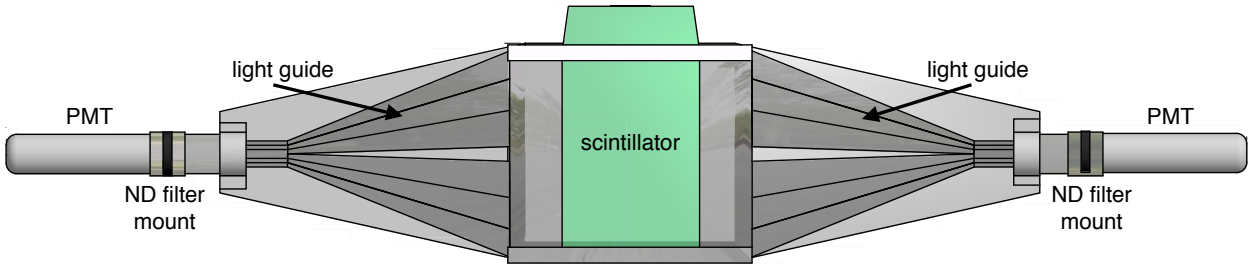


Figure 2.13: Diagram of the T0 detector. Particles entering the ring pass through the scintillator. Light guides transport the light to the PMTs on each end. Neutral density filters control the amount of light entering the PMTs. The detector is placed vertically in the path of the beam.

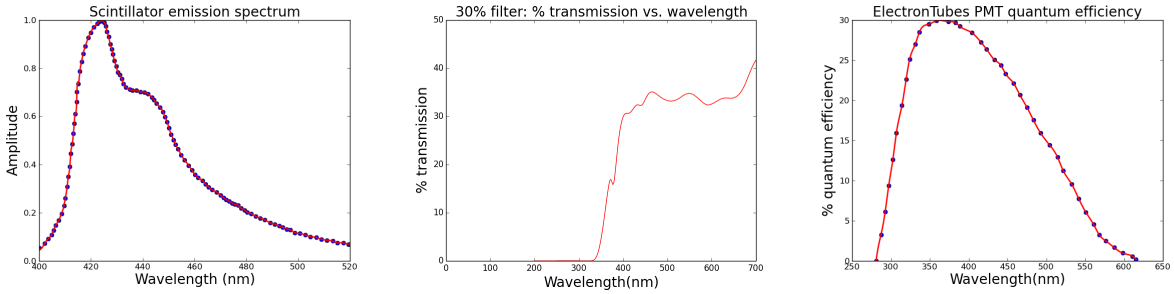


Figure 2.14: Scintillator emission spectrum, sample ND filter transmission curve, and PMT quantum efficiency as a function of wavelength for the T0 start-time detector. These three curves were multiplied to get an effective transmission percentage due to the ND filter.

2.4 Injection

When a polarized muon beam is injected into the storage ring, it first passes through the T0 start time detector, which defines $t = 0$ for the measurement period. The detector consists of a 1mm thick piece of plastic scintillator through which the beam passes, as seen in Figure 2.13. The scintillator is coupled via light guides to two photomultiplier tubes (PMTs), called PMT-A and PMT-B. Between the light guide and the PMT, there is a hinged mechanism for holding neutral density (ND) filters, which limit the amount of light entering the PMTs. A laser fiber is coupled to the filter mounts in order to provide synchronization pulses from the laser system, so that T0 can be synchronized with other detectors in the experiment. The scintillator emission spectrum, sample ND filter transmission curve, and PMT quantum efficiency curve as a function of wavelength for the T0 detector are seen in Figure 2.14. These three curves were multiplied to find the effective transmission percentage of the neutral density filter placed in front of the PMT.

A sample beam pulse in the T0 detector is seen in Figure 2.15, where the time in shown in clock ticks

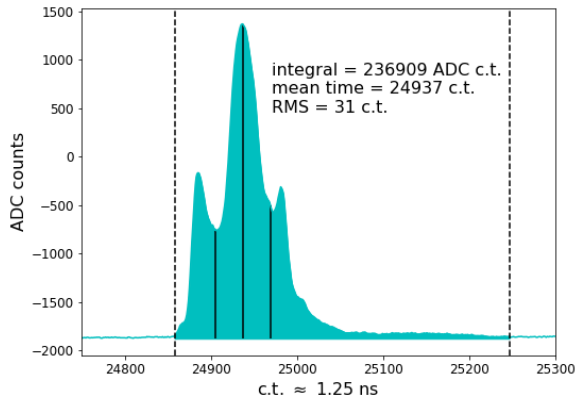


Figure 2.15: Sample T0 start-time detector pulse, showing reconstruction of time, integral, and root mean square (rms). The time shown is a weighted average time of the pulse and determines $t = 0$ for the experiment. The characteristic "W" shape arises from the way the pulse is created in the Fermilab accelerator complex. The values are given in clock ticks (c.t.), where 1 c.t. $\approx$ 1.25 ns. The vertical dashed lines mark the integration period for the T0 pulse, which is fixed for all T0 pulses.

(c.t.), with 1 c.t. $\approx$ 1.25 ns. The characteristic time width of a beam pulse is about 125 ns. The pulse time (calculated from a weighted average), root mean square (rms), and pulse integral (not calibrated) are shown. The vertical dashed lines indicate the integration period for the T0 pulse, which is standard for all pulses. The T0 integral is a measurement of beam intensity and provides a denominator for storage efficiency. The average T0 time determines $t = 0$ for the entire experiment. The rms is a measurement of the time width of the beam, which is later used for data quality control purposes to ensure that the time structure of the beam is not changing. The T0 integral, timing, and pulse shapes are available in real time as data is being taken, making T0 a critical tool for tuning and monitoring experiment systems. The detector is also used to create averaged beam profiles of each of the 8 incoming beam pulses pulses separately, as their shapes are distinct, due to the way they are created in the accelerator complex. The averaged pulse shapes are shown in Figure 2.16. Ultimately, the T0 detector provides the first information on the beam as it is injected into the ring.

After the T0 detector, the beam passes through the inflector beam monitoring system (IBMS), a set of scintillating fiber detectors that measure the horizontal and vertical beam profile. Finally, the beam passes through a superconducting inflector, which cancels out the field of the storage magnet so that the muons can pass through a hole in the magnet into the storage region. A third IBMS detector can be inserted to measure

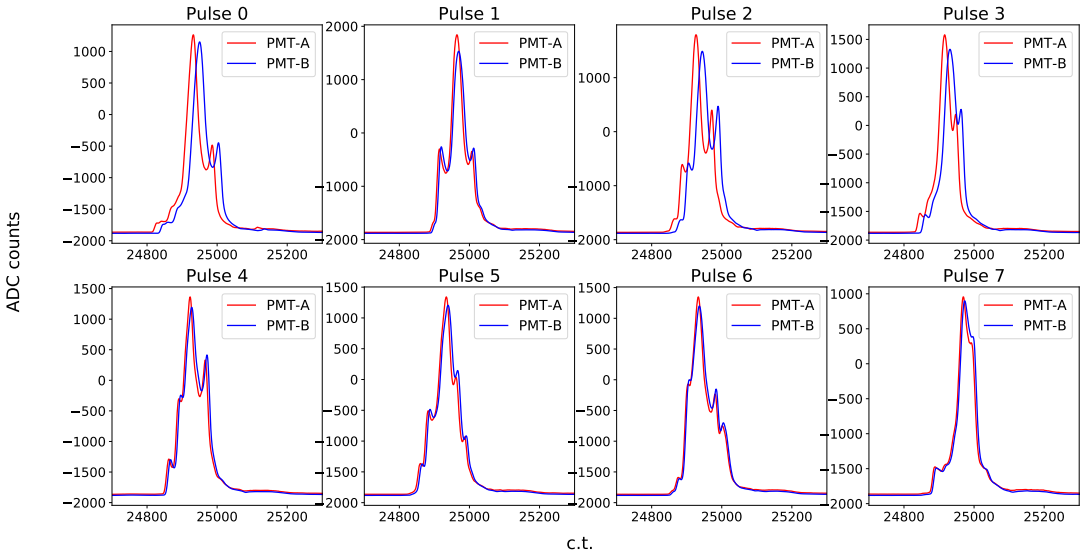


Figure 2.16: Averaged pulse shapes for the two PMTs (PMT-A and PMT-B) for each of the 8 beam pulses in a typical accelerator injection sequence. Traces are aligned in time to a single pulse to correct for timing jitter and averaged over 100 pulses. The horizontal axis is in clock ticks, where 1 c.t. ≈ 1.25 ns.

the beam as it exits the inflector, but it is retracted during normal data taking as it destroys the stored beam over time. The muons are ultimately injected radially outward from the ideal central orbit.

2.5 Beam storage

Because the muons are injected off-axis from the center of the storage region, they must be centered by a magnetic kick. Three magnetic kickers, located in a single section of the ring, deflect muons horizontally onto their proper orbit. A schematic of this procedure can be seen in Figure 2.17 (left), where the location of the kickers is shown in gray. If the muons were not kicked, they would follow their original orbit, the dashed red line, and hit the back of the inflector after orbiting once around the ring. Therefore, the muons need to be kicked on the green orbit in order to be stored for the measurement period. The kicker pulse and the incoming T0 pulse are shown together in Figure 2.17 (right). Because of the relatively wide kicker pulse relative to the T0 pulse, different portions of the incoming beam may experience different kicks. A sufficient kick to center the beam in the storage region could not be achieved in Run-1, Run-2, and part of Run-3 due to hardware challenges. The weaker kick resulted in the beam being off-center on average, with

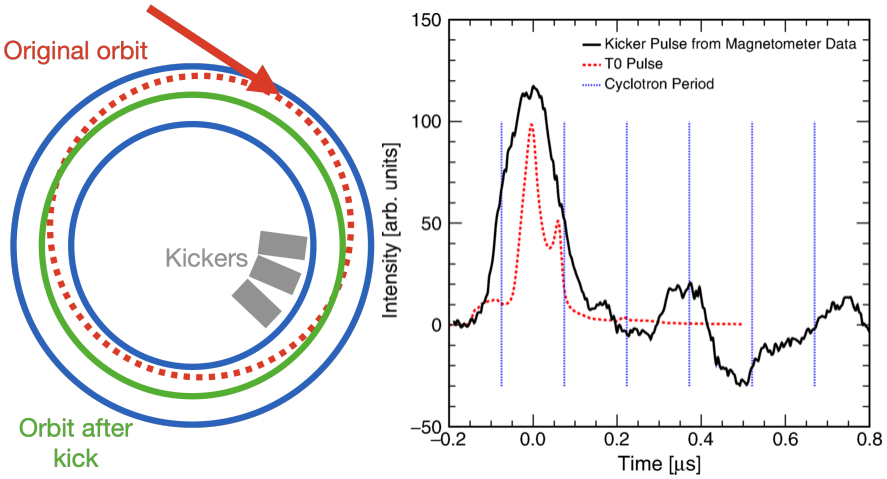


Figure 2.17: Left: Schematic (not to scale) of the result of the magnetic kick on the muon orbit. Muons are injected off-orbit, and would follow the red dashed orbit if not kicked. A magnetic kick centers them on the ideal orbit, in green. Right: kicker pulse overlaid with the T0 pulse, showing their relative widths. Not all muons experience exactly the same kick strength. The cyclotron period of the beam is shown in the vertical blue lines. Reproduced from [4].

a large oscillation around the central orbit. This is the reason for the off-center beam location visible in Figure 2.8. The ideal kick was finally achieved at the end of Run-3.

The magnetic field provides horizontal focusing, but vertical focusing must be provided by electrostatic quadrupoles (ESQs) in order to store muons. The ESQs provide a linear restoring force in the vertical direction at the expense of a slight horizontal defocusing [7]. The simulated quadrupole field is shown in Figure 2.1. The ESQs are located in four locations shown in Figure 2.11, meaning that the focusing electric field is not equal at all points around the ring. These electric fields result in beam dynamics that are accounted for in the precession frequency analysis. Before the fill begins, the ESQ voltage is changed to move the beam so that it is scraped against collimators, reducing the number of muons which might be lost later during the measurement period. The ESQ voltage should be constant when physics data is being taken, so that the beam is not moving. However, in Run-1, there were two damaged ESQ resistors that were not discovered until after data taking, resulting in slower recovery of the voltage. A comparison between the voltage of nominal and damaged resistors as a function of time is shown in Figure 2.1 (right). Because of the voltage recovery after 30 μs, the beam position and shape were still changing at the start of the measurement period, resulting in a systematic uncertainty discussed in subsection 3.2.6. This issue was fixed before Run-2 and is not relevant for the Run-2 and Run-3 analysis.

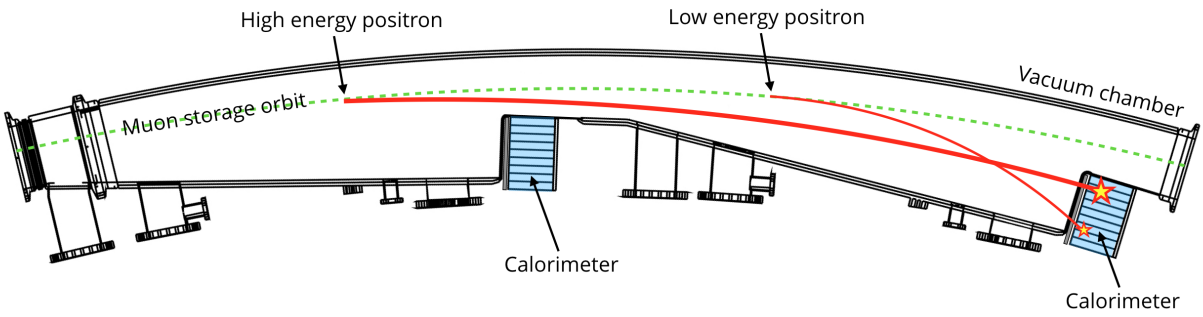


Figure 2.18: Schematic of a ring section shown from above. Muons decay into positrons, which spiral inward to be measured in tracking and calorimeter detectors. Reproduced from [65].

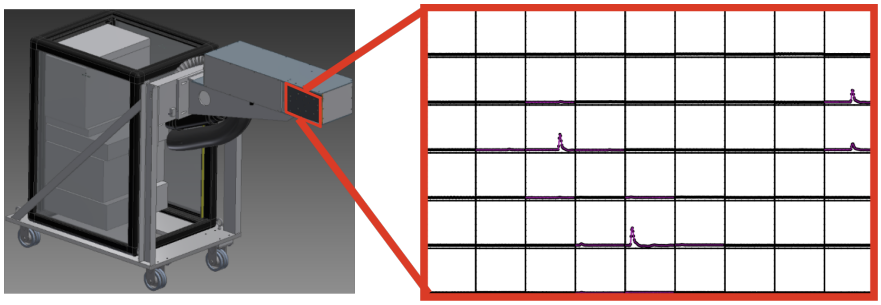


Figure 2.19: Diagram of the calorimeter crate. The location of the 54 calorimeter channels is highlighted in red. A sample event showing several positron hits in an array of channels is shown in the right.

2.6 Calorimeters

When muons decay into positrons, they spiral inward from the storage region. They are then measured in 24 calorimeter detectors equally spaced around the inside of the ring, as seen in Figure 2.11. A schematic drawing of a section of the ring looking from above is shown in Figure 2.18, showing the muon storage orbit as a green dashed line. High and low energy positron arcs are shown in red lines, and the calorimeter detectors in blue.

Each calorimeter is composed of 54 lead fluoride (PbF_2) crystals arranged in a 6×9 array [66]. Each crystal is attached to a silicon photomultiplier (SiPM) sensor. A diagram of a single calorimeter crate is shown in Figure 2.19, with the location of the crystal array highlighted in red. A sample event with positron hits in the array of 54 channels is shown on the right. A positron may deposit energy in one or several crystals. There are 1296 calorimeter channels (crystal with SiPM and associated electronics) in the full experiment. SiPMs are chosen because of their small size and ability to operate in a high magnetic field

region and because they do not perturb the magnetic field [66]. The downside is that their gains are highly temperature sensitive, resulting in the need for a sophisticated laser system for gain correction.

The physics data (number of decay positrons with a certain energy as a function of time) is extracted from the calorimeter detectors. A set of complex correction, calibration, and reconstruction procedures are required to extract high quality positron data, which will be discussed in later chapters.

2.7 Laser system

A laser system provides extremely consistent laser pulses to all detector systems. This allows for detectors to be synchronized in time with each other, so that for example all 1296 calorimeter channels are synchronized with the T0 detector. Further, the laser system provides pulses of extremely similar intensity to calorimeter channels so that their gains can be monitored. The gain corrections applied using the laser will be discussed in section 4.6.

2.8 Trackers

Two straw tracking detectors reconstruct the original transverse location of the muon decays. The position of the straw trackers relative to the storage region is shown in Figure 2.20. As decay positrons curl inward, they can pass through gas-filled straws, depositing ions and positrons [67]. Hits in subsequent straws allow for the reconstruction of the decay positron track, which can be used to reconstruct the momentum of the particle, given the magnetic field. They can also be used to estimate the transverse location of the original muon decay by extrapolating the track back to the location of azimuthal tangency in the storage region. The tracking detectors are essential for measuring the beam distribution in the storage region, which is used as an input to $\tilde{\omega}_p$. They further measure beam oscillations, providing valuable input into the precession frequency analysis and some systematic uncertainty analyses.

2.9 Current status

At the time of this thesis, Run-5 data taking has just been completed. Run-6 will start in the fall of 2022 and will be the final experimental run, focusing especially on reducing systematics. Figure 2.21 shows the

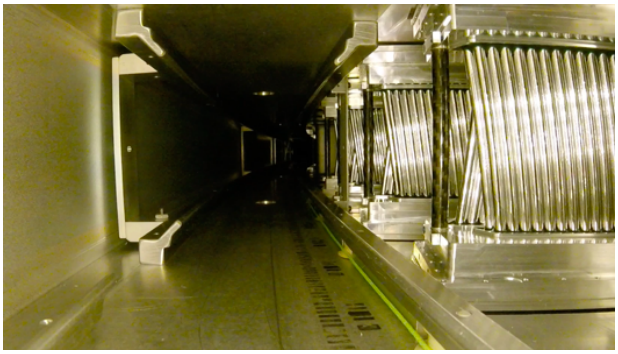


Figure 2.20: Photograph of a straw tracking detector taken from inside the vacuum chamber. The stored beam would be located at the center of the image. As stored muons decay into positrons, their tracks are measured in tracking detectors, allowing for their momentum and decay position to be reconstructed.

data taken through June 2022 in units of total positrons collected at the Brookhaven Muon $g-2$ experiment. Through Run-5, about 18 times the number of positrons collected at Brookhaven have been recorded, which will ultimately allow the experiment to achieve its statistical uncertainty goal of 100 ppb.

Many hardware upgrades have been implemented since the Run-1 analysis in order to improve operations and reduce systematic uncertainties, including upgrades to the kicker system, improved hall cooling, special detectors to measure transient fields and other sources of systematic error, and a quadrupole radio frequency (RF) system to suppress beam oscillations. The hardware improvements that have allowed for systematic uncertainty reductions for the Run-2/3 analysis will be discussed in more detail in the next section. An analysis of Run-2 and Run-3 data will be published next, with the publication expected around this winter.

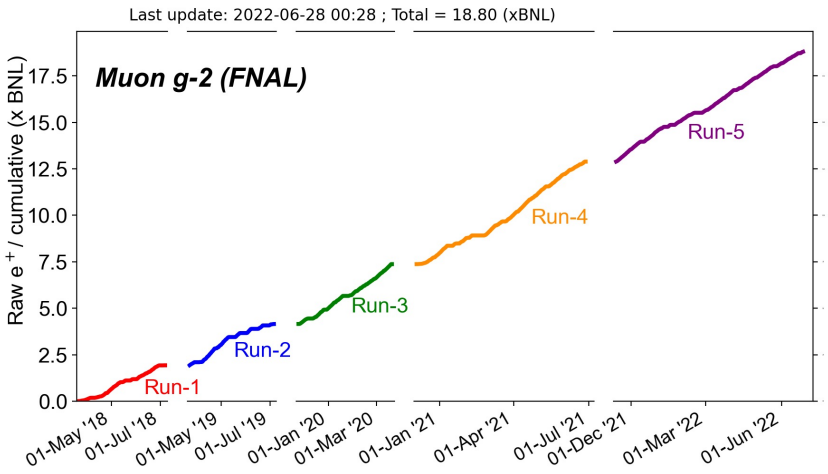


Figure 2.21: Cumulative positrons collected through Run-5 of data taking. The vertical axis is in terms of the total number of positrons collected at the Brookhaven Muon $g - 2$ experiment. This thesis concerns Run-2 and Run-3.

Chapter 3

Analysis summary: from Run-1 to Run-2/3

This chapter will summarize the full a_μ analysis using the framework established for the published Run-1 analysis. It will also describe some improvements made for the Run-2/3 analysis. The basis for the measurement of a_μ is the ratio of the precession frequency ω_a to the weighted Larmor precession frequency of free protons $\tilde{\omega}_p$. However, additional corrections to both ω_a and $\tilde{\omega}_p$ are required by experimental realities such as beam dynamics and fast transient fields. Systematic uncertainties are associated both with the corrections and with the ω_a and $\tilde{\omega}_p$ analyses. In general, corrections are acceptable (and frequently necessary) in the analysis as long as the corrections are known precisely.

The master equation for a_μ is as follows [2]:

$$a_\mu \propto \frac{f_{\text{clock}}\,\omega_a\,(1 + C_e + C_p + C_{ml} + C_{pa} + C_{dd})}{f_{\text{calib}}\,\langle \omega_p(x,y,\phi) \times M(x,y,\phi)\rangle\,(1 + B_k + B_q)}. \tag{3.1}$$

The numerator contains ω_a and its corrections, and the denominator contains $\tilde{\omega}_p$ and its corrections. Each term in the equation will be briefly discussed below, with more focus given to the C_{ml} (muon loss) correction, because it was this author's contribution to the Run-1 analysis, although it is no longer relevant for Run-2/3. More details on the Run-1 analysis can be found in [2–5].

The Run-1 corrections and uncertainties are shown in Figure 3.1. The Run-1 result was highly dominated by ω_a statistical uncertainty, which was 434 ppb relative to the combined total uncertainty of 460 ppb. The estimated Run-2/3 ω_a statistical uncertainty of 200 ppb is shown for comparison. Additionally, in the bottom section, the technical design report (TDR) goal is shown for the full lifetime of the experiment. 100 ppb is

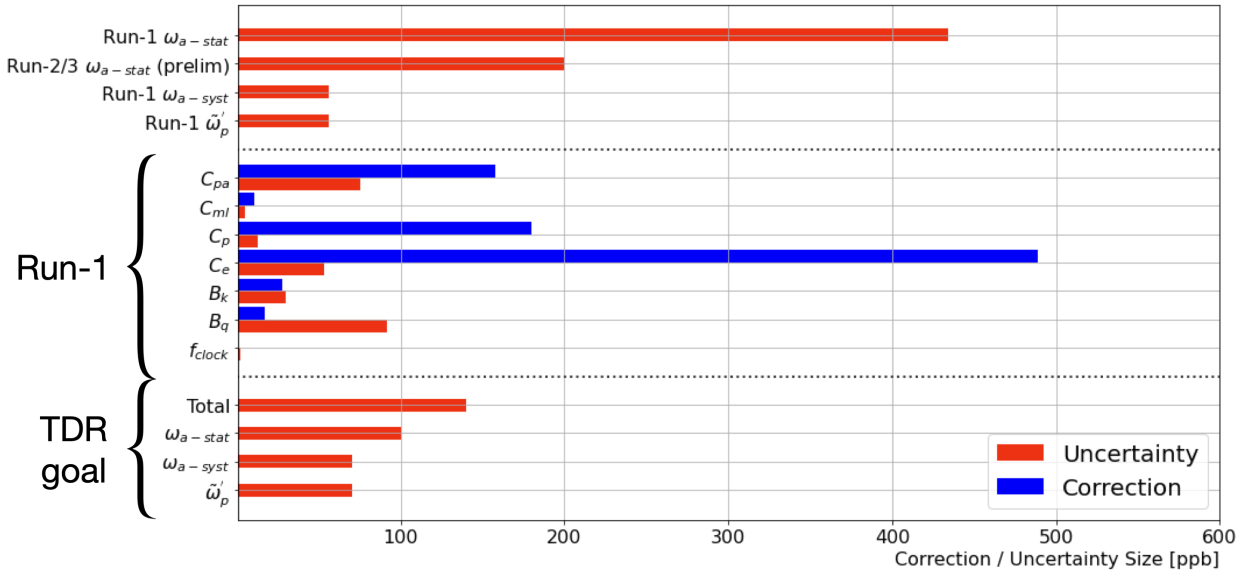


Figure 3.1: Chart of Run-1 corrections and uncertainties for the master equation given in Equation 3.1. The Run-1 result was highly dominated by the ω_a statistical uncertainty, with the largest systematic uncertainties resulting from C_{pa} (phase-acceptance correction) and B_q (ESQ transient field). The technical design report (TDR) goal for the full lifetime of the experiment and the expected Run-2/3 ω_a systematic uncertainty are also shown [2–5, 7].

ultimately allotted to the ω_a statistical uncertainty and 70 ppb each to ω_a and $\tilde{\omega}_p$ systematics. In order to achieve these systematic uncertainty goals, uncertainties on the scale of 70 ppb must be reduced, including the two largest sources of systematic uncertainty for Run-1, the phase acceptance correction C_{pa} and the quadrupole transient field correct B_q. The Run-1 ω_a systematic uncertainty should also be reduced going forward, which is a primary focus of this thesis.

3.1 Hardware improvements in Run-2 and Run-3

Several key hardware improvements were made during Run-2 and Run-3 to improve operating conditions and reduce sources of systematic uncertainty. The most significant improvement after Run-1 was the replacement of the damaged ESQ resistors. The large size of the phase-acceptance correction C_{pa} and its corresponding uncertainty was due to the beam motion induced by the long ESQ voltage recovery time in Run-1. Because the voltage did not recover as quickly as expected, the beam's vertical width and radial position changed during the measurement period, as shown in Figure 3.2 in black. This induced a large

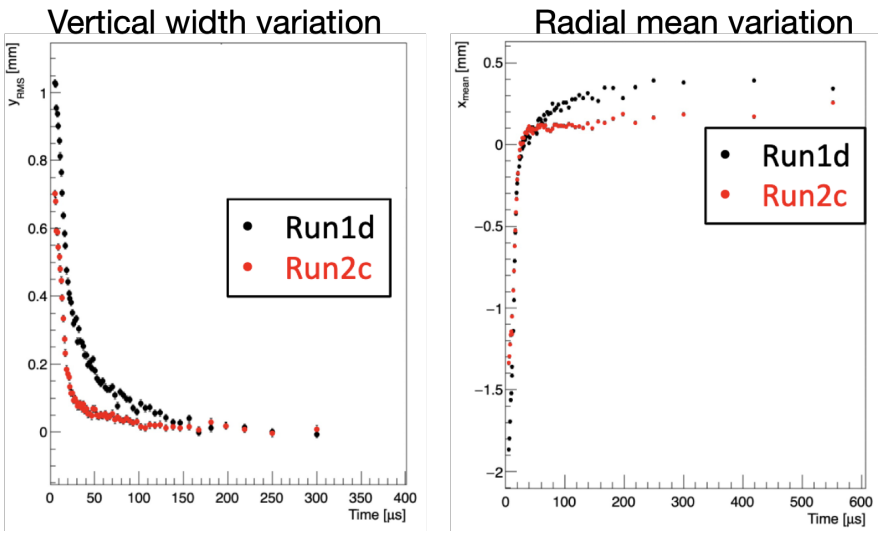


Figure 3.2: Beam motion in Run-1 vs. Run-2 as a function of time in fill. Left: variation in the vertical width or vertical waist. Right: Variation in the radial mean. After the damaged ESQ resistors were replaced, the beam motion was significantly reduced, which resulted in a decrease in the required phase-momentum correction C_{pa}. Reproduced from E. Bottalico.

systematic bias, called the phase-acceptance correction. After the resistors were replaced, the beam motion was much less, as seen in red, and this correction and its associated uncertainty are expected to be smaller in Run-2/3. C_{pa} is discussed further in subsection 3.2.6. Additionally, the damaged ESQ resistors increased the number of muons lost from the storage ring before decaying, which affected the size of the muon loss correction C_{ml}. Replacing the resistors decreased the muon loss rate by an order of magnitude, as will be discussed in subsection 3.2.5.

Insufficient hall cooling caused high temperature fluctuations in Run-1 and Run-2. These fluctuations affected the physical shape of storage ring magnet and caused changes in the magnetic field, resulting in changes in the beam behavior, as discussed further in subsection 5.3.2. The temperature fluctuations also resulted in large fluctuations in detector gain. Figure 3.3 shows the SiPM gain and temperature as a function of time for a Run-2 dataset, showing that the gain and temperature are highly correlated, meaning that the large temperature changes in the hall resulted in large changes of detector gain. Insulation was added to the magnet before Run-2 to reduce the effect of temperature changes on the magnet. The improved short-term stability of one of the field multipoles after the insulation was installed is shown in Figure 3.4 (left), although longer term temperature drifts remained a problem, as seen in Figure 3.4 (right). The long term temperature

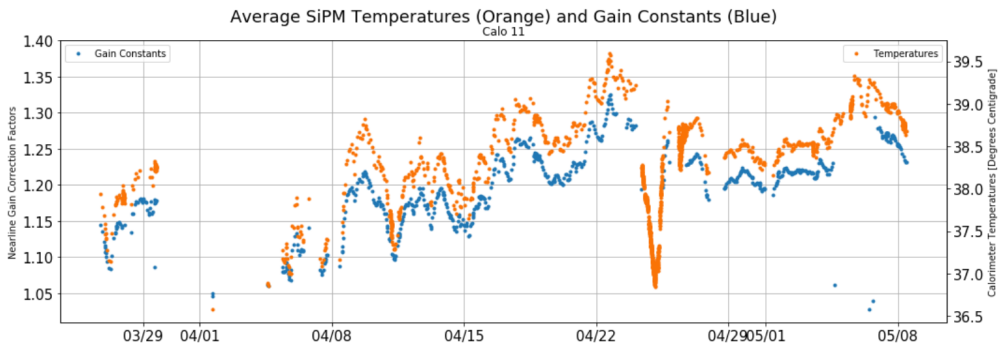


Figure 3.3: SiPM temperatures and gain as a function of time for a sample channel before sufficient hall cooling was installed, showing that SiPM gain is highly effected by temperature. Reproduced from J. Labounty.

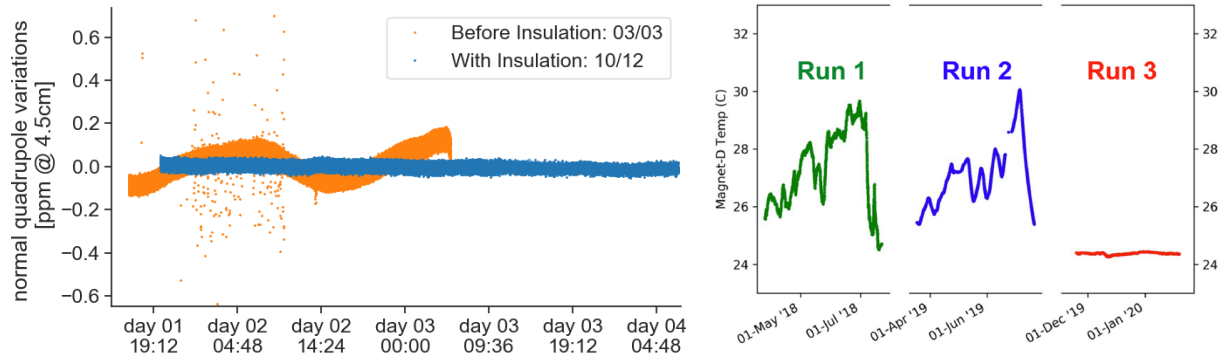


Figure 3.4: Left: improved stability of one of the field multipoles after adding insulation to the magnet. Right: Magnet temperature vs. run period showing large temperature fluctuations in Run-1/2 and minimal fluctuations in Run-3 after additional hall cooling was installed.

behavior did not improve until Run-3, when additional cooling was installed, at which point the temperature became very consistent. The reduced field fluctuations resulted in more consistent beam conditions and detector gains.

In Run-1 and Run-2, the beam was underkicked relative to its ideal orbit, so that oscillations around the center of the storage region were large. Because of hardware problems, the kicker could not reach a high enough voltage without sparking. Many hardware upgrades were made to improve the kicker performance, and an ideal kick was finally achieved partway through Run-3. Therefore, Run-3 is divided into two datasets (Run-3BM and Run-3NO), before and after the kick was improved.

After the Run-1 publications, additional special measurements were undertaken in order to reduce systematic uncertainties. For example, further measurements of the quadrupole transient correction B_q were

taken, and an additional detector was designed to measure a source of uncertainty to the electric field correction C_e. These improvements will be described briefly below.

3.2 ω_a and its corrections

The precession frequency ω_a is extracted by a fit to a histogram of the positrons above an energy threshold, as shown in Figure 2.4. Corrections to ω_a are generally related to the dynamics of the beam in the storage region, as described in this section. The ω_a fitting procedure will be discussed in great detail in Chapter 5, and the uncertainties associated with the fitting procedure will be evaluated in Chapter 6, and so will not be described further here.

3.2.1 Time changing spin phase

Many ω_a systematic uncertainties are due to a time changing spin phase. The standard five parameter fit function for the precession frequency oscillation Equation 2.12 assumes that the stored muons which will ultimately decay into measurable positrons have a constant phase ϕ_0, as determined by the fit function. This phase corresponds to the fit's assumption about the spin phase, or position of the spin vector with respect to the momentum vector, at $t = 0$. Here times have been corrected so that $t = 0$ corresponds to the mean time of the T0 detector pulse, as described in section 2.4. If, however, the phase determined by the fit for $t = 0$ is time dependent, $\phi = \phi(t)$, then the Taylor expansion of $\phi(t)$ gives

$$\phi(t) = \phi_0 + \frac{d\phi}{dt}t + \frac{d^2\phi}{dt^2}t^2 + ... \tag{3.2}$$

Plugging $\phi(t)$ into the five parameter fit function Equation 2.12 results in a shift to ω_a, $\Delta\omega_a = \frac{d\phi}{dt}$. This effect is indistinguishable from the true ω_a. Because the precession frequency fit is performed over many fills added together, only effects that occur coherently over the time of the fill (the ≈30-650 μs measurement period), termed "early-to-late" effects, result in a changing phase. On the other hand, effects occurring over a long time scale, such as a slowly varying detector gain due to diurnal changes, generally wash out due to the fitting procedure and are not as concerning.

For example, detector gain changes induced by the intense beam splash at the beginning of the fill

are considered an early-to-late effect. As the detectors recover from the splash, the gain changes coherently every fill. Because the ω_a analysis uses an energy threshold, a changing energy scale affects which positrons enter the final histogram, which induces a changing phase as a function of time. Early-to-late effects also include phase shifts associated with the beam rate. The rate of positron hits is higher at the beginning of the fill than at the end as the muons decay away at the muon lifetime. Rate-dependent reconstruction errors and gain changes due to two pulses arriving close in time are examples of rate-dependent, and therefore potentially early-to-late, effects. Reconstruction algorithms are designed to minimize rate-dependent biases, and gain and pileup corrections are applied to reduce the remaining effect. Lastly, beam changes can cause different portions of the muon beam to be measured as a function of time. If the beam's average measured ϕ_0 changes as a function of time, or if muons with certain phases are more likely to be measured as a function of time, a bias will occur. Two corrections discussed in this section, the muon loss correction C_{ml} and the phase-acceptance correction C_{pa}, fall under this last category.

3.2.2 Clock unblinding factor f_{clock}

f_{clock} is the clock unblinding conversion factor. Muon $g-2$ is a blinded experiment, meaning that the true a_μ is not known when performing the analysis. In order to ensure the experiment remains blinded, the 40 MHz master clock for the experiment is detuned by independent Fermilab staff to a frequency between 39997 Hz and 39999 Hz. This conversion factor is only revealed when the collaboration has approved the decision to unblind [3]. Additional software blinding factors are also used in the ω_a and $\tilde{\omega}_p$ analyses so that different analyzers are blinded with respect to each other. The clock unblinding factor has a negligible uncertainty.

3.2.3 Electric field correction C_e

Because a quadrupole electric field is used to vertically focus the beam, the electric field experienced by the muons is nonzero. The majority of the electric field term in Equation 2.5 is cancelled by the choice of the "magic momentum" of 3.1 GeV/c, but a small correction is still necessary because of the finite width of the momentum distribution [4], meaning that not all muons are at the magic momentum. This correction is referred to as C_e, the electric field correction.

If a muon is oscillating around an equilibrium radius x_e, then the electric field correction is [4]

$$C_e \approx 2n(1-n)\beta_0^2 \frac{\langle x_e^2 \rangle}{R_0^2}, \tag{3.3}$$

where n is the field index responsible for the weak focusing in the vertical direction, which depends on the ESQ voltage. Also, $\beta_0^2 = p_0^2/[m^2c^2 + p_0^2]$, where p_0 is the magic momentum and R_0 is the corresponding magic radius in the magnetic field. Calculating C_e requires measuring the radial distribution of the beam $\langle x_e^2 \rangle$. The electric field depends on both the mean and width of the distribution, since $\langle x_e^2 \rangle = \sigma_{x_e}^2 + \langle x_e \rangle^2$ [4].

The "fast rotation" analysis is used to measure the momentum distribution, and therefore the radial distribution, of the beam. Because of the correlation between muon momentum and radial position, the higher frequency (small radius) muons will gradually lap the lower frequency (large radius) muons as a function of time in the fill. By performing a fast Fourier transform (FFT) of the oscillating cyclotron signal in the calorimeters, the radial distribution can be extracted. The fast rotation distributions for Run-1 and Run-2/3 can be seen in Figure 3.5 on the left and right, respectively. An equilibrium radius of 0 on these plots corresponds to the magic radius R_0. The Run-1, Run-2, and Run-3BM distributions are asymmetric with respect to the R_0, because the kick did not center the muons in the storage region, whereas Run-3NO has a more symmetric distribution, because the kick strength was increased.

For Run-1, the value of C_e was +489 ppb. The largest contribution to the C_e uncertainty was due to a correlation between the time and momentum of the beam after it was kicked. Methods of calculating the momentum distribution rely on the assumption that the momentum of each muon is uncorrelated with its time in the pulse. However, because the kicker strength varied in time, as seen in Figure 2.17 (right), a correlation is introduced, which causes an error in the reconstruction of the beam momentum distribution, as seen in Figure 3.6. Muons with a time-momentum correlation were simulated, and their equilibrium was reconstructed using the standard analysis, as shown in black. The true momentum distribution is shown in red. Since the exact degree of the phase-momentum correction was not known from data for Run-1, a relatively large uncertainty of 50 ppb was assigned to C_e.

Efforts are underway to measure the time-momentum correlation of the beam going forward in Run-2/3, including building a new detector at UW. This minimally intrusive scintillating fiber detector (MiniSciFi)

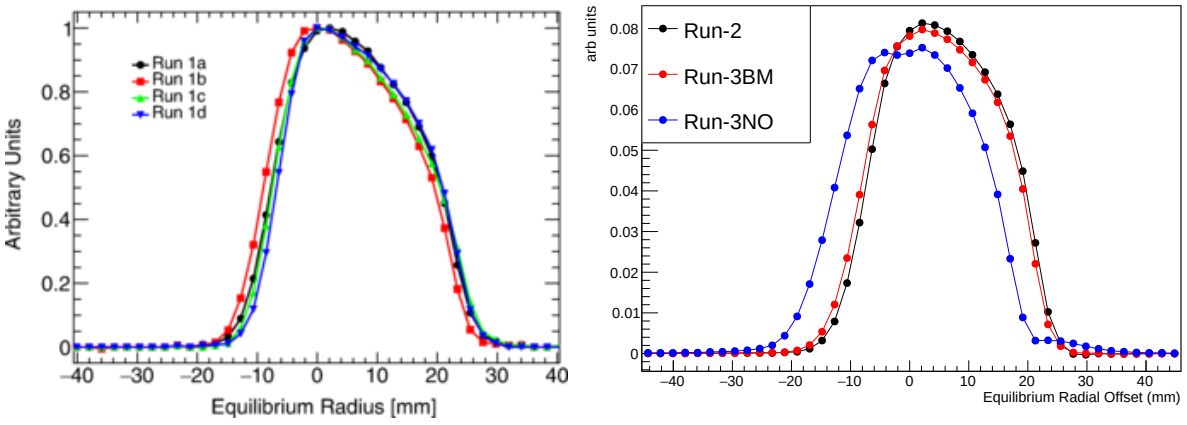


Figure 3.5: Equilibrium radius distributions from Run-1 (left, reproduced from [4]) and Run-2/3 (right, courtesy of T. Barrett) calculated using the Fourier method. The zero point on the horizontal axis is set to be the magic radius. The asymmetric distribution all datasets except Run-3NO is due to an insufficient magnetic kick. Run-3NO has a more centered distribution due to the improved kick.

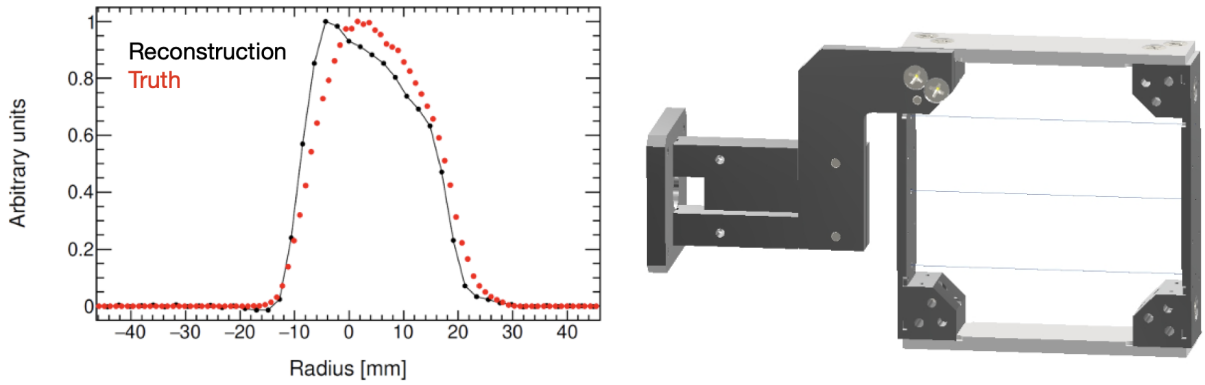


Figure 3.6: Momentum reconstruction error due to the time-momentum correlation, from simulation. Reproduced from T. Barrett. Right: minimally intrusive scintillating fiber detector (MiniSciFi) designed to measure the time-momentum correlation. The beam will pass through the middle of the detector. Reproduced from B. MacCoy, C. Claessens, et. al.

will measure narrow time slices of the beam without significantly perturbing the beam itself. Figure 3.6 (right) shows a diagram of the MiniSciFi detector. These measurements will help reduce the uncertainty on the mis-reconstruction of the beam momentum going forward to reduce the uncertainty on C_e.

3.2.4 Pitch correction C_p

The middle term in Equation 2.5 requires a "pitch" correction, C_p. The muons oscillate vertically, and therefore they travel vertically in the magnetic field at an angle that is nonzero on average [4], that is, $\vec{\beta} \cdot \vec{B} \neq 0$. The pitch correction is

$$C_p \approx \frac{n}{2} \frac{\langle y^2 \rangle}{R_0^2} = \frac{n}{4} \frac{\langle A^2 \rangle}{R_0^2}, \tag{3.4}$$

where y is the vertical position of the beam and A is the vertical oscillation amplitude, which is measured by the tracking detectors. The uncertainty on C_p is less than 15 ppb, with the dominant uncertainty coming from the tracker reconstruction and alignment. The uncertainty on this correction is already sufficiently low for future analyses. For Run-1, C_p was +180 ppb.

3.2.5 Muon loss correction C_{ml}

The lost muon correction C_{ml} was this author's main contribution to the Run-1 analysis. It is no longer a factor in the Run-2/3 analysis. Therefore the correction is discussed in some detail in the following section and in full detail in Appendix A.

Lost muons are muons that hit aperture-defining collimators and exit the storage region before decaying into positrons. They are minimum ionizing particles (MIPs), depositing a small and well-known amount of energy as they pass through multiple adjacent calorimeters. They have a characteristic flight time of ≈ 6.2 ns between adjacent calorimeters, as seen in Figure 3.7. Measuring the time-dependent number of lost muon hits in the calorimeter gives only the shape of the losses throughout the fill; the overall efficiency scaling factor comes from a many-parameter ω_a fit. The incorporation of lost muons into the ω_a fit will be described in subsection 5.3.2.

If the lost muons have a different phase from the average stored muon population, they can induce an early-to-late effect. If their phases are the same, then the lost particles only change the number of stored particles, not their average phase. If, however, they differ in phase, they will cause the phase of the beam

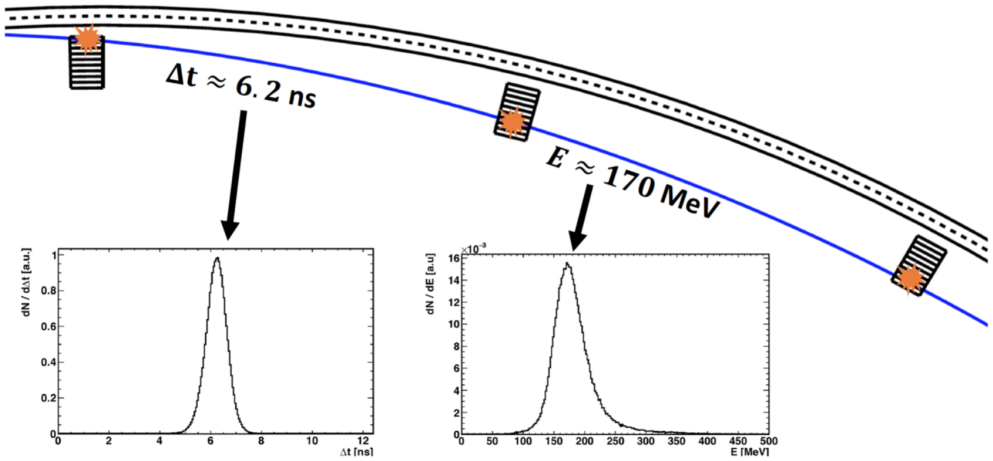


Figure 3.7: Lost muons passing through adjacent calorimeters are minimum ionizing particles and deposit known amounts of energy. They also have a characteristic flight time between detectors. This information is useful for timing corrections and energy calibrations, in addition to generating a lost muon spectrum for precession frequency fitting. Reproduced from [63].

(and therefore the measured positrons) to change coherently. One possible way for this to happen is if

$$\frac{d\langle\phi\rangle}{d\langle p\rangle} \neq 0, \tag{3.5}$$

meaning that the average stored muon phase and momentum are correlated, and

$$\frac{d\langle p\rangle}{dt} \neq 0, \tag{3.6}$$

meaning that the stored momentum is changing early-to-late. This could occur if muons are lost in a momentum dependent way. For example, if low momentum muons were more likely to be lost than high momentum muons, the average stored momentum would change over time. Because the momentum and phase are correlated, the phase would change early-to-late as well. Then

$$\frac{d\langle\phi\rangle}{d\langle p\rangle} \cdot \frac{d\langle p\rangle}{dt} = \frac{d\langle\phi\rangle}{dt} = \Delta\omega_a \neq 0, \tag{3.7}$$

and there will be a bias to ω_a. The size of the bias scales with the phase-momentum correlation, the momentum dependence of the losses, and the absolute loss rate. If the loss rate is higher, then the effect will be magnified.

In Run-1, there was significant concern surrounding this source of uncertainty. Firstly, simulations predicted a phase-momentum correlation arising from the beam dynamics of the particles in the delivery ring. It is also intuitively likely that losses are momentum-dependent, since particles on the low and high edges of the momentum distribution have lower and higher average radii in the storage region, respectively, and are more likely to hit a collimator. Further, losses in Run-1 were very high due to damaged ESQ resistors. Because there was no data about the momentum dependence of the losses, the systematic uncertainty was calculated for the worst-case scenario where all muons were lost from one edge of the momentum distribution, which suggested that the bias could be $>$ 100 ppb [63] in Run-1. Although losses in Run-2 and Run-3 are $\mathcal{O}(10)$ times lower than in Run-1, the Run-1 muon loss correction analysis informs the Run-2 and Run-3 analysis, so a brief summary of the measurement and analysis procedure for the Run-1 result is given below, followed by a discussion of the phase-momentum correlation systematic in Run-2 and beyond.

The Run-1 muon loss correction measurement involved taking two datasets under special conditions. The purpose of the first half of the study was to establish the phase-momentum correlation, and the purpose of the second was to measure the momentum dependence of the losses.

The first study involved changing the magnetic field by $\pm 0.67\%$ so that different slices of the incoming momentum distribution were stored. The phase was then extracted from an ω_a fit to the data. The phase measured at three different field settings can be seen in Figure 3.8 in black, with a linear fit applied to the data. Results from an end-to-end beamline simulation of the correlation are shown in blue, showing good agreement with the data. This study established the phase-momentum correlation of -10.0 ± 1.6 mrad$/\%\Delta p/p_0$.

The second study used the upstream delivery ring collimators to select momentum-biased slices of the nominal momentum distribution. For each biased distribution, the lost muon spectrum was also measured in order to determine any asymmetry in the number of losses. Sample momentum distributions (determined from the fast rotation analysis) are shown in Figure 3.9. The black dashed line is the nominal momentum distribution, and for example "1/5 low" means that the beam intensity has been reduced by scraping off the high momentum edge of the distribution so that a low momentum slice of the full distribution is taken. A comparison of the loss spectra from the lowest and highest momentum slice is shown in Figure 3.9 (right). A small asymmetry in the number of losses at early times is evident, suggesting that lower momentum muons are slightly more likely to be lost.

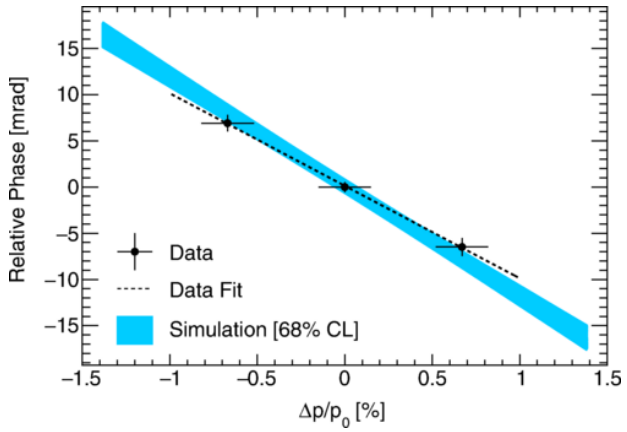


Figure 3.8: Relative phase from the five parameter fit as a function of the percent change in stored beam momentum relative to the nominal (magic) momentum p_0. Data from a dedicated systematic study is shown in black, compared to results from an end-to-end simulation shown in blue. There is good agreement between data and simulation. Reproduced from [4].

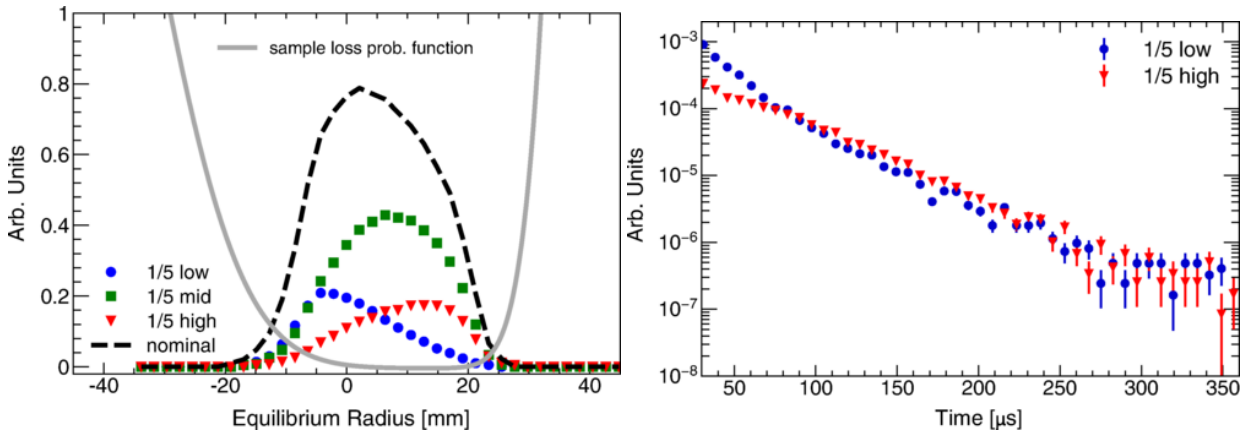


Figure 3.9: Left: biased momentum distributions used for C_{ml} studies. The nominal momentum distribution is shown in black, with momentum-biased slices of the distribution shown in blue, green, and red. A sample loss probability function at early times is shown in gray, which has been calculated using the momentum distributions and loss spectra shown on the right. The loss probability is asymmetric, with low momentum muons slightly more likely to be lost at early times. Right: triple coincidences as a function of time, normalized to the beam intensity. The blue and red points correspond to the momentum distributions visible on the left. At early times, low momentum muons are more likely to be lost. Reproduced from [4].

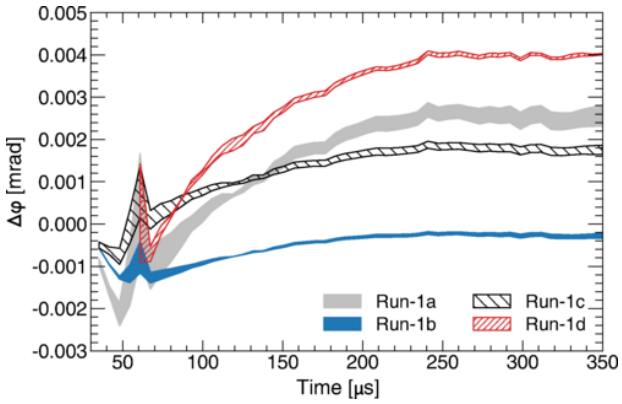


Figure 3.10: Phase shift as a function of time for the Run-1 datasets. It is calculated by time evolving the momentum distribution throughout the fill and converting the momentum change to a phase shift using the time-momentum correlation. A time changing spin phase induces a bias to ω_a, which is corrected by the muon loss correction factor C_{ml}. The shaded regions correspond to uncertainties due to the choice of muon loss function. Reproduced from [4].

An analytical model was used to describe the momentum dependence of the losses as a function of time in fill based on the momentum distributions and corresponding muon loss spectra. An example loss probability function at early times is shown in Figure 3.9 (left) in gray, representing the loss probability asymmetry as a function of momentum. As the losses depleted certain portions of the momentum distribution (which are shown for the full Run-1 datasets in Figure 3.5), the average stored muon momentum changed as a function of time. The phase-momentum correlation from the first part of the study was then used to convert the average momentum of the stored muons to the average phase of the stored muons. The phase change as a function of time can be see in Figure 3.10. By incorporating a time changing phase into a toy model of the precession frequency histogram, the shift due to the phase-momentum correlation of the lost muons could be calculated.

This correction to ω_a, C_{ml}, varied between -3 and -17 ppb for Run-1 datasets, largely depending on the number of losses in each dataset [4]. Uncertainties on the correction were around 6 ppb and arose from three sources: uncertainty on the loss function, uncertainty on the choice of analytical model for the loss probability, and uncertainty on the phase-momentum correlation. Ultimately, the muon loss correction was much smaller than the worst-case scenario, because the muons were lost relatively symmetrically in momentum, with only a small asymmetry present. The uncertainties were also very small.

In Run-2 and Run-3, the losses were an order of magnitude lower than in Run-1, as seen in Figure 3.11.

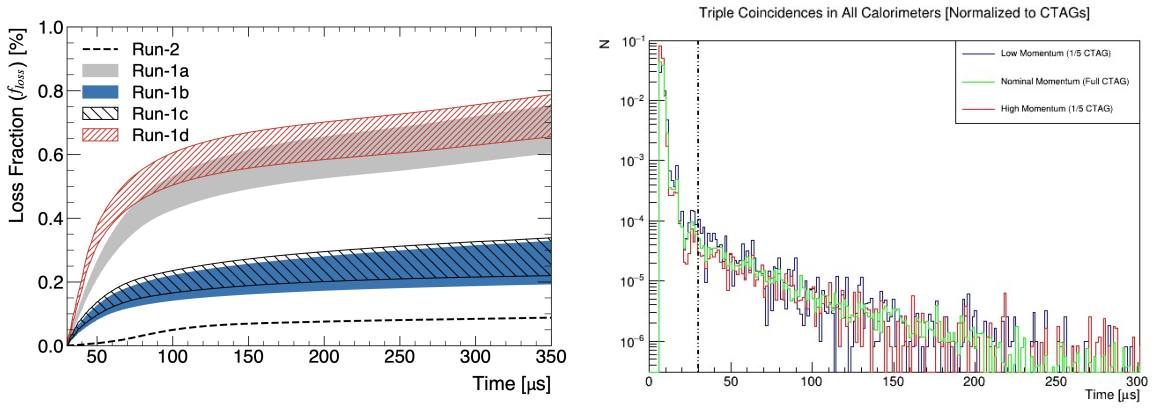


Figure 3.11: Left: fractional losses as a function of time in Run-1 datasets and all of Run-2 showing much lower losses in Run-2. Right: normalized muon loss spectrum for low and high momentum distributions in Run-2 showing no asymmetry in the loss rates.

This plot shows the loss fraction as a function of time in fill for the four Run-1 datasets, with the error bars representing the result of different reconstruction and analysis choices. The black dashed line represents the Run-2 loss fraction (Run-3 losses are lower still). Given that Run-1b had $C_{ml} = -3$ ppb because of relatively lower losses compared to other Run-1 datasets, C_{ml} in Run-2 and beyond can be assumed to be negligible. However, in order to ensure that the loss probability as a function of muon momentum did not get more asymmetric in Run-2, an abbreviated version of the second study was repeated with Run-2 conditions (i.e., without the bad resistors) as seen in Figure 3.11 (right). An even smaller asymmetry was observed with Run-2 conditions. Therefore C_{ml} and the uncertainty associated with it are considered negligible in Run-2 and Run-3.

3.2.6 Phase-acceptance correction C_{pa}

The term C_{pa} is the phase-acceptance correction. Calorimeters have a nonuniform acceptance, meaning that decay positrons have a different probability of being measured depending on where the original muon decayed in the storage region. A simulation of the measured phase as a function of decay position can be seen in Figure 3.12 (left). As long as the beam position is constant through the fill, the dependence of the acceptance on the muon location in the storage region is irrelevant. However, in Run-1, because of the damaged ESQ resistors discussed in section 2.5, the beam moved significantly throughout the measurement period, as shown in Figure 3.2. Figure 3.12 (right) shows a schematic of the changing phase as the beam's

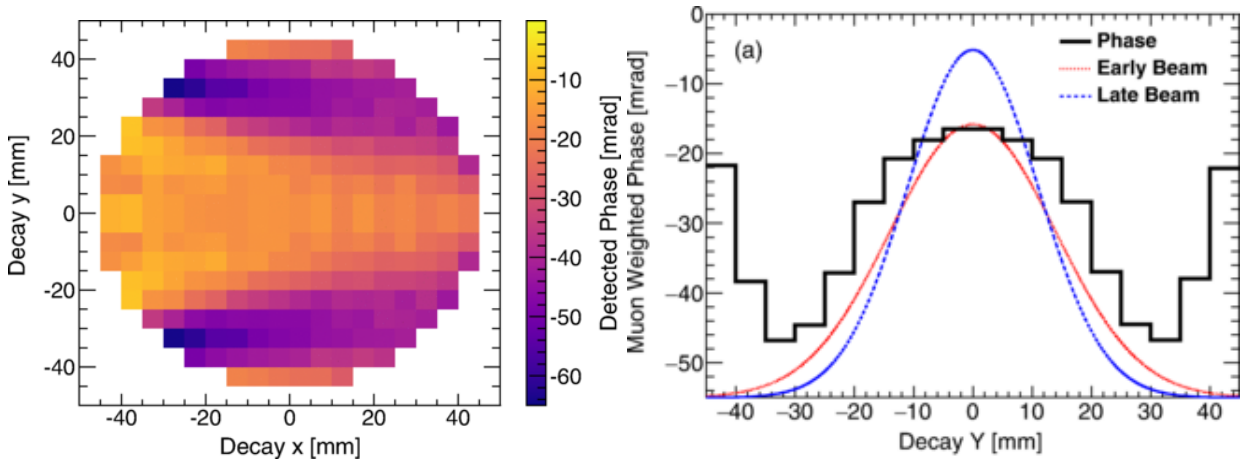


Figure 3.12: Left: detected phase as a function of the decay position of the muon, as calculated in simulation. As the beam moves in the storage region, the detected phase changes, inducing a systematic error. Right: the phase as a function of vertical position is shown in black. As the vertical beam width changes over time (exaggerated, in red and blue), a different part of the underlying phase vs. decay position distribution is sampled, resulting in a different detected phase. Both figures reproduced from [4].

vertical width decreases as a function of time. Since the measured phase is different depending on the vertical decay position, as the beam decreases in width, the measured phase changes. This time changing phase results in a bias to ω_a.

The C_{pa} correction and its uncertainty were large in Run-1, with $C_{pa} = -158$ ppb and σ_{pa} = 75 ppb. Uncertainties arose due to the tracker analysis used to reconstruct the beam's position, the simulations used to generate acceptance maps, and beam dynamics models used to extract calorimeter distributions from tracker distributions [4]. However, in Run-2 and beyond, the beam motion was significantly reduced, because the ESQ resistors had been replaced. Therefore, the correction for the Run-2/3 analysis will be about a third of the size as it was in Run-1, with the uncertainty on the correction lower as well.

3.2.7 The differential decay correction C_{dd}

In subsection 3.2.5, a correlation between muon phase and momentum was established. As muons were preferentially lost from the lower momentum side of the distribution, the average momentum of the stored beam changed with time, resulting in a bias to ω_a. A similar type of error, called differential decay, arises because the muon lifetime is dependent on its momentum, with lower momentum muons having shorter lifetimes than higher momentum muons. Over time, the low momentum side of the distribution will be

depleted relative to the high momentum side of the distribution, and the phase-momentum correlation will couple the momentum change to a phase change of the stored muons. This correction was not included for Run-1, but analysis for Run-2/3 is underway. The correction is expected to be around 10-20 ppb, and the uncertainty on the correction, which will arise from uncertainty on the phase-momentum correlation and measured momentum distributions, is expected to be small.

3.3 $\tilde{\omega}_p$ and its corrections

The denominator of Equation 3.1 contains the field measurement and its corrections. The first term in the denominator, $f_{\text{calib}} \langle \omega_p(x, y, \phi) \times M(x, y, \phi) \rangle$, is the calibrated weighted average of ω_p as experienced by the stored muon distribution $M(x, y, \phi)$. The final two terms, B_k and B_q, are fast transient fields due to the electrostatic quadrupoles and kickers, respectively, which require a correction as they are not present during the trolley runs used to measure the field [5]. Because the fixed probes are shielded from these changes due to the skin depth of the vacuum chamber, and because the transient fields occurred on a faster time scale than the sampling period of the fixed probes, new measurement systems were designed for both transients.

For Run-1, the combined uncertainty of the field calibration (described below) and the muon weighted field was 56 ppb, arising from uncertainties in the interpolation between trolley run, as visible when comparing the blue and orange lines in Figure 2.7; from calibration uncertainties; and from the tracking detector input to the field weighting. The uncertainty on $f_{\text{calib}} \langle \omega_p(x, y, \phi) \times M(x, y, \phi) \rangle$ was significantly lower than the transient field uncertainties for Run-1.

3.3.1 Field calibration factor f_{calib}

The term f_{calib} is the absolute calibration of the trolley NMR probes used to measure ω_p. The final magnetic field measurement is ultimately presented in terms of $\tilde{\omega}'_p(T_r)$, the proton spin precession frequency in a spherical water sample at 34.7 $^\circ C$, weighted by the muon position in the storage region. The trolley probes require corrections due to differences in diamagnetic shielding of petroleum compared to water, sample shape, and magnetization of the trolley materials. In order to calibrate the trolley, a calibration probe is repeatedly swapped into the locations of the trolley probes. The calibration probe is then cross calibrated with a spherical water sample and a ^{3}He magnetometer [2, 5]. The uncertainty of f_{calib} is combined with the

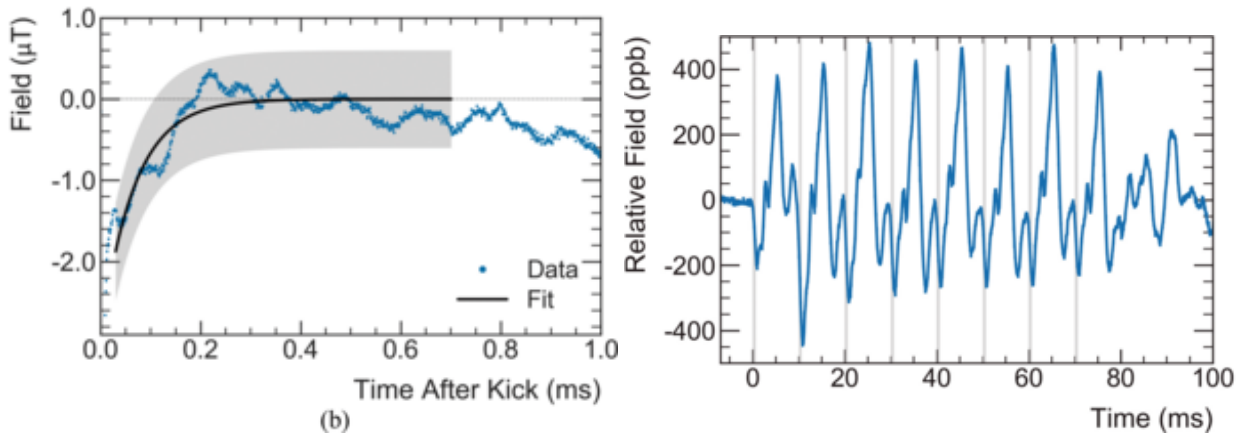


Figure 3.13: Left: kicker transient field measured by the fiber magnetometers. A fit to the transient and the uncertainty on the fit (in gray) are shown. The length of the gray section corresponds to the ω_a fitting period. Right: electrostatic quadrupole transient field. The gray vertical lines correspond to pulsing of the ESQs, during which physics data can be taken. Reproduced from [5].

uncertainty on $\tilde{\omega}_p$.

3.3.2 Kicker transient B_k

When the kickers pulse, they induce eddy currents in the surrounding metal, resulting in a field perturbation. The term B_k, or the transient field correction, accounts for this perturbation. Two fiber magnetometers were designed to measure B_k. This measurement can be seen in Figure 3.13 (left). A fit has been applied to the data, and the uncertainty on the fit is shown in gray. The length of the gray section corresponds to the ω_a fitting period. Uncertainties on B_k arise from the fit function, the background subtraction, and the calibration of the magnetometer [5]. B_k was -27 ppb in Run-1.

3.3.3 Quadrupole transient B_q

The term B_q is the transient field due to the electrostatic quadrupoles (ESQs). The pulsing of the ESQs induced mechanical vibrations in the charged plates, which generated a transient field. Dedicated measurements were performed using trolley probes sealed in plastic tubes, as seen in Figure 3.13. The gray region marks times when the ESQs were charged, during which muons could be stored. Measurements of the transient in Run-1 were primarily performed in a single ESQ section, as it was the most accessible. B_q was -17 ppb in Run-1, but with a systematic uncertainty of 92 ppb. This large uncertainty was due to the time delay

between Run-1 and the transient field data, which was taken in Run-3, and because only one of 12 ESQ sections was fully measured [3]. For the Run-2/3 analysis, full mapping of the ring with the probes was performed. Therefore the systematic uncertainty can be expected to significantly improve going forward.

Chapter 4

From raw data to physics

This chapter discusses the steps to convert raw positron hits in calorimeters into physics-quality data for the precession frequency analysis. First, the set of algorithms and procedures that reconstruct the time and energy coordinates from raw traces will be described. Then, the data quality control (DQC) procedure, which removes low quality data before analysis, will be discussed. Procedures performed on the physics-quality data to prepare it for fitting, such as pileup corrections and weighting, will be discussed in the next chapter.

A schematic of the reconstruction procedure is shown in Figure 4.1. In summary, positrons hit a calorimeter, which has 54 channels arranged in a 6×9 array. The positrons may hit a single channel or multiple channels. In order to avoid very high data rates, traces are only saved if the voltage in the channel goes above a certain threshold; this is called island chopping. These saved traces are fit using pulse templates generated from data. Based on dedicated calibration runs and timing alignment measurements, the template fits are converted into (time, energy) coordinates for each channel. A gain correction is also applied. This information is stored in a "crystal hit" data product. The crystal hits are then partitioned into "clusters," where ideally one cluster is equivalent to one positron. The (time, energy) coordinates of the clusters ultimately enter the precession frequency histogram from which ω_a is extracted. This chapter will summarize the full reconstruction chain and then discuss an improvement made to the clustering algorithm by this author for the Run-2/3 analysis.

Two main reconstruction chains are used by different analyzers. These two algorithms are referred

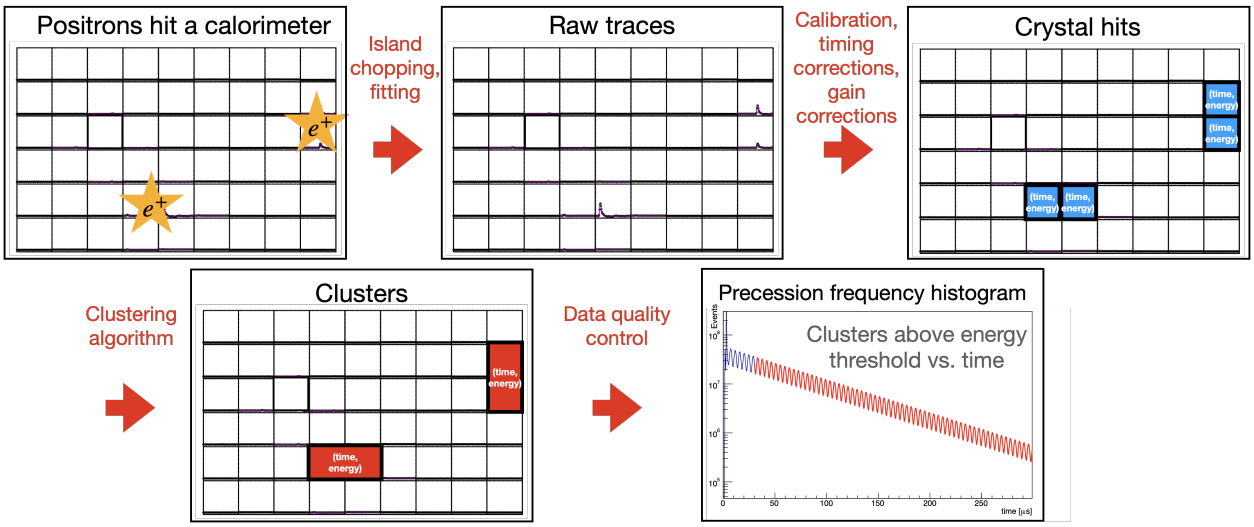


Figure 4.1: Summary of the reconstruction procedure to convert raw traces into physics quality data. First, traces above a certain threshold are saved and fit to templates generated from data. Then, calibration, timing corrections, and gain corrections are applied to convert the traces to (time, energy) coordinates in each channel. Then, the clustering algorithm groups crystal hits together to reconstruct the time and energy of the original positron. Data quality control is applied to remove low quality data. Then the clusters above a certain energy threshold are plotted.

to as "Recon West" (RW), because it was created by A. Fienberg at UW [63] and "Recon East" (RE), because it was created by D. Sweigart at Cornell [68]. This analysis inherits many of the procedures used by A. Fienberg and therefore uses the RW algorithm. The differences between the two will particularly become important when discussing the reconstruction-induced slow term in subsection 5.6.1, so they will be discussed in the relevant sections below.

4.1 Island chopping

The data rate received by the calorimeters is very high. The data are digitized at 800 MSPS (mega-samples per second), and muons arrive in 700 μs fills at a frequency of about 12 Hz. If all of the data were saved, this would result in a data rate of about 18 GB/s. Therefore, graphical processing units (GPUs) search for voltages above a certain threshold in the calorimeter channels. Samples before and after this voltage peak and samples in adjacent calorimeter channels are "chopped" and saved for analysis [7, 63, 69]. These portions of traces are referred to as "islands" and can be of varying lengths depending on the number of pulses above threshold in the calorimeter [63]. In Run-1, when a calorimeter channel exceeded a voltage

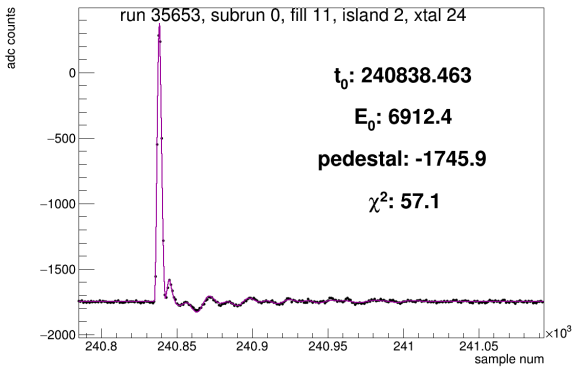


Figure 4.2: Sample pulse fit showing good fit quality, with ADC (analog-to-digital converter) counts on the vertical axis and sample number (or clock ticks) on the horizontal axis. The pedestal has not yet been subtracted, and the energy is uncalibrated at this stage. The black points are the raw samples and the purple line is a template fit to the data.

threshold, islands from all channels in that calorimeter would be chopped. In Run-2 and beyond, in order to reduce the data rate, only data from the adjacent 3×3 square of channels surrounding the above-threshold channel was chopped.

4.2 Pedestal corrections

The chopped islands then undergo corrections before they can be fit. The waveform has a baseline offset, or pedestal, which must be subtracted so that the minimum value of the waveform is zero. The pedestal is visible in a sample trace in Figure 4.2, which shows a sample trace, where it is about -1750 ADC (analog-to-digital converter) counts, where 1 ADC count $\approx$ 0.25 mV. Further, because the digitizers used in the experiment have two chips which have slightly offset pedestals, every other sample has a slight offset with respect to its adjacent sample. An "even-odd" correction is applied to correct this problem, and then the pedestal value is subtracted off of the trace before fitting [63, 70].

4.3 Pulse fitting

The traces are then fit using pulse templates generated from data. This template-based procedure allows fitting to occur with only two parameters per pulse: the pulse time and the pulse integral [63].

Each calorimeter channel has its own characteristic pulse shape, and the pulse shape varies between

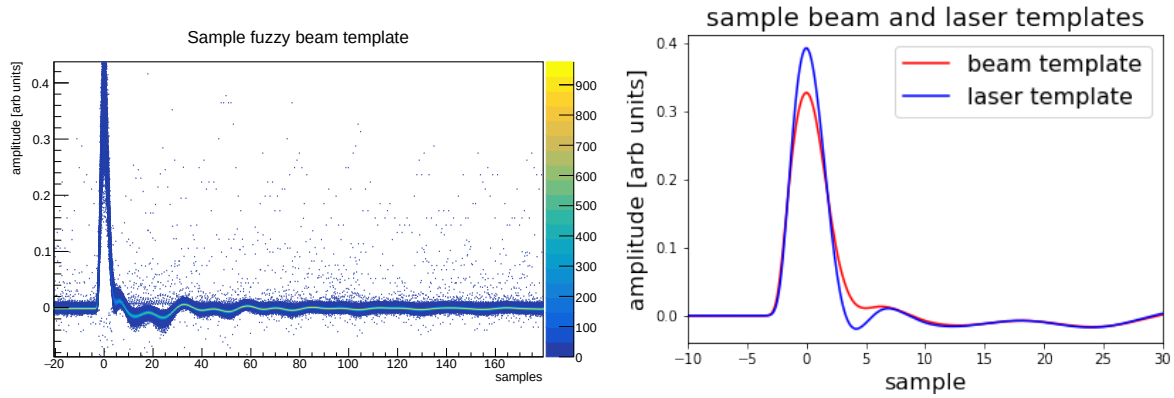


Figure 4.3: Left: sample "fuzzy" beam template for a single channel created from many beam pulses. A spline fit will be applied to the fuzzy template. Right: comparison of beam and laser template splines in the same channel showing that the laser template shape is narrower than the beam template shape.

beam (positron) pulses and laser pulses. Therefore, pulse templates are created individually for all 1296 channels and for beam and laser pulses separately. This procedure is repeated at the beginning of every year of data taking.

For each channel, many traces are accumulated, time aligned, and overlaid to create a "fuzzy" template. The traces are longer than usual islands in order to capture the ringing behavior after the main pulse. Figure 4.3 (left) shows a fuzzy template in blue and yellow. A spline interpolates between points on the fuzzy template histogram to generate a continuous template function. Template shapes have been shown to be relatively insensitive to pulse energy, rate, and time [63, 65, 71]. This template is compared with a laser template in the same channel in the plot in Figure 4.3 (right), showing that laser template is narrower than the beam template.

A fit using the template fitting procedure can be seen in Figure 4.2. The energy shown here is still uncalibrated. A high quality fit is achieved, with a good fit of the long tail after the positron pulse. The importance of updating the templates at the beginning of each run is demonstrated by Figure 4.4. Fits were performed on Run-3 data using the old Run-2 templates and the updated Run-3 templates. Better fit quality was achieved using the updated templates. Because several damaged channels were replaced between Run-2 and Run-3, the new channels had different average pulse shapes than the old channels and new templates were required.

The fitter also performs chain fitting, where two pulses in the same channel are fit together, so that the

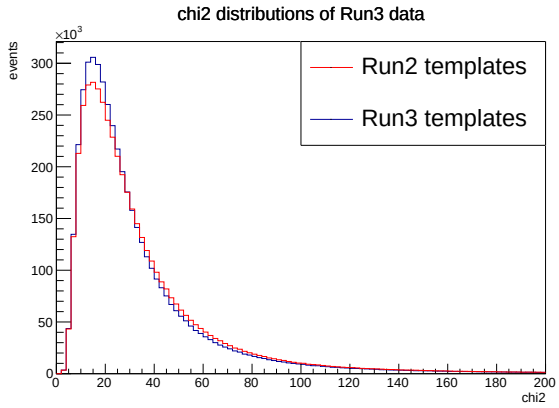


Figure 4.4: Comparison of χ^2 distribution of fits using Run-2 templates and Run-3 templates on Run-3 data, showing better fit quality with the updated templates.

tail of the template can overlap the next pulse. Chain fitting may occur across multiple islands, as seen in Figure 4.5. This sample fit shows that three pulses have been fit together. The second shape is fit to two templates 5 ns apart, showing the power of the fitter to resolve pulses close together in time.

After fitting, each trace has an associated uncorrected time and uncalibrated pulse integral. Calibrations, timing corrections, and gain corrections are required to convert this information to physics-quality time and energy coordinates.

The RW and RE algorithms have a major difference in their approach to fitting. The RW algorithm checks each channel separately to determine whether it is above the primary fitting threshold. After the first fit is performed, it checks if the residuals of that fit are above a secondary fitting threshold and then may perform another fit. On the other hand, the RE algorithm uses a "global" fitting approach, where every channel in a square around the above-threshold channel is fit, regardless of the adjacent traces' amplitudes. The fits are permitted to be negative. This difference will have implications discussed in subsection 5.6.1.

4.4 Timing corrections

In order to perform precession frequency fits using 1296 independent channels, they must be precisely aligned in time. Several timing corrections are required. Ideally, $t = 0$ would correspond to the time that the beam is injected into the ring and the measurement period begins. This injection time is measured by the T0 detector, as discussed in section 2.4. As the beam passes through the T0 scintillator, the time profile

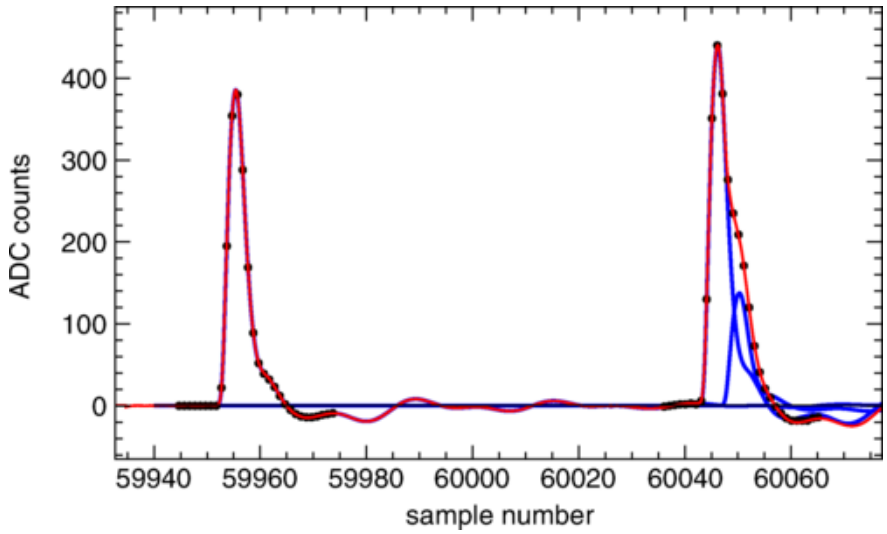


Figure 4.5: Sample template fit showing chain fitting across multiple islands. The black points are the digitized samples, and the red curve shows the full fit result. Three underlying pulses in blue have been fit to the data. The later pulse shape shows a double pulse fit, resolving overlapping pulses 5 ns apart. Reproduced from [3].

of the incoming beam is measured. The weighted average time of that pulse is defined as the start of the measurement period.

In order to time align the T0 detector time with 1296 calorimeter channels, each channel, plus the T0 detector itself, receives a synchronization laser pulse at the beginning of each fill. The location of the sync pulse relative to the beam arrival is shown in Figure 4.6. The in-fill laser pulses used for gain corrections are also shown in this diagram.

The pulse time can then be corrected so that injection is at $t = 0$ for all channels [70]:

$$t_{corr} = t_{raw} - t_{\text{sync}} - (t_{\text{T0, beam}} - t_{\text{T0, sync}}), \tag{4.1}$$

where t_{corr} is the corrected time, t_{raw} is the uncorrected time, t_{sync} is the time of the sync pulse in that channel, $t_{\text{T0, beam}}$ is the time of the beam pulse in the T0 detector, and $t_{\text{T0, sync}}$ is the time of the sync pulse in the T0 detector.

However, this simple correction assumes that the time between beam injection and the sync pulse time is the same for all channels [63]. This assumption is not correct, as the six different laser heads may have slightly different timing, and the optical fibers bringing laser pulses to the channels have different lengths.

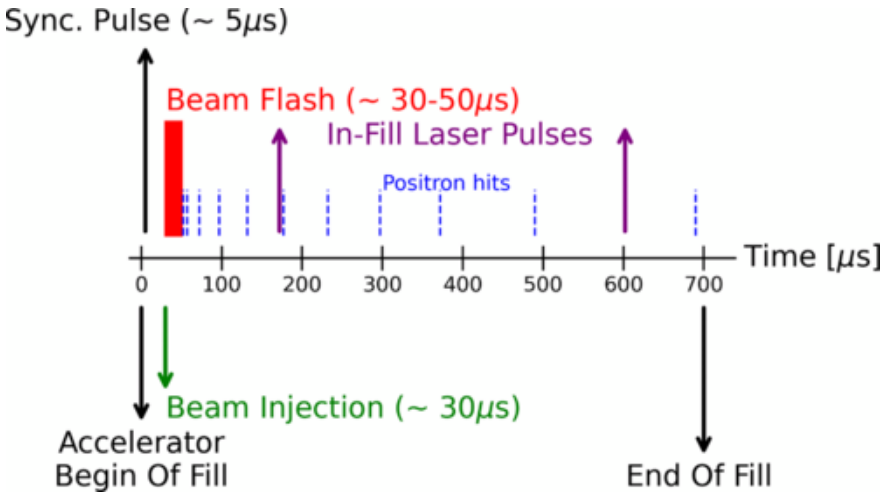


Figure 4.6: Timing of the laser synchronization pulse relative to the beam injection and the rest of the fill. The timing of in-fill pulses used for gain corrections is also shown. Reproduced from [3].

These factors affect the timing of the sync pulse arrival [63, 70]. Therefore, two additional sets of timing corrections are required: one intra-calorimeter and one inter-calorimeter. Intra-calorimeter timing corrections are generated by aligning the peak distribution of positron pulses arriving in neighboring channels. When a cluster contains multiple high energy crystal hits, the time difference between crystal hits is plotted. To perform the time correction, the time difference is set to zero for horizontally and vertically adjacent channels. Sample intra-calorimeter timing corrections for a single calorimeter are plotted in Figure 4.7, showing three groups corresponding to three lengths of optical fibers [70].

Inter-calorimeter corrections are generated using lost muons, which were discussed in subsection 3.2.5. Destructive detectors are inserted to scatter the beam and generate many lost muons. Coincident hits in three adjacent calorimeters are used to align these calorimeters with respect to each other in time [70].

The time correction in Equation 4.1 then becomes

$$t_{corr} = t_{raw} - t_{\text{sync}} - (t_{\text{T0, beam}} - t_{\text{T0, sync}}) - \delta_{\text{SiPM}} - \delta_{\text{calo}}, \tag{4.2}$$

Where δ_{SiPM} and δ_{calo} are the intra- and inter-calorimeter corrections, respectively.

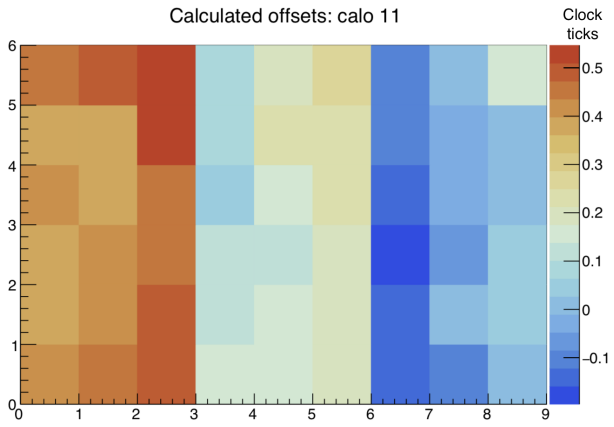


Figure 4.7: Intra-calorimeter time corrections for a sample calorimeter. The three blocks correspond to three different optical fiber lengths delivering the laser sync pulses. Differences are ≈ 0.3 c.t. Reproduced from [70].

4.5 Calibration

Energy calibration is required to convert pulse integrals into consistent energy units. Lost muons are also used for energy calibrations. Since lost muons are minimum ionizing particles (MIPs) and deposit a consistent amount of energy in each calorimeter they pass through, the energy peak for coincidences in 3 adjacent calorimeters is plotted and fit to a Gaussian function, as seen in Figure 4.8. For each channel, the value of the fitted MIP energy peak is defined as 170 MeV in order to align all channels in energy [70]. This value is reasonable based on calculations from the Particle Data Group [15]. However, it is ultimately arbitrary, and the energy threshold for fitting is determined based on a scan to optimize the statistical precision on ω_a.

4.6 Gain correction

Gain is defined as the ratio of a sensor's output to its input. For the calorimeters, a positron of a certain energy generates an output signal of a certain size in the SiPM sensors. The output is then converted to useful energy units by calibration procedures. Ideally, this ratio remains constant in time so that the calibration remains accurate. However, the gain of the calorimeters can vary with effects such as temperature and hit rate. In particular, SiPM sensors are very temperature sensitive, and seasonal or diurnal temperature changes cause the gain to vary. Figure 3.3 shows an example of SiPM gain tracking temperature deviations before the additional hall cooling was installed to maintain consistent temperatures. Further, positron hits induce a

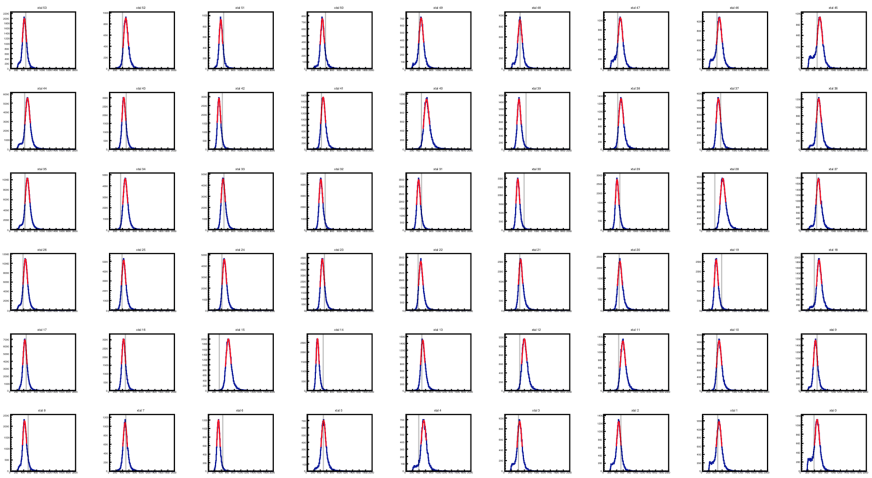


Figure 4.8: Minimum ionizing particle (MIP) peaks in 54 channels of a calorimeter. Each peak is fit to a Gaussian function. In order to calibrate the channels with respect to each other, the peak of each fit is set to 170 MeV. Reproduced from [70].

gain sag in the detectors, so that a positron arriving shortly after another positron will be measured with a smaller signal. Gain changes are particularly important to calculate because they can cause "early-to-late" effects as discussed in subsection 3.2.1 if they are related to the rate of positrons hitting the calorimeters.

The laser system discussed in section 2.7 supplies laser pulses to each of the calorimeter channels so that gain can be corrected. Since the laser pulses are a constant size, they can be used to normalize the detector response by comparing the laser pulse measured during data taking to a reference laser pulse. Three types of gain correction are required at different time scales [72]:

- Short-term double pulse (STDP) [ns-level]: This gain sag occurs when two pulses hit the same SiPM in short succession. Because the SiPM has a finite number of pixels, each of which has a finite recovery time, the second pulse will be measured with a lower energy than the first pulse. The STDP correction is extracted from special runs where two lasers fire pulses within a few ns of each other to map out the SiPM recovery time following the earlier pulse. The gain function calculated for a single channel is seen in Figure 4.9 (left) showing a lifetime of 15 ns and a gain drop of 11.3% [3].

- In-fill gain (IFG) [μs-level]: This gain sag occurs across the calorimeter as a result of the intense beam splash at the start of the fill. The average lifetime of this correction is a few μs, corresponding to the recovery of the calorimeter electronics. The IFG correction is measured using laser pulses fired every 11 fills during normal data taking . The approximate timing of these pulses is shown in Figure 4.6.

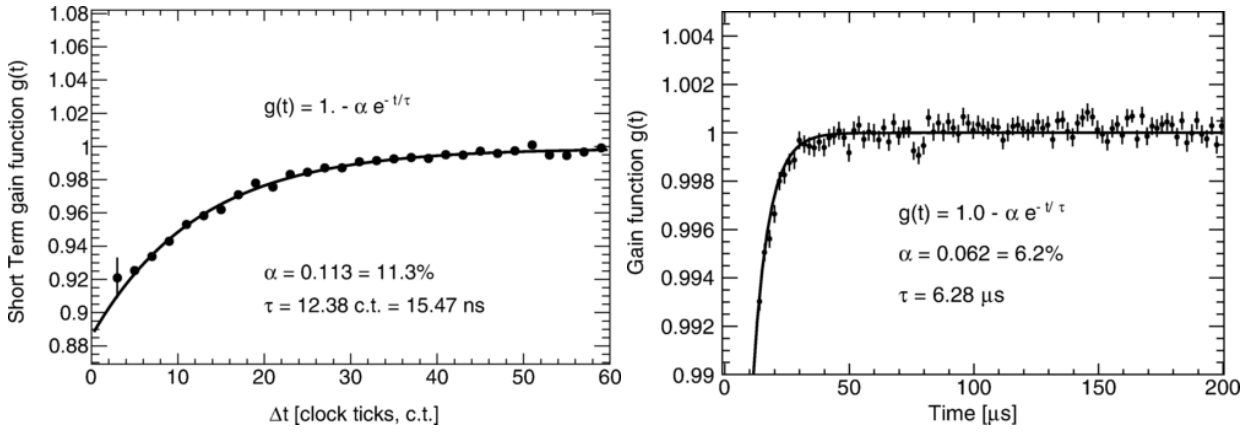


Figure 4.9: Left: sample short term double pulse (STDP) gain drop as a function of time in a single channel showing a 15 ns recovery time following an initial pulse. Right: sample in-fill gain (IFG) drop as a function of time in a single channel showing a 6 μs recovery time. The precession frequency fit starts at 30 μs. Reproduced from [3].

These fills are currently excluded from the physics analysis so that the positron hits are not perturbed by the laser pulses. The IFG function for a single channel is shown in Figure 4.9 (right) showing a lifetime of 6 μs and a gain drop of 6.2% [3]. Note that the fit is started at 30 μs to avoid the large gain sag as the detectors are bombarded with particles at the beginning of the fill. The IFG drop has the largest effect on the precession frequency analysis.

- Out of fill (OOF) [long term]: This correction concerns long time scale gain drifts, which are primarily due to temperature. Laser pulses at the end of every muon fill are compared to laser pulses from a reference run. This correction is generally less significant to the precession frequency analysis, because it does not occur coherently on a muon fill level.

The systematic error due to gain corrections will be discussed in more detail in section 6.2. A further gain-like effect due to details of the reconstruction algorithm will be discussed in subsection 5.6.1.

4.7 Clustering

At this point in the reconstruction, crystal hits have meaningful time and energy information. However, the time and energy of the decay positron has still not been reconstructed, because one positron could generate multiple crystal hits. Clustering algorithms convert a list of crystal hits into clusters, where ideally a single

cluster corresponds to a single positron. These data products enter the precession frequency histogram from which the physics information is extracted. In general, the goal of a clustering algorithm is to group crystal hits as correctly as possible while not introducing time dependent behavior. That is, the clustering algorithm should make the same decisions about the same pulses early in the fill, when the positron rate is higher, and later in the fill, when the positron rate is lower. Several clustering algorithms are in use by the collaboration, associated with the Recon East and Recon West reconstruction algorithms. RW only uses time partitioning (crystal hits are only grouped by their proximity in time), whereas RE also uses spatial partitioning (crystal hits closer together are more likely to be grouped into a cluster than those far apart)

First, the nominal RW clustering algorithm created by A. Fienberg for the Run-1 analysis [63] will be described. Because this algorithm was associated with a relatively large systematic uncertainty for reasons discussed below, an improvement to this basic algorithm was developed by the author of this thesis for the Run-2/3 analysis. General characteristics and goals of clustering algorithms that informed this improvement will be discussed. Finally, the implementation and results of the updated clustering algorithm will be described.

4.7.1 Time-based clustering algorithm

The simplest clustering algorithm divides crystal hits into clusters using time only, without taking spatial information into account. When a positron hits the calorimeter, it induces an electromagnetic shower over a relatively short time scale. Clustering by time partitioning has the advantage of simple and predictable behavior: the algorithm will generally make consistent choices regardless of the time in fill, and no assumptions are made about the spatial distribution of positron hits.

The nominal or "Δt" algorithm is outlined in Figure 4.10. The clustering module takes two input parameters, the low and high time cutoffs, nominally set to 2 and 3 clock ticks (c.t.) respectively, where 1 c.t. $\approx$ 1.25 ns. First, a list of crystal hits is ordered in time. Looping through that list, each hit is either included in the previous cluster or begins a new cluster based on a comparison between its time and the running time of the existing cluster. The running time is an energy-weighted average of the crystal hit times in the existing cluster. If the time difference between the crystal hit time and the running time is less than the low time cutoff, it is always included in the previous cluster; if the time difference is greater than high time

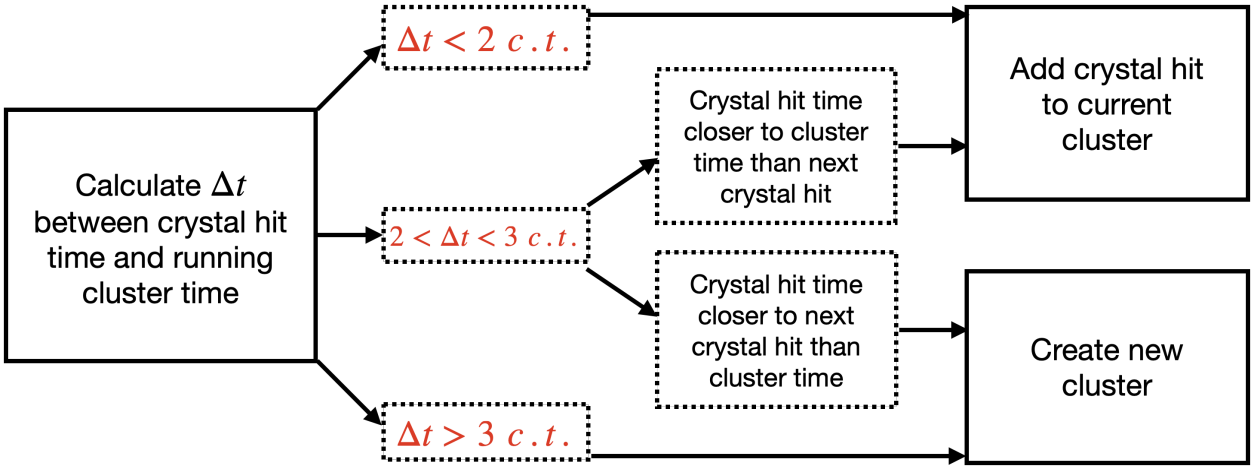


Figure 4.10: Flowchart of standard Recon West clustering algorithm that uses the time difference between hit time and cluster time to separate crystal hits into clusters. 1 c.t. (clock tick) $\approx$ 1.25 ns.

cutoff, it always begins a new cluster; if it is in between, it is included in the previous cluster if it is closer to the running time than the time of the next crystal hit in the list. Hits in the same crystal are combined after the clustering. At the end of the loop, the official "cluster time" is set to be the time of the crystal hit with the highest energy in the cluster.

This algorithm was used effectively in the Run-1 analysis, as discussed in [3] and [63]. Because of its relatively large time window, it errs on the side of simple behavior rather than perfect correctness.

4.7.2 Clustering and pileup

The primary clustering error of concern is pileup, which is when multiple positrons are accidentally reconstructed as a single positron. In the case of time-based clustering algorithms, this error is easy to imagine. If two positrons arrive in the same calorimeter within 2 c.t. of each other, they will be clustered together and enter the precession frequency histogram as a single positron. This type of mis-reconstruction results in assigning an incorrect phase to the reconstructed positron, as two low-energy positrons will generally have a different phase than a single high-energy positron. A schematic of the phase bias resulting from pileup is shown in Figure 4.11. The lower energy blue calorimeter hits will generally originate from positrons with a different phase than the higher energy orange calorimeter hit. If the blue hits are mis-reconstructed as a single higher energy hit, they will enter the precession frequency histogram with the incorrect phase. Pileup is also a rate-dependent effect, with more pileup early in the fill when more positrons are arriving at the

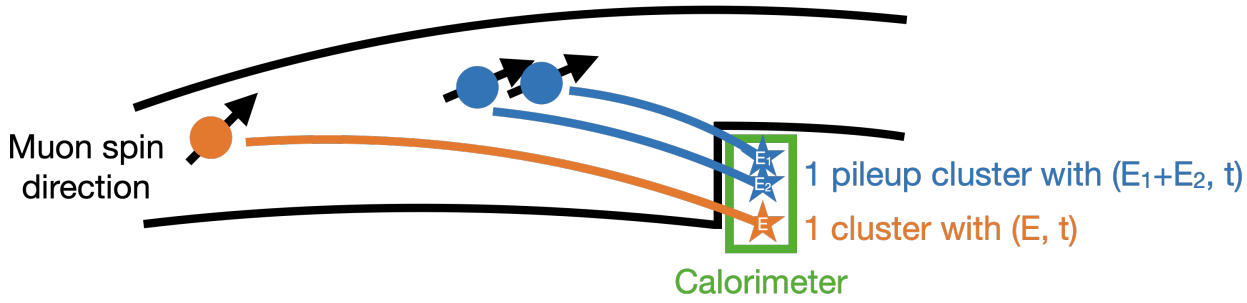


Figure 4.11: Schematic of a pileup event (in blue) measured in a calorimeter compared to a single event (in orange). The blue positron hits will generally originate from muons with a different spin phase than the orange positron hit, resulting in the pileup event having the wrong phase for its energy.

calorimeters. Therefore, pileup is considered an "early-to-late" type effect as discussed in subsection 3.2.1 and can significantly bias ω_a if left uncorrected. The amount of pileup present in an energy spectrum can be determined by looking for events above 3.1 GeV, since decay positrons cannot have a larger energy than the original muon.

The goal for the pileup uncertainty over the lifetime of the experiment is 40 ppb [7]. If pileup remains uncorrected, the shift to ω_a can be hundreds of ppb [63]. Therefore a pileup subtraction routine must be applied to remove as much pileup as possible from the histogram before fitting. In general, pileup subtraction routines generate a pileup spectrum from data and then subtract that from the uncorrected histogram. The corrected spectrum is then used for the physics analysis.

For Run-1, the uncertainty due to pileup in the UW analysis was around 50 ppb [63]. This was due both to the relatively high amount of pileup resulting from the reconstruction algorithm and the details of the pileup subtraction. Since this was above the final goal uncertainty of 40 ppb, it motivates improvements to the clustering and pileup procedures for Run-2 and beyond. In general, reducing pileup in the clustering algorithm will reduce the uncertainty due to pileup, as the fit will be less sensitive to changes in the pileup correction. Further details of the pileup correction, including why it was updated for Run-2/3, will be discussed in subsection 5.2.2.

The following sections will discuss options for improving the clustering algorithm to reduce pileup and discuss the implementation of the algorithm used for the final analysis.

4.7.3 Spatial separation and false pileup

Because the calorimeters are segmented, clustering using spatial separation of crystal hits is an obvious option that is used by some groups. This strategy depends on the assumption that crystal hits grouped close together in space are likely to originate from the same decay positron, whereas crystal hits separated in space are likely to originate from different positrons. By incorporating spatial information into the clustering, spatially separated groups of crystal hits would be separated into different clusters.

In order to test the validity of this assumption, this author performed a study on spatially separated events. The goal of the study was to determine whether clusters that are spatially separated (i.e. clusters containing crystal hits from opposite sides of the calorimeter) are likely to originate from different positrons, as previously assumed.

To study the population of spatially separated events, a spatially separated event is defined as having at least 3 crystal lengths between the highest energy crystal and the farthest crystal from that. The "primary" cluster is defined as a 3×3 square around the highest energy crystal hit, and the "secondary" cluster is defined as a 3×3 square around the farthest crystal hit from the highest energy crystal. Sample spatially separated events are shown in Figure 4.12, with the primary and secondary clusters outlined in red and blue, respectively. Note that each of these events would be clustered together using the nominal RW clustering algorithm.

True pileup events are expected to

- Not be time coincident with the rest of the hits in the cluster, and
- Disappear late in time when event rates are low.

For each spatially separated event, the times of the primary and secondary crystal hits are separately averaged. A histogram of these average times is shown in Figure 4.13, where each entry in each histogram represents a single nominal cluster. The primary cluster distribution is relatively narrow, with a width of < 1 c.t. Under the assumptions described in the previous section, the secondary cluster would be expected to originate from a different positron than the primary cluster and therefore not be time coincident with it. Contrary to expectations, the double humped feature between -1 and 1 c.t. suggests that these secondary clusters are time coincident with the primary cluster. (The double humped shape is due to the fact that the

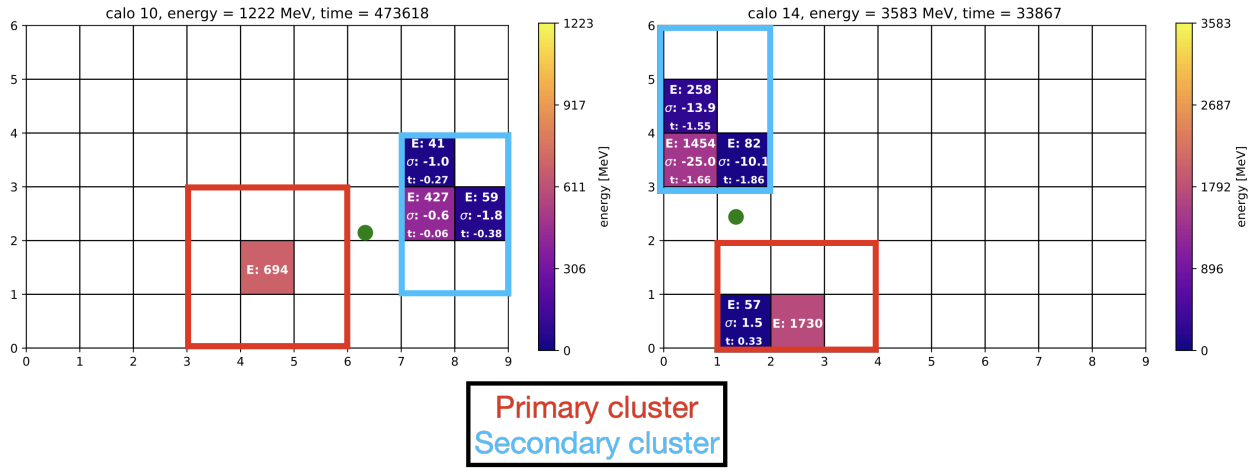


Figure 4.12: Sample spatially separated event displays showing distribution of hits in a calorimeter. The primary and secondary clusters are marked in red and blue, respectively. For the highest energy crystal hit, only the energy in MeV is given; for the others, the energy, $\Delta t'$ (labeled as σ), and Δt (labeled as t) with respect to the highest energy crystal hit are given.

primary cluster contains the crystal that is determining the cluster time, so a section of the distribution has been removed.) Underlying the time coincident structure is a flat non-coincident background. Late in time, the flat background has decreased, but time-coincident secondary clusters remain.

This suggests that the early time distribution contains both a population of true pileup events (the flat background, highlighted in gray) and a much larger population of time coincident "false" pileup events, where false pileup is here defined as an event originating from the same positron with spatially separated crystal hits; that is, an event which would be considered two different positrons if spatial separation were applied. These time coincident secondary hits are presumably induced by the same positron as the primary hits, meaning that ideally they would not be separated into two clusters. Notably, the secondary cluster time distribution is wider than the primary cluster time distribution, although narrower than the window allowed by the RW clustering algorithm.

The percentage of spatially separated events that are true pileup as a function of time was estimated by taking the integral of the flat true pileup region from -2 to 2 c.t. and subtracting it from the integral of the whole histogram in that range. In Figure 4.13, the true pileup is indicated by the shaded gray region, with the integration bounds marked with vertical dashed gray lines. The percentage of spatially separated events that are true pileup is plotted as a function of time in Figure 4.14, and an exponential decay is applied. Even at early times, when true pileup is highest, only about 16% of spatially separated events are true pileup, and

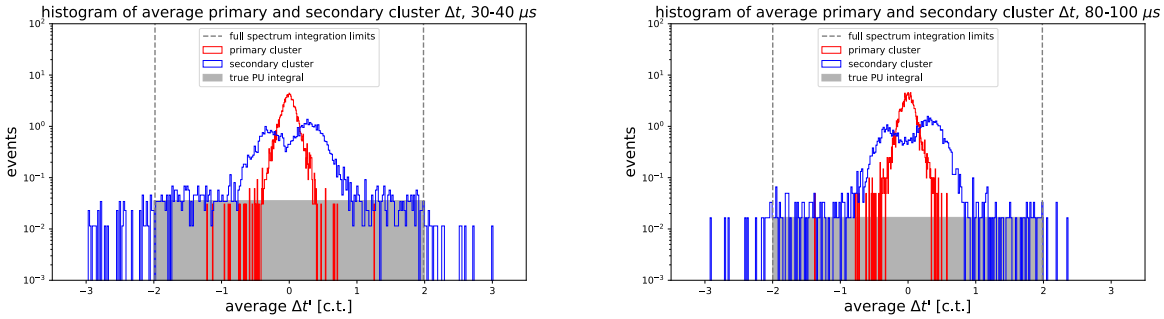


Figure 4.13: Time difference distributions of the primary cluster in red and secondary or spatially separated cluster in blue. The secondary cluster distribution shows false pileup, where a part of the distribution is time coincident with the primary cluster. The gray region represents the flat underlying true pileup distribution, which is decreasing as a function of time.

that number decreases exponentially.

To understand a possible mechanism behind these false pileup clusters, the percent of clusters after 250 ns that are spatially separated was plotted as a function of calorimeter number (Figure 4.15). All of these clusters were measured at late times, when the true pileup rate is low, so all spatially separated clusters should be false pileup. This data suggests that around 1.5% of all events are false pileup. Second, the points have been color coded in accordance with whether that calorimeter is behind a tracker, kicker, or ESQ plate. The association between material in front of the calorimeter (particularly trackers) and number of false pileup events suggests that positrons hitting material may kick out another particle that hits the calorimeter in spatially separated location from the primary positron hit. Similar events have been found in simulation.

Because the vast majority of spatially separated events are false pileup, and because only true pileup should be separated in the reconstruction algorithm, the decision was made to not utilize spatial separation for this analysis.

However, this study did suggest that improved time discrimination could decrease pileup. All the clusters entering the histograms in Figure 4.13 are "nominal" Δt clusters with an allowed Δt range of 2-3 c.t. Because hits originating from the same positron were within $\pm$ 1 c.t. of the cluster time, the time discrimination can be improved to reduce the amount of pileup.

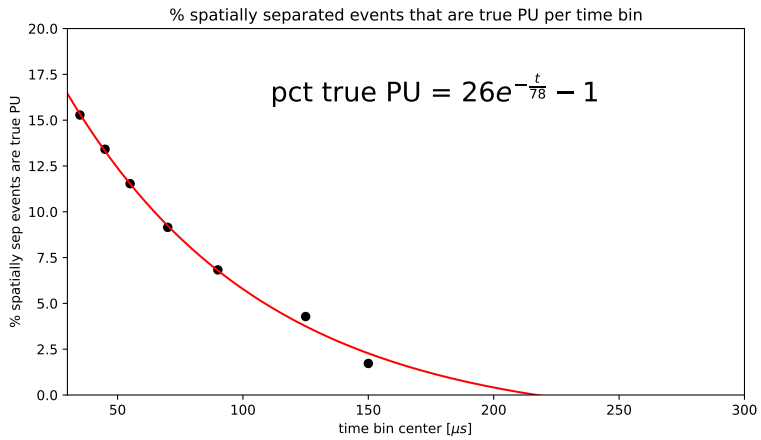


Figure 4.14: Percentage of spatially separated events that are true pileup as a function of time in the fill, as determined by taking the ratio of non-time coincident events to all events. At early times, about 16% of spatially separated events are true pileup.

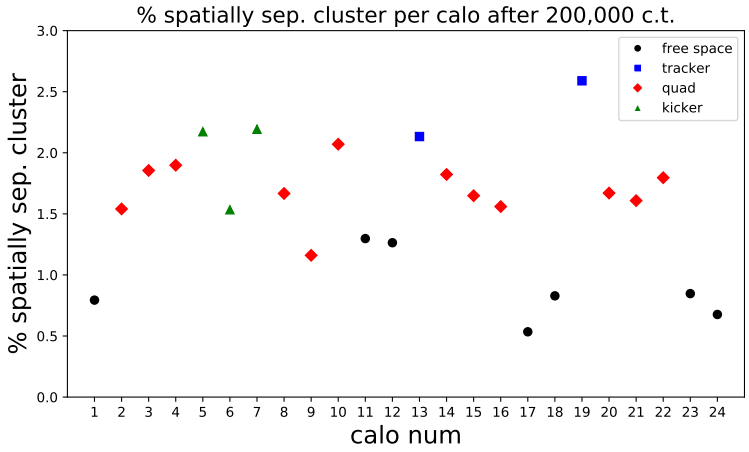


Figure 4.15: Percentage of spatially separated clusters as a percentage of all clusters vs calorimeter number. Points are color coded based on whether that calorimeter is behind a tracker, electrostatic quadrupole plate, kicker plate, or nothing. A higher rate of spatially separated clusters is measured in calorimeters behind material.

4.7.4 Energy-dependent time resolution

This section discusses a method of improving the time discrimination by incorporating the energy dependence of the calorimeter channel time resolution, which was measured by early calorimeter tests at the Stanford Linear Accelerator Center (SLAC) test beam [65]. At SLAC, a prototype calorimeter received laser pulses of different energies, and the time resolution was measured by histogramming the time difference between the same laser hit arriving in different channels. The time difference histogram was fit to a Gaussian and the standard deviation of the fit, $\sigma_{\Delta t}$, was extracted. Figure 4.16 (left) shows the standard deviation as a function of effective energy, which takes into account the energy E_1 and E_2 of the pulses being compared and is defined as

$$E_{\text{eff}} = \frac{E_1 E_2}{\sqrt{(E_1^2 + E_2^2)/2}}. \tag{4.3}$$

The blue and red curves in the plot correspond to whether the compared pulses were measured in the same or different waveform digitizers (WFD5s). The timing measurement of crystal hits in the calorimeters is affected by electronics effects, statistical fluctuations of photons, and the physics of the particle shower induced by the particle. Electronics jitter causes the time difference distribution between channels digitized by different digitizers to be wider. The standard deviations $\sigma_{\Delta t}$ as a function of E_{eff} were fitted to

$$\sigma_{\Delta t}(E_{\text{eff}}) = \sqrt{2C_T^2 + \frac{S_T^2}{E_{\text{eff}}/\sigma_n}}, \tag{4.4}$$

where C_T is a constant term coming from electronics noise, S_T is the stochastic term related to photostatistics, and σ_n is the noise term, set to 5 npe (number of photoelectrons) due to digitizer noise level. Energy in the SLAC analysis was measured in npe.

In summary, calorimeter tests performed at the SLAC test beam demonstrated that time resolution is energy-dependent. The timing of low energy pulses in the calorimeter crystals is known less precisely than the timing of high energy pulses. This analysis did not take into account the time spread of a positron shower, since it used laser pulses only. However, it provides a starting point to motivate incorporating the time resolution information into physics analyses.

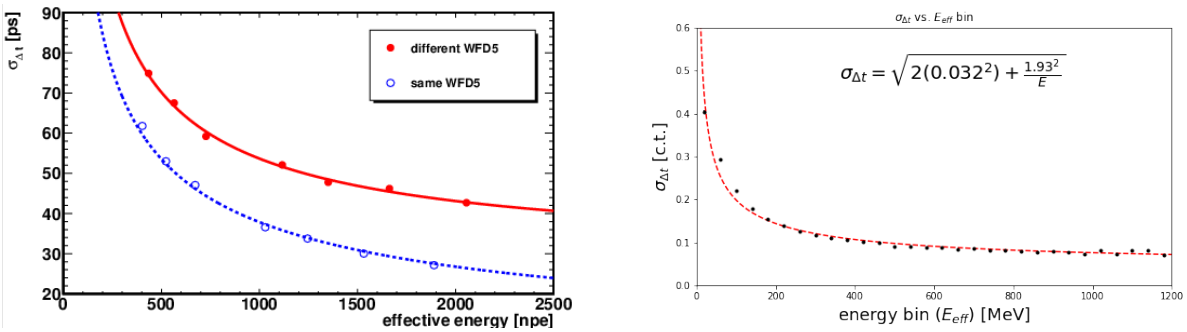


Figure 4.16: Left: Standard deviation of the time difference distribution between laser pulses measured in different channels, depending on whether they are digitized from the same or different waveform digitizers (WFD5s), as measured at the Stanford Linear Accelerator Center (SLAC) test beam. Reproduced from [65]. Right: similar pseudo-time resolution curve generated from real beam data. A fit to the data is shown using the same function form as on the left. In general, time resolution improves as a function of energy.

Time resolution calculation using beam data

A comparable empirical timing resolution vs. energy function can be extracted from real beam data. In order to incorporate information about timing resolution into the clustering algorithm, the empirical time spread of crystal hits from a single positron can be measured in different effective energy bins. Notably, this will incorporate both statistics and electronics effects, as measured in the SLAC test beam, and the spread in times induced by the physics of the positron shower. Therefore it is not strictly a time resolution measurement.

For a given cluster, the effective energy and time difference were calculated between each crystal hit and the maximum energy crystal, which determines the cluster time. These time differences were added to a histogram for that E_{eff} bin. By looping over many clusters, a set of distributions of time differences was created for each energy bin. A Gaussian fit was applied to each energy-binned histogram. The example distributions in Figure 4.17 show a wider time distribution at early times than at later times, as expected.

As mentioned previously, spatially separated crystal hits in a "false pileup" event, which are presumably induced by a particle created when the positron passes through material, have a larger time spread than the crystal hits induced by the positron itself. This slightly wider time distribution should still be grouped together as a single cluster in order to accurately reconstruct the decay positron energy.

A plot of the standard deviation of the Gaussian distribution versus the central value of the E_{eff} bin is shown in Figure 4.16 (right). The data are fit to Equation 4.4 to replicate the study performed on the test

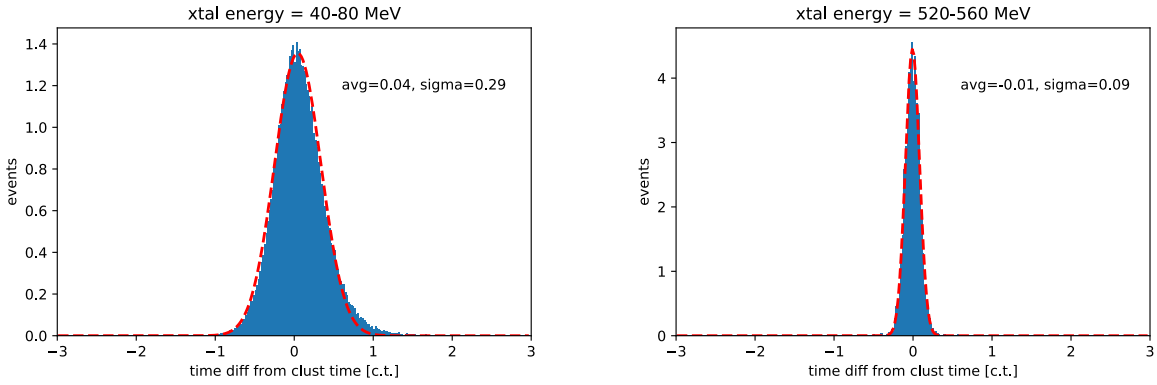


Figure 4.17: Sample time difference histograms binned by effective energy showing that the distribution is narrower, and therefore the time resolution is better, at higher energies. Time differences for each crystal hit in a cluster were taken with respect to the cluster time. The left plot shows the 40-80 MeV bin, and the right plot shows the 520-560 MeV bin. The standard deviation of each distribution is also shown.

beam data shown on the left. S_T^2 and σ_n are combined into a single parameter, as σ_n is not known in this case. The fit is not ideal in lower energy bins, possibly due to the asymmetric tail visible in the low energy distribution. However, the functional form is still sufficient to incorporate the time resolution information into the clustering algorithm.

Implementation

The energy-dependent time resolution can then be incorporated into an energy-dependent metric to discriminate clusters. The metric

$$\Delta t'(E_{\text{eff}}) \equiv \frac{\Delta t}{\sigma_{\Delta t}(E_{\text{eff}})} \tag{4.5}$$

replaces the original Δt metric in the clustering algorithm (Figure 4.10), where $\sigma_{\Delta t}(E_{\text{eff}})$ in the denominator was extracted in the previous section. Essentially, $\Delta t'$ is the number of standard deviations that the two pulses are from each other in time, including information about how precisely those two times are known. For example, returning to the sample events in Figure 4.12, the energy, $\Delta t'$, and Δt are shown for each channel, in that order. In the left event display, the secondary cluster is time coincident with the primary cluster; both Δt and $\Delta t'$ are relatively small, suggesting that this is a false pileup cluster. In the right event display, although the secondary cluster Δt values are within a 2 c.t. window of the cluster time, the $\Delta t'$

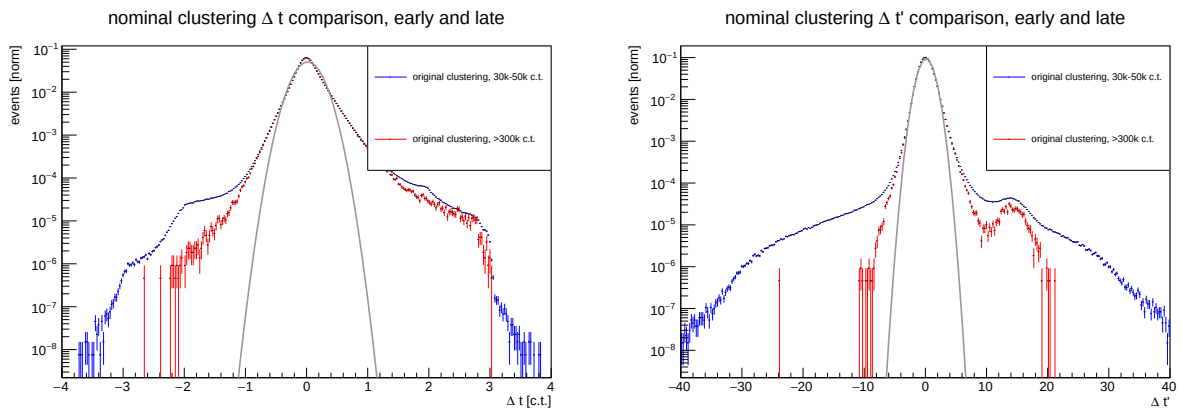


Figure 4.18: Distribution of crystal hits with respect to the cluster time early and late in the fill using the usual time difference metric (left) and the time difference weighted by energy resolution (right). The bump at $\approx 15\Delta t'$ is due to a fitting error discussed in the text.

metric makes it clear that these crystal hits are many standard deviations outside of the cluster time. This, combined with the energy above 3.1 GeV, suggests that this is a true pileup cluster.

The utility of this metric to distinguish pileup events is demonstrated by Figure 4.18. The plot on the left compares the Δt of crystal hits in a cluster early in time, when there is pileup, and late in time, when there is no pileup. The plot on the right shows the distribution of $\Delta t'$. In an ideal clustering algorithm, cuts would be applied so that the early time distribution resembles the late time pileup-free distribution. These cuts are easier to define using the $\Delta t'$ distribution. A Gaussian fit has been applied to the $\Delta t'$ distribution. The underlying distribution is slightly wider than a Gaussian due to the time structure of a positron shower.

The prominent bump at $\approx 15\Delta t'$ is due to a fitting error wherein the fitter decides to fit two pulses to a single positron trace. Figure 4.19 shows an example of this situation. For occasional high energy pulses, the template does not accurately fit the falling tail of the trace, and the fitter prefers a second low energy pulse 2-4 c.t. later in time. The vertical gray lines represent the times of the two pulses. This fitting error could be evidence of nonlinearities in the pulse shape as a function of energy. Because this fitting error is not time-dependent to first order, it is disregarded when determining the best selection window for $\Delta t'$ (i.e., the low energy secondary pulse will be split off). This will cause a slight energy reconstruction error of order 50 MeV to the high energy pulses but does not cause any major problems.

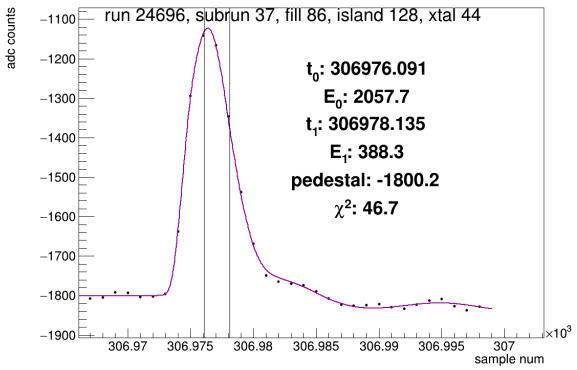


Figure 4.19: Example of fitting error where two pulses were fit to a single positron hit. The gray lines mark the times of the two pulses.

Determining the clustering window

The ideal clustering window would:

- Distort the energy spectrum as little as possible, keeping in mind that our current "nominal" reconstruction algorithm is not fully correct.
- Reduce pileup as much as reasonably possible. This may be prioritized over a reconstruction accuracy as long as no early-to-late effects are introduced.

A variety of clustering windows were tried. The distributions of $\Delta t'$ early and late in time are shown in Figure 4.20. All clustering windows are successful in reducing the pileup tails visible at early times. At late times, the more restrictive windows cut into the pure singles distribution, suggesting they may be too tight. Note that the high $\Delta t'$ bump caused by a fitting error will be excluded by all the possible windows.

The energy spectrum with the different windows can be seen in Figure 4.21 (top left). Compared to the original clustering algorithm, all windows reduce the pileup, as measured by the number of events with nonphysical energies (above 3.1 GeV). The pileup reduction factor, quantified by the ratio of integrated events above 3.1 GeV in the nominal clustering method to integrated events in the $\Delta t'$ clustering method, is listed for each time window in Table 4.1. Even the loosest window reduces pileup by more than a factor of 3.

However, it is important to check that no artifacts have been introduced into the data by artificially

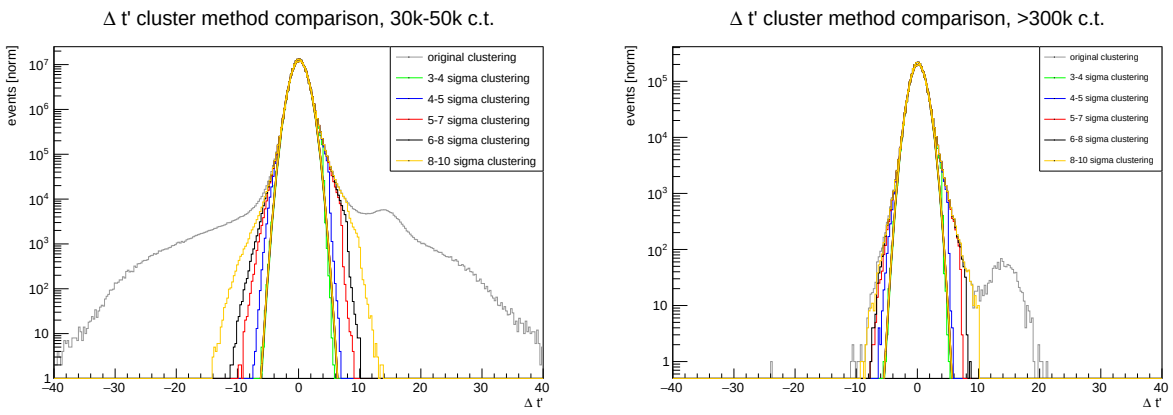


Figure 4.20: Behavior of time resolution-corrected clustering algorithm using different $\Delta t'$ windows at early (20k-40k c.t.) and late ($>$ 300k c.t.) times. A 6-8 $\Delta t'$ window was ultimately chosen for the analysis.

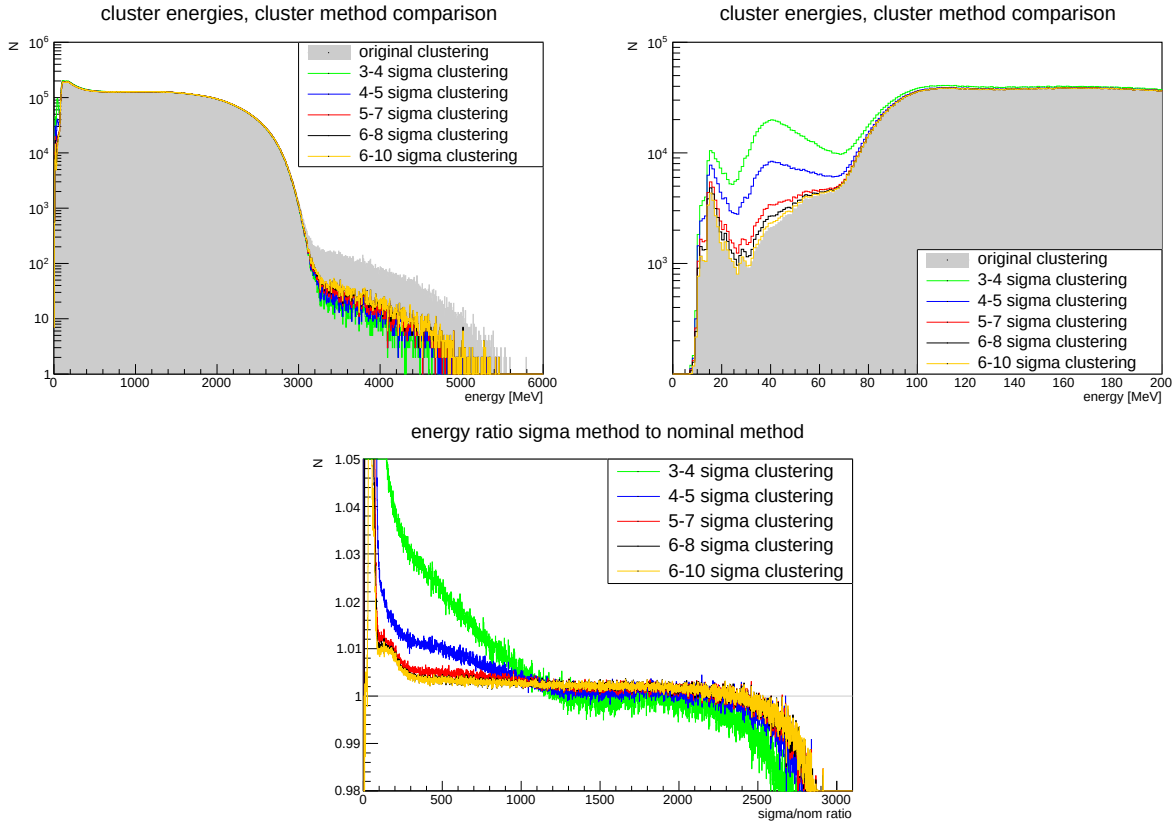


Figure 4.21: Top left: Comparison of the energy spectrum for the different clustering methods showing pileup reduction in the high energy tail. Top right: comparison of the energy spectrum in the low energy region. The more restrictive $\Delta t'$ windows induce a large bump at low energies relative to the original clustering. Bottom: ratio between the different clustering methods and the nominal method.

$\Delta t'$ window	Pileup reduction factor
3-4	5.9
4-5	4.8
5-7	4.0
6-8	3.5
6-10	3.4

Table 4.1: Pileup reduction factors for the different $\Delta t'$ windows, as measured by the ratio of integrals of events with nonphysical energy (over 3.1 GeV).

tightening the clustering window too much. The effect on the low energy region of the energy spectrum is seen in Figure 4.21 (top right). The more restrictive time windows induce a large bump at energies $<$ 100 MeV relative to the original clustering method. Although the original method is wrong at some level, these artifacts are likely due to the clustering algorithm splitting off crystal hits from the main cluster and is an indication that these windows are too restrictive. The less restrictive windows follow the shape of the nominal clustering method more closely.

The ratios between the $\Delta t'$ based clustering methods and the nominal clustering method are shown in Figure 4.21 (bottom). Because a single pileup event is composed of multiple lower energy events, an algorithm that accurately separates pileup should have fewer events in the high energy region and more events in the lower energy region. As expected, all windows have this overall behavior. The more restrictive windows have a strong slope in the intermediate region, probably the result of clustering that is too restrictive. However, the less restrictive windows are relatively flat at intermediate energies, with an increase in events of around 0.3%. This behavior strengthens the case for the use of the 6-8 window for the $\Delta t'$ clustering algorithm.

The 6-8 $\Delta t'$ window was implemented for the final analysis. The functional form extracted from Figure 4.16 is used to calculate the $\Delta t'$ metric for each crystal hit in the list. A $\sigma_{\Delta t}$ cutoff is applied at low energies so that it cannot be higher than the value at 40 MeV in order to avoid divergent behavior.

Event display case studies

In order to compare the behavior of the two clustering algorithms on the same data, the reconstructed energy of the $\Delta t'$ method was compared to the nominal method (Figure 4.22). Clusters within 4 c.t. of each other in the same calo were matched under the assumption that these clusters would have originated from the same

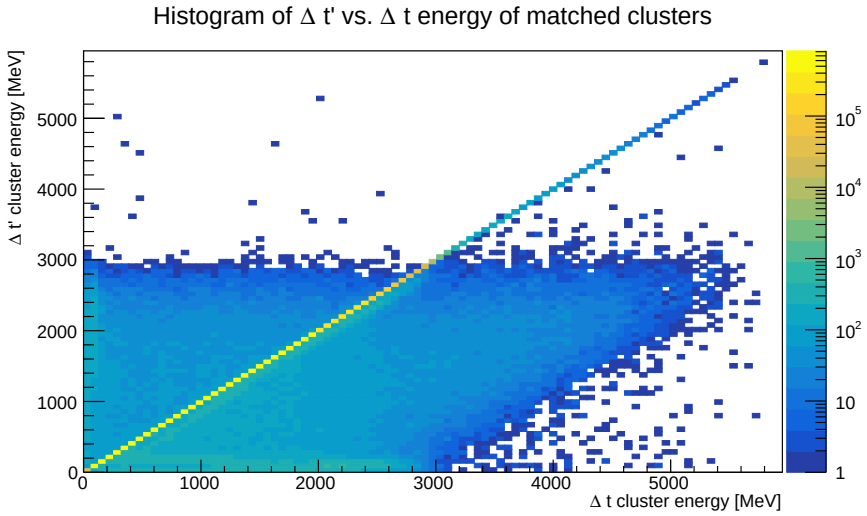


Figure 4.22: Comparison of the energy of clusters reconstructed with the $\Delta t'$ and nominal (Δt) methods. Clusters within 4 c.t. of each other in the same calorimeter were matched.

set of crystal hits.

The diagonal line represents clusters where the same energy was reconstructed with both algorithms. The area below the line represents clusters where the nominal clustering method reconstructed a higher energy than the 6-8 $\Delta t'$ method and vice versa for the area above the line. The former behavior is expected, as the $\Delta t'$ method is more restrictive overall and should be separating pileup events. The pileup-reducing effect of the new method can be seen in the reduction of events above 3.1 GeV.

A sample set of event displays showing the $\Delta t'$ method splitting a single cluster into two clusters is shown in Figure 4.23. The top line of the text in each crystal gives the energy in MeV; the second line the $\Delta t'$ value; and the third line the normal Δt. Both metrics are with respect to the cluster time. The highest energy crystal hit, from which the cluster time is derived, has only the energy labeled. Events that occur before the cluster time have a negative Δt and $\Delta t'$, whereas events that occur after have positive values.

The default clustering method produces a cluster with energy 3.9 GeV (a pileup cluster). The 1.3 c.t. time separation of the upper left cluster indicates that it can be clearly separated from the lower right cluster; this is evident in the bottom two plots. In the energy spectrum, a single 3.9 GeV hit would be correctly replaced with a 2.1 GeV hit and a 1.8 GeV hit.

A more ambiguous case can be seen in Figure 4.24. The original cluster is also a pileup cluster with

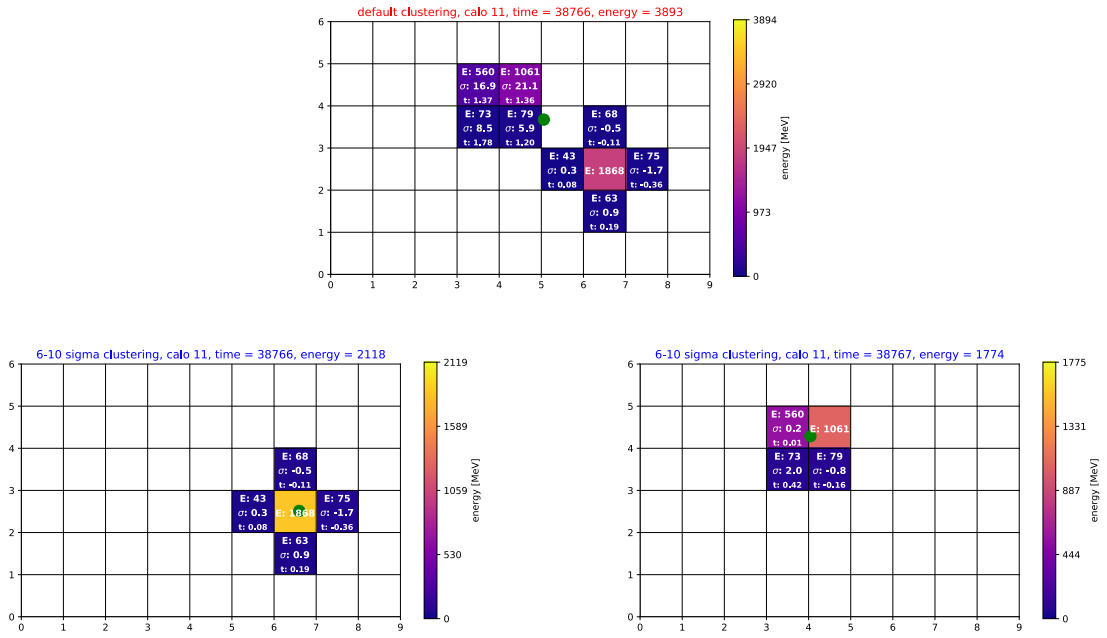


Figure 4.23: Set of event displays representing the $\Delta t'$ clustering algorithm splitting a single pileup cluster from the nominal algorithm (top) into two lower energy clusters (bottom left and right).

energy 4.2 GeV. The two high energy crystals with energy 2051 and 1403 MeV are 0.74 c.t. apart, which is outside the norm for an event created by a single positron. However, because no spatial information is incorporated, because the low energy crystals with energies 52 and 58 are given a lot of leeway in the $\Delta t'$ calculation, and because the 2051 MeV crystal is earlier in time, they end up being included in the top cluster rather than the bottom. This is probably not a physical result based on the time and spatial distributions of the crystal hits. However, the result is still more correct than the original clustering, because a pileup event is correctly split into two clusters.

The events above the diagonal line are less expected, as the new clustering method is more restrictive overall. A sample set of event displays can be seen in Figure 4.25. In this case, an extra crystal hit has been included in the cluster in the $\Delta t'$ case versus the nominal case. It is unclear why this happened, but possibly this crystal hit was closer in $\Delta t'$ to this cluster but closer in Δt to the previous cluster.

In summary, the new clustering method is effective at separating pileup. It does not always make a fully correct decision, but in general it is more correct than the previous clustering. The effect that the new

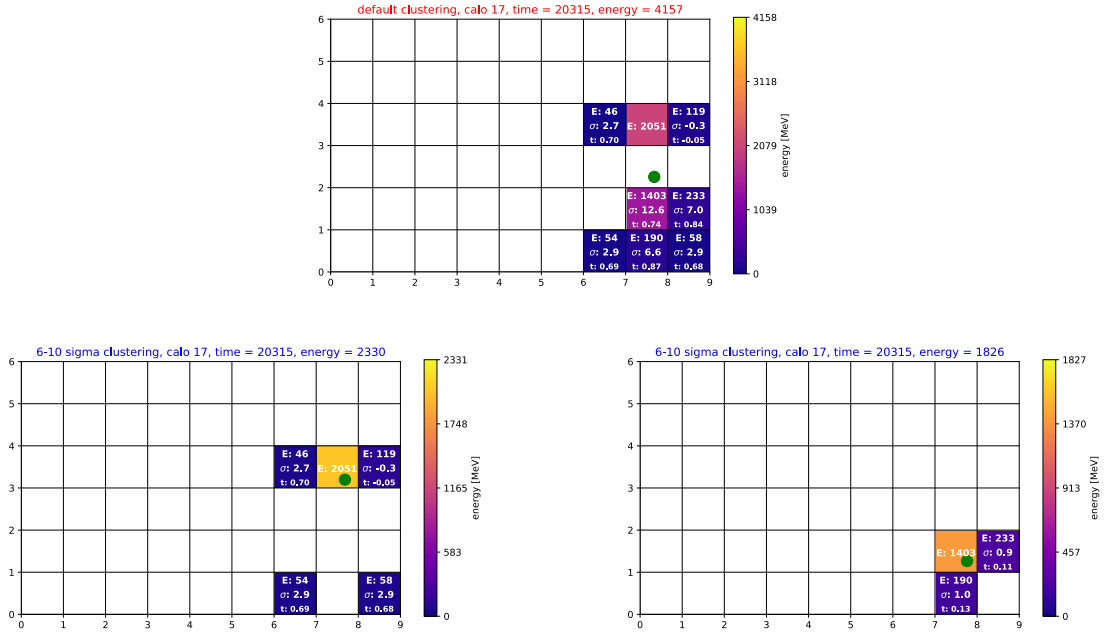


Figure 4.24: Set of event displays representing the $\Delta t'$ clustering algorithm splitting a single pileup cluster from the nominal algorithm (top) into two lower energy clusters (bottom left and right). This case is more ambiguous, as the two low energy crystal hits likely should have been included in the lower right cluster.

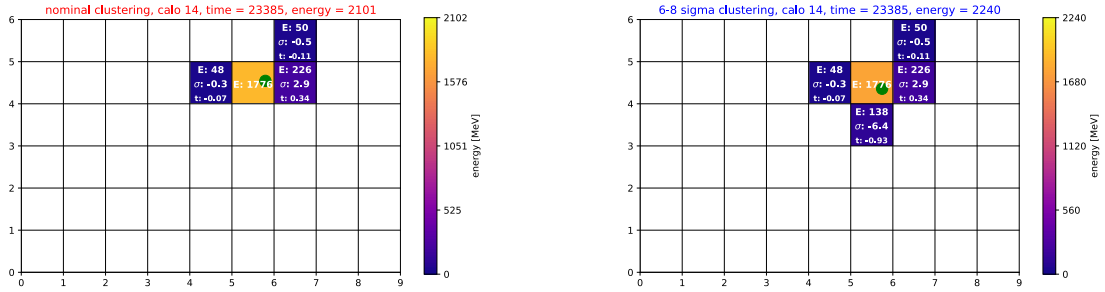


Figure 4.25: Set of event displays representing a situation where the $\Delta t'$ clustering method (right) reconstructs a higher energy than the nominal clustering method (left). An extra crystal hit is included in the cluster.

clustering algorithm has on pileup subtraction and pileup uncertainty will be discussed in subsection 5.2.2.

4.8 Data quality control

The final step in creating physics-quality data is data quality control (DQC). DQC ensures that the data that enters the final physics result is high quality. A high quality fill has functional detector and laser systems, a high intensity incoming beam, and good beam storage. Metrics from calorimeters and auxiliary detectors are used to filter out fills with the following problems:

- Fills that are empty (have no beam at all)
- Fills where insufficient information is present to prepare the data for physics, such as fills that lack laser synchronization pulses
- Fills where the beam delivered to the storage ring is changing or unusual, which could cause subtle problems when stored
- Fills with poor beam storage, such as when the kicker or ESQs spark. This would result in unusual beam dynamics conditions
- Fills in time periods where most fills are failing data quality, indicating a long term problem is likely

DQC is performed both at the fill level, meaning that a metric will be checked every fill, and at the subrun level, where a subrun is about 6 seconds of data. This author implemented a per-fill DQC filter based on metrics from the T0 start time detector, which was described in section 2.4. T0 provides the first and simplest measurement of the beam as it enters the ring, and it provides valuable filtering capabilities. Relevant T0 metrics for DQC are plotted in Figure 4.26, with the DQC cuts in red. The T0 beam integral gives the intensity of the incoming beam, and the DQC cut ensures that empty fills with no beam are removed. In general, the DQC errs on the side of permitting low intensity fills to be included in the final analysis, as they are unlikely to cause any systematic problems. No upper limit is placed on this cut. The T0 beam time cut ensures that the beam is not arriving at an unusual time. Shifts in the timing generally indicate that conditions are changing upstream of the storage ring, which could cause problems for beam storage. The T0 RMS cut measures the overall shape of the beam time intensity profile. Similarly, changes to the beam

shape could indicate changes upstream or systematic studies that were accidentally included in the physics data. The T0 sync integral and sync time measure the intensity and timing of the T0 sync pulse, respectively. This sync pulse is essential for the timing alignment scheme discussed in section 4.4, so a sync pulse must be present.

Additionally, cuts are placed on the number of ctags (a proxy for the number of decay positrons before full reconstruction is performed) and the ratio of ctags to T0 integral, both of which are a measure of storage quality. If few positrons are measured, but the beam intensity is also low, there is no storage problem. However, if few positrons are measured and beam intensity is high, a storage problem, such as an ESQ spark, is likely.

Per-fill DQC is also performed using other systems. Laser cuts ensure that every calorimeter channel has a laser sync pulse, which is necessary for timing and gain corrections. Kicker cuts ensure that the kicker is behaving normally, which affects beam storage. On the subrun-level, cuts are applied on the number of muon losses, which could also indicate a storage problem; the ratio of ctags/T0 integral, with stricter cuts than the per-fill DQC; and on the number of fills passing all DQC cuts. DQC is also performed for the field measurement, and precession frequency and field data is matched so that only data that is high quality for both measurements is accepted. After the DQC has been applied, the data is finally ready to be analyzed for physics.

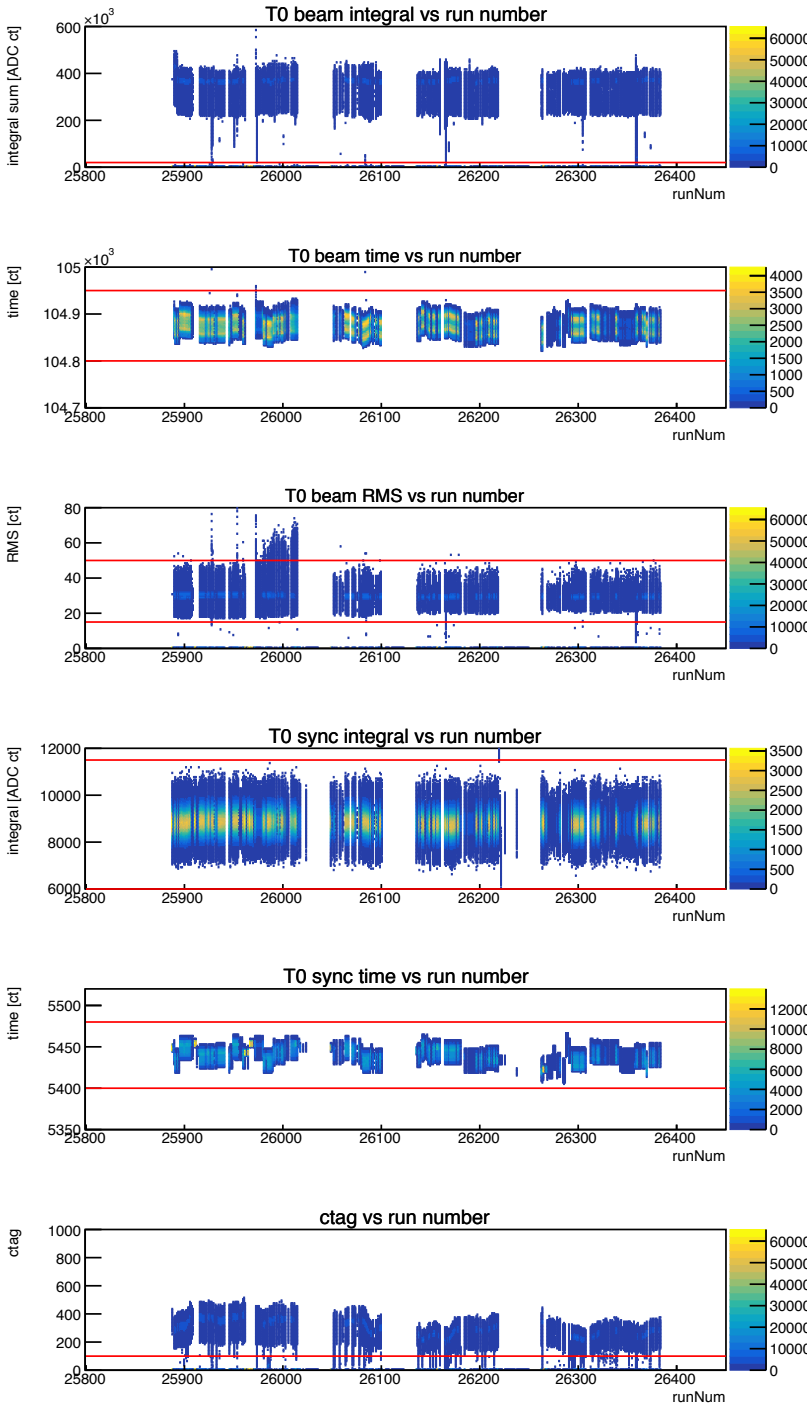


Figure 4.26: Data quality control (DQC) parameters as a function of run number for a Run-2 dataset. Red lines indicate the bounds of the DQC cuts.

Chapter 5

The precession frequency analysis

This thesis presents a precession frequency analysis of the Run-2 and Run-3 data. At the end of the analysis, a value of ω_a will be extracted from a fit to the data. This chapter will discuss the preparation of the data for fitting and the choices made for fitting procedures. Chapter 6 will discuss systematic uncertainty estimates. Other ω_a related systematic uncertainties that are not related to the fitting procedure have been discussed in Chapter 3.

Seven groups are performing analyses of the Run-2/3 data. Different choices are made by each analyzer regarding reconstruction algorithms, pileup corrections, parameterization of the fit function, randomization, and so on. This thesis presents one of set of choices used by this writer. At the time of this writing, the analyses have not been unblinded with respect to each other, nor has the hardware blinding been removed, meaning that the final numbers do not yet contain physics information about ω_a.

5.1 Datasets

In Run-1, the datasets were distinguished by different kicker and ESQ conditions. In Run-2 and Run-3, these conditions were generally consistent across the data taking period. Further, the damaged quadrupole resistors were repaired, resulting in lower muon losses and less beam motion. However, in Run-2, large temperature variations in the hall caused changes in the magnetic field, which affected the behavior of the lost muons, which will be described further in subsection 5.3.2. Between Run-2 and Run-3, hall cooling was improved, resulting in much more consistent behavior of the magnet. Partway through Run-3, the kicker

hardware was upgraded so that it could achieve a sufficient kick to center the beam in the storage region.

The data are fit in 3 large datasets: Run-2, Run-3BM, and Run-3NO. Run-2 had 7 sub-datasets defined for the offline data production (Run-2B through Run-2H), with datasets primarily defined by magnet cycles bracketed by trolley runs, corresponding to periods of changing muon losses. However, data from all of Run-2 was successfully combined and fit as a single Run-2 dataset. Run-3 had 13 sub-datasets (Run-3B through Run-3O, excluding Run-3H), which were divided into two large datasets based on the kicker conditions, with Run-3B through Run-3M having the lower kick and Run-3N through Run-3O having the higher kicker. Full fit results, scans, and random seeds are presented for each of the three larger datasets; the results will later be averaged together. Some results will also be presented for all of the sub-datasets. Run-2 and Run-3 had different hardware blindings so their data cannot be combined. However, in principle data from Run-3BM and Run-3NO can be compared and fit together.

5.2 Preparing the precession frequency histogram

Precession frequency histograms contain clusters, which are generated by the procedure described in section 4.7. Each cluster now has a corrected and calibrated time and energy coordinate, which ideally corresponds to the time and energy of a decay positron. Clusters are further prepared for fitting by:

1. Randomizing the clusters to remove residual effects of the cyclotron motion
2. Applying a pileup correction
3. Applying a hardware and software blinding so that the frequency reported does not reveal physics information
4. Correcting the remaining gain-like slow term due to reconstruction issues
5. Creating a one dimensional histogram as a function of time using two different weighting schemes

The time histogram can then be fit. Each of these topics will be discussed in turn in the following sections. Number 4, the slow term, will be discussed later in this chapter because of the way in which it is discovered and corrected.

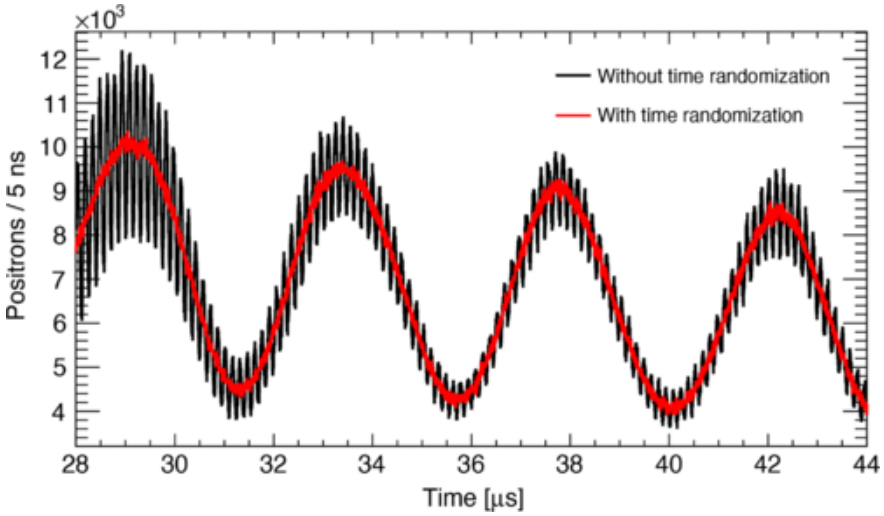


Figure 5.1: Positron intensity in a single calorimeter as a function of time, showing the slower ω_a oscillation with the faster cyclotron oscillation underlying (in black). The randomization procedure described in the text removes the cyclotron oscillation and produces the curve in red, which is ultimately fit to extract ω_a. Reproduced from [3].

5.2.1 Randomization

The muon beam enters the ring with a finite time width of 120 ns and debunches early in the fill as the lower momentum muons at small radii lap the higher momentum muons at high radii [3]. By the fit start time of 30 μs, the beam is almost uniformly distributed in the ring; however, a residual oscillating signal at the cyclotron period is still visible in individual calorimeters. Figure 5.1 shows the measured positron intensity in a single calorimeter. The slower oscillation is the ω_a frequency, and the faster oscillation is the cyclotron frequency. The cyclotron oscillation largely cancels when summing the data from all calorimeters together, since the calorimeters are out of phase with each other, and it is further reduced by binning at the cyclotron period $T_C = 149.2$ for the fit. However, the remaining cyclotron oscillation causes oscillation in the value of R as a function of calorimeter number, as discussed further in section 5.5.

In order to remove this effect, randomization is performed uniformly across the cyclotron period. Every fill, a random number is chosen from the interval $(-T_C/2, T_C/2)$ and added to the cluster times. This washes out the effect of the cyclotron oscillation, as seen in Figure 5.1 in red. To avoid bias from the randomization procedure, fits are performed over an ensemble of 100 random seeds, and the final frequency value is taken from an average of all the random seed results. These randomization results are discussed in

section 5.4.

5.2.2 Pileup correction

The concept of pileup was introduced in subsection 4.7.2 in the context of updates to the clustering algorithm. To summarize, pileup occurs when multiple decay positrons are accidentally mis-reconstructed as a single positron. Detectors have an inherent dead time below which two hits are indistinguishable, with a higher dead time resulting in more pileup. The dead time depends on both the hardware characteristics of the detector and the fitting and clustering algorithms used to reconstruct positrons. In general, a low energy positron and a high energy positron measured at the same time in the same calorimeter will originate from muons with a different phase, as represented schematically in Figure 4.11. When multiple low energy positrons are accidentally reconstructed as a single high energy positron, the single positron will enter into the precession frequency histogram with the wrong phase. Because pileup is a rate-dependent effect, meaning that there is more pileup at the beginning of the fill when the muon intensity is higher, it can bias ω_a if left uncorrected.

In order to achieve the pileup uncertainty goal of < 40 ppb for the Muon g-2 experiment, pileup must be subtracted from the histogram before fitting. A variety of pileup subtraction algorithms have been implemented by different analyzers: a "shadow window" method that overlays clusters to create pileup clusters, an "empirical" method that generates pileup traces and then fully reconstructs them using the usual reconstruction procedure [68], and an energy density method used by the UW analysis in Run-1 that generates the pileup spectrum from the 2D time-energy histogram itself [3, 63]. Systematic uncertainties are associated with the size and shape of the pileup subtraction, which will be discussed further in section 6.3. A pileup corrected spectrum is ultimately analyzed, with the corrected spectrum ρ_{corr} defined as

$$\rho_{\text{corr}} = \rho_{\text{uncorr}} - \rho_{\text{pu}}, \tag{5.1}$$

where ρ_{uncorr} is the uncorrected distribution that emerges from the data and ρ_{pu} is the pileup distribution that will be generated using one of the above procedures.

The largest contribution to ρ_{pu} is referred to as double pileup, when two positrons are reconstructed as a single positron. In order to correct the spectrum for double pileup, the two lower energy clusters

should be added back in and the single high energy cluster removed. Triple pileup refers to three positrons being reconstructed as a single positron, and so on to higher pileup orders. Pileup corrections are generally performed for double and triple pileup, with higher orders being irrelevantly small. Equation 5.1 then becomes

$$\rho_{\text{corr}} = \rho_{\text{uncorr}} - \rho_{\text{pu,d}} - \rho_{\text{pu,t}}, \tag{5.2}$$

with $\rho_{\text{pu,d}}$ and $\rho_{\text{pu,t}}$ referring to double and triple pileup, respectively.

Energy density pileup correction

For the University of Washington analysis in Run-1, an "energy density" method was used to generate ρ_{pu} [3, 63]. For this pileup correction method, rather than overlaying single crystal hits or clusters, the entire two dimensional energy-time histogram is taken as an estimate of the pileup corrected spectrum ρ_{corr}. It can then be integrated to construct an estimate of the pileup spectrum. If $\rho(E,t)$ is the ideal positron hit distribution with no detector dead time or pileup, then the first order double pileup perturbation is

$$\rho_{\text{pu,d}}(E,t) = r(t)\Delta t \cdot \Big(\int \big(\rho(E-E_2,t)\cdot\rho(E_2,t)dE_2\big) - 2\rho(E,t)\int \rho(E_2,t)dE_2\Big), \tag{5.3}$$

where Δt is the detector dead time and $r(t)$ is the positron rate. The first term corresponds to the double pileup hits added to the spectrum, and the second term corresponds to the true hits lost from the spectrum. More details of this calculation can be found in A. Fienberg's thesis [63].

This method of determining the pileup spectrum ignores reconstruction non-linearities. If two positrons would be reconstructed with energies E_1 and E_2 if they were single hits but with combined energy $E \neq E_1 + E_2$ when they pile up, the pileup spectrum reconstructed using this method would not be correct. Additionally, Equation 5.1 only integrates over the time and energy dimensions; if the pileup is position-dependent, the reconstructed pileup spectrum would also be incorrect.

Such a position dependence can arise if the cluster dead time Δt_c is lower than the fitter dead time Δt_f. If $\Delta t_c < \Delta t_f$, then two positrons arriving at the same calorimeter with a time difference t, $\Delta t_c < t < \Delta t_f$, may become a pileup event if the two positrons are spatially coincident but may be separated if the two

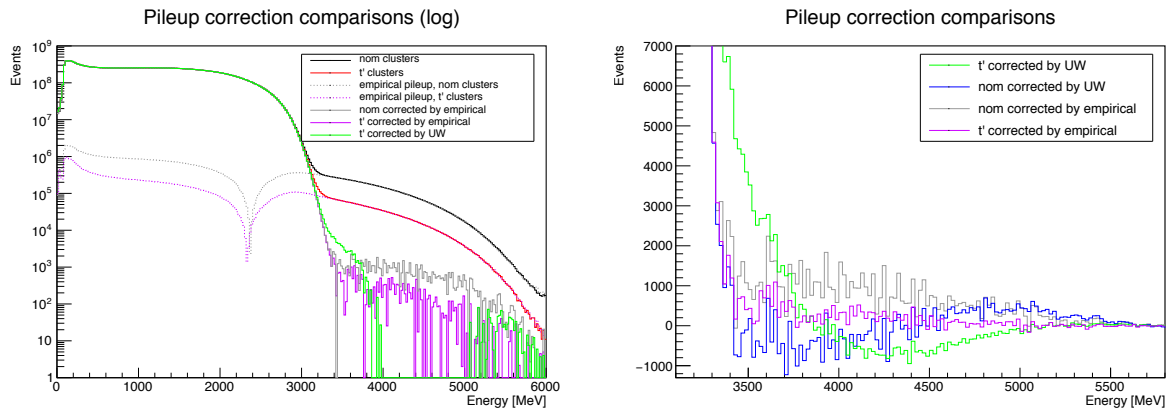


Figure 5.2: Comparison of clustering and pileup subtraction methods, with the full energy spectrum on a log scale shown on the left and the pileup region on a linear scale shown on the right. Combinations of clustering (nominal and $\Delta t'$) and pileup corrections (nominal and empirical) are shown. On the left plot, the black and red lines are the original energy spectrum before pileup subtraction, showing that pileup is reduced for $\Delta t'$ clusters. When the energy density method is applied to the $\Delta t'$ clusters, the constructed pileup spectrum has the wrong shape (in green). The purple line shows the pileup subtracted spectrum used for this analysis, with $\Delta t'$ clusters and the empirical pileup correction. It is an improvement over the combination which would have been used for the Run-1 analysis, shown in blue.

positrons are spatially separated. In Run-1, $\Delta t_c \approx 2.5$ ns, so the energy density method was sufficient. However, the $\Delta t'$ clustering method as discussed in section 4.7 decreases the effective Δt_c to about 0.75 ns.

After generating the precession frequency histogram with the $\Delta t'$ clustering algorithm, the energy density pileup correction was applied, using the same procedure as was used in Run-1. Figure 5.2 shows the Δt and $\Delta t'$ clustering energy spectra corrected by different pileup subtraction algorithms. The left plot shows the full energy range on a log scale, and the right plot shows the pileup region only on a linear scale. Applying the energy density pileup method to the new $\Delta t'$ clusters gives the green curve. In the pileup region, the corrected spectrum has the wrong shape, with the corrected spectrum overshooting and then undershooting the zero line indicating a successful pileup subtraction. This shape mismatch indicates that the energy density method is not adequately capturing the true pileup behavior in the data, likely due to the spatial dependencies introduced by decreasing the cluster dead time in the $\Delta t'$ method. Therefore, rather than adapt the energy density method to integrate over the spatial dependence, the empirical pileup subtraction method was used.

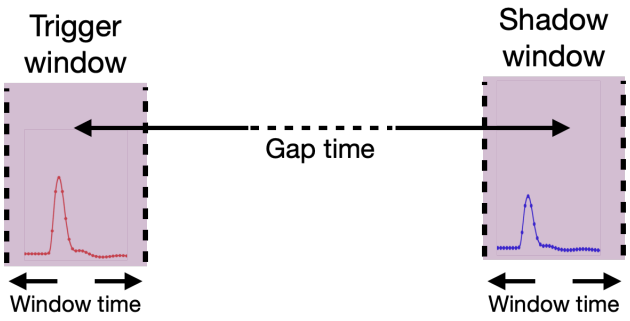


Figure 5.3: Schematic of the empirical pileup method. A data trace in a certain time window triggers a search for another trace a gap time away. If another trace is found, they are overlaid and a pileup trace is generated.

Empirical pileup correction

The empirical pileup method, originally developed by D. Sweigart [68] and adapted for the time-based reconstruction algorithm by S. Foster [73], uses traces from data to generate the pileup spectrum. Similarly to the shadow window method, which was commonly used for the Brookhaven analysis, a cluster in a trigger window prompts a search for a second cluster a gap time Δt_g away in a shadow window (Figure 5.3). Unlike in the shadow method, the traces themselves are overlaid to generate a pileup trace. This trace is then refit and reclustered using the same algorithms as the regular data to produce a pileup cluster, as seen in Figure 5.4, which shows two traces being overlaid and refit to create a pileup island. Therefore this method automatically incorporates the details of the fitting and clustering, including the fitting and cluster dead times δt_f and δt_c. If two positron hits would normally be reconstructed as two clusters, the same will occur for the two artificial pileup positron hits. This allows for accurate construction of a pileup spectrum with no scaling factor or assumptions about the reconstruction required.

The shadow window is taken to be ± 5 ns around the gap time Δt_g, well above the fitter dead time, so that all pileup candidates (traces that might possibly become pileup) are included [68]. The gap time is taken to be the cyclotron period, so that the rate of positrons between the trigger and shadow times is as close as possible. A longer window was chosen over a shorter window so that the trigger trace would not affect the trace in the shadow window. Additional details about how exactly the traces are generated and overlaid are given in D. Sweigart's thesis [68].

The empirical pileup method begins by creating pileup candidate islands. For a double pileup event, the

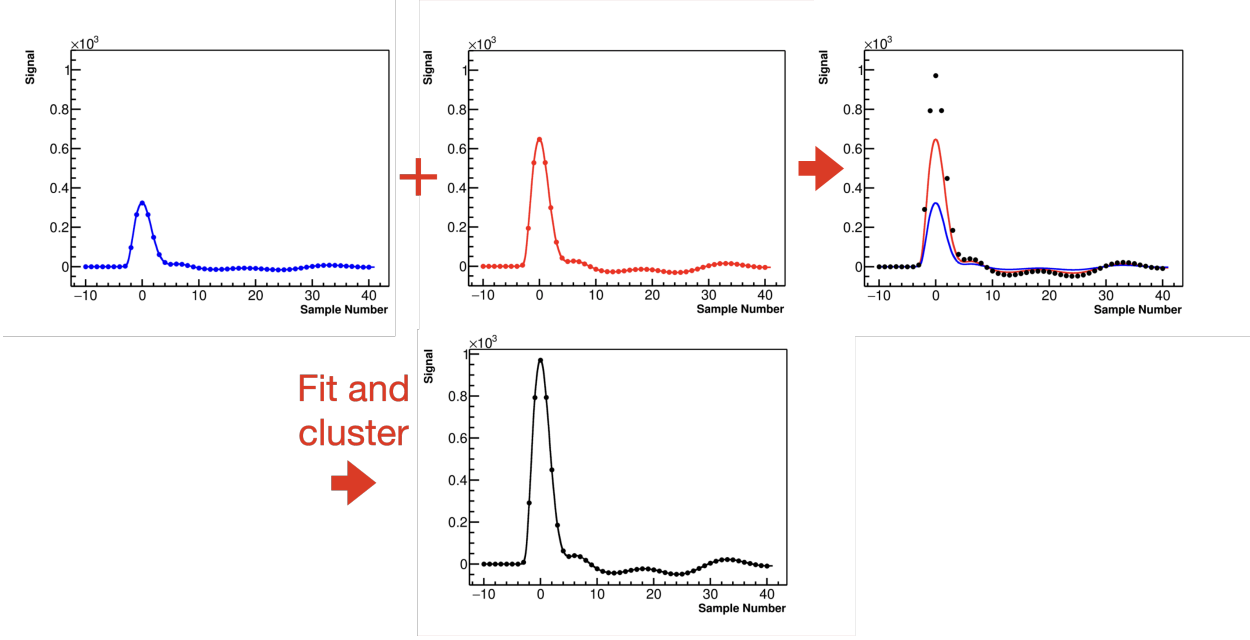


Figure 5.4: Schematic of pileup spectrum generation using the empirical pileup method. Two traces from the data (in blue and red) are overlaid. The traces are added together to create a pileup trace. Then the pileup trace is clustered and fit using the standard algorithms used for the data. The (time, energy) coordinate of this fit will enter the pileup histogram.

pileup information is comprised of three islands: the trigger island, the shadow island a gap time later, and the combined island with the trigger and shadow island overlaid. All three of these islands are then passed through the full reconstruction chain to produce clusters, so that the behavior of the fitting and clustering algorithms in the regular data is replicated.

The combinatorics required to combine the trigger, shadow, and combined islands must be calculated in order to generate the correct pileup spectrum. This calculation follows S. Foster's procedure and notation [73]. The first step in the calculation is to determine the probability of single, double, and triple events occurring in the real data. The goal of the pileup construction is to generate equivalent events by overlaying constituent islands and then calculate the combinatorics needed to replicate the true probability of each type of event. The constructed pileup spectrum can then be subtracted from the full spectrum, leaving only singles. This calculation relies on the assumption that positron hits are independent of each other and therefore obey Poisson statistics. The average number of hits in a time island $\mu(t)$ is given by $\mu(t) = r(t)\Delta t_{\text{island}}$, where $r(t)$ is the time dependent rate of positron hits and Δt_{island} is the width of the island. The probability of measuring k events given the rate $\mu(t)$ is then

$$f(\mu, k) = \frac{e^{-\mu}\mu^k}{k!}. \tag{5.4}$$

The rate $r(t)$ is assumed to be the same for all constituent islands for this calculation (the trigger, shadow, and combined island), which is approximately correct because of the choice of the cyclotron period as the gap time. The error associated with this rate assumption will be discussed in section 6.3.

For islands containing one, two, or three positron hits, $k = 1, 2, 3$. Note that islands with more than one hit are only pileup *candidates*, and that the reconstruction algorithm may be able to distinguish the hits at a later stage of the reconstruction. For each value of k the equation becomes:

$$f(\mu, 1) = \mu e^{-\mu} \tag{5.5}$$

$$f(\mu, 2) = \frac{1}{2}\mu^2 e^{-\mu} \tag{5.6}$$

$$f(\mu, 3) = \frac{1}{6}\mu^3 e^{-\mu}. \tag{5.7}$$

These three equations are expanded using the power series definition of the exponential. Because μ is small, terms above $\mathcal{O}(\mu^4)$ are ignored. The equations become:

$$f(\mu, 1) = \mu - \mu^2 + \frac{1}{2}\mu^3 + \mathcal{O}(\mu^4) \tag{5.8}$$

$$f(\mu, 2) = \frac{1}{2}\mu^2 - \frac{1}{2}\mu^3 + \mathcal{O}(\mu^4) \tag{5.9}$$

$$f(\mu, 3) = \frac{1}{6}\mu^3 + \mathcal{O}(\mu^4). \tag{5.10}$$

The calculation requires accounting for single, double, and triple islands separately. Therefore the equations are rewritten in a basis where where $\hat{x}_k$ indicates an island with k true hits in it, so that for example $\hat{x}_2$ would indicate a double island. The previous equations can then be rewritten as a single equation, which gives the entire uncorrected spectrum:

$$\rho_{\text{uncorr}} = \left(\mu - \mu^2 + \frac{1}{2}\mu^3\right)\hat{x}_1 + \left(\frac{1}{2}\mu^2 - \frac{1}{2}\mu^3\right)\hat{x}_2 + \frac{1}{6}\mu^3\hat{x}_3. \tag{5.11}$$

The ideal corrected spectrum contains only single islands, so the rate of positrons, μ, multiplies the single island vector:

$$\rho_{\text{corr}} = \mu\ \hat{x}_1. \tag{5.12}$$

This is subtracted from Equation 5.11 to find the pileup spectrum in the data:

$$\rho_{\text{pu}} = \rho_{\text{uncorr}} - \rho_{\text{corr}} = \left(-\mu^2 + \frac{1}{2}\mu^3\right)\hat{x}_1 + \left(\frac{1}{2}\mu^2 - \frac{1}{2}\mu^3\right)\hat{x}_2 + \frac{1}{6}\mu^3\hat{x}_3. \tag{5.13}$$

Note that the pileup spectrum contains both double and triple islands, as expected, and single islands which must be added back into the data to account for the singles that were mis-reconstructed as doubles.

In order to subtract off the pileup events from the full uncorrected spectrum, Equation 5.13 must now be created from combinations of constructed pileup events. To construct the pileup spectrum from these islands, the Poisson probabilities are calculated for the trigger and shadow windows individually, and then for the two islands combined. This calculation will allow matching of coefficients between the constructed trigger and shadow windows and the desired pileup spectrum in the data.

The first order pileup candidates arise from overlaying two candidate islands. For these combined islands, the probability of measuring a combined window with k_1 and k_2 events (for the trigger and shadow windows) is the probability of measuring an island with k_1 events times the probability of measuring an island with k_2 events, since the probabilities of measuring an event in both windows are independent:

$$f(\mu, k_1, k_2) = f(\mu, k_1)f(\mu, k_2). \tag{5.14}$$

This combined island is considered a double island because the two islands have been overlaid. The trigger and shadow islands are required by the algorithm to have at least one cluster, since otherwise they would not combine to create a pileup candidate. Therefore k_1 and k_2 cannot be 0. So $k_1 = 1, 2$ and $k_2 = 1, 2$. The probabilities up to $\mathcal{O}(\mu^4)$ are then

$$f(\mu, 1, 1) = \mu^2 - 2\mu^3 \tag{5.15}$$

$$f(\mu, 1, 2) = \frac{1}{2}\mu^3 \tag{5.16}$$

$$f(\mu, 2, 1) = \frac{1}{2}\mu^3, \tag{5.17}$$

A double, triple, and triple island respectively; triple islands can be created by overlaying two constituent islands if one of the constituent islands is a double island. Using the $\hat{x}_k$ notation, this can be rewritten as

$$\rho^{12}_{\text{first}} = \sum_{k_1=1,2;k_2=1,2} f(\mu, k_1, k_2)\hat{x}_{k_1+k_2} = (\mu^2 - 2\mu^3)\hat{x}_2 + \mu^3\hat{x}_3, \tag{5.18}$$

the subscript "first" indicates the first order pileup, and the superscript "12" indicates the combined 1+2 (shadow + trigger) island.

The distributions of the trigger and shadow window events alone must also be calculated. Again, the probability of measuring a trigger island with k_1 is conditional on there being at least one event in the shadow island. This means that the probability of the trigger island having one hit is not the same as in Equation 5.5. Instead, it is given as follows:

$$\rho^{1}_{\text{first}} = \sum_{k_1=1,2;k_2=1,2} f(\mu, k_1, k_2)\hat{x}_{k_1} \tag{5.19}$$

Note that the requirement of a shadow island with $k_2 > 0$ affects the probability of measuring the trigger island, but the shadow island still counts as a single or double island if it has $k_1 = 1, 2$ (respectively) if it is not combined with the shadow island. Similarly, the probability of measuring single and double islands in the shadow island is given as

$$\rho^{2}_{\text{first}} = \sum_{k_1=1,2;k_2=1,2} f(\mu, k_1, k_2)\hat{x}_{k_2}. \tag{5.20}$$

Calculating these sums gives

$$\rho^{1}_{\text{first}} = \left(\mu^2 - \frac{3}{2}\mu^3\right)\hat{x}_1 + \frac{1}{2}\mu^3\hat{x}_2, \tag{5.21}$$

and the same for ρ^{2}_{first}. The pileup spectrum is the difference between Equation 5.18, which corresponds to the combined pileup event that was added to the distribution, and Equation 5.19 and Equation 5.20, which correspond to the two single events that were lost. So the first order pileup spectrum $\rho_{\text{pu,d}}$ is proportional to their difference:

$$\begin{aligned} \rho_{\text{pu, first}} &\propto \rho^{12}_{\text{first}} - \rho^{1}_{\text{first}} - \rho^{2}_{\text{first}} \\ &= (-2\mu^2 + 3\mu^3)\hat{x}_1 + (\mu^2 - 3\mu^3)\hat{x}_2 + \mu^3\hat{x}_3. \end{aligned} \tag{5.22}$$

Technically, the first order pileup correction contains both double and triple pileup candidates (islands which have 2 and 3 events). This equation can be compared to Equation 5.13, the pileup spectrum in the data. To order μ^2,

$$\rho_{\text{pu}} \approx (-\mu^2)\hat{x}_1 + \frac{1}{2}\hat{x}_2 \tag{5.23}$$

$$\rho_{\text{pu, first}} \propto (-2\mu^2)\hat{x}_1 + \mu^2\hat{x}_2. \tag{5.24}$$

So by inspection, a factor of 1/2 is required so that the constructed spectrum matches the expected spectrum in data. The first order correction is then given by

$$\begin{aligned}\rho_{\text{pu, first}} &= \frac{1}{2}(\rho^{12}_{\text{first}} - \rho^{1}_{\text{first}} - \rho^{2}_{\text{first}}) \\ &= \left(-\mu^2 + \frac{3}{2}\mu^3\right)\hat{x}_1 + \left(\frac{1}{2}\mu^2 - \frac{3}{2}\mu^3\right)\hat{x}_2 + \frac{1}{2}\mu^3\hat{x}_3,\end{aligned} \tag{5.25}$$

which is correct up to $\mathcal{O}(\mu^2)$. The first order correction leaves some terms in the pileup spectrum unaccounted for:

$$\rho_{\text{pu}} - \rho_{\text{pu, first}} = -\mu^3\hat{x}_1 + \mu^3\hat{x}_2 - \frac{1}{3}\mu^3\hat{x}_3., \tag{5.26}$$

which will be accounted for by the second order pileup correction, which overlays a trigger island with two shadow islands.

The second order pileup correction can then be calculated, overlaying islands 1, 2, and 3 (trigger window, first shadow window, and second shadow window) to create candidate triple pileup events. Following the same combinatorics procedure for the second order correction up to $\mathcal{O}(\mu^4)$ gives

$$\rho^{123}_{\text{second}} - \rho^{1}_{\text{second}} - \rho^{2}_{\text{second}} - \rho^{3}_{\text{second}} = \mu^3\hat{x}_3 - 3\mu^3\hat{x}_1 \tag{5.27}$$

$$\rho^{12}_{\text{second}} - \rho^{1}_{\text{second}} - \rho^{2}_{\text{second}} = \mu^3\hat{x}_2 - 2\mu^3\hat{x}_1 \tag{5.28}$$

$$\rho^{13}_{\text{second}} - \rho^{1}_{\text{second}} - \rho^{3}_{\text{second}} = \mu^3\hat{x}_2 - 2\mu^3\hat{x}_1 \tag{5.29}$$

$$\rho^{23}_{\text{second}} - \rho^{2}_{\text{second}} - \rho^{3}_{\text{second}} = \mu^3\hat{x}_2 - 2\mu^3\hat{x}_1, \tag{5.30}$$

Where the first equation is the combined 1+2+3 island and the next three the double islands created from all combinations of the three islands.

In order to match the remaining terms in the pileup correction as given in Equation 5.26, the last three equations are subtracted from the first. This gives

$$\begin{aligned}\rho_{\text{pu, second}} \propto &(\rho^{123}_{\text{second}} - \rho^{1}_{\text{second}} - \rho^{2}_{\text{second}} - \rho^{3}_{\text{second}}) - (\rho^{12}_{\text{second}} - \rho^{1}_{\text{second}} - \rho^{2}_{\text{second}}) \\ &- (\rho^{13}_{\text{second}} - \rho^{1}_{\text{second}} - \rho^{3}_{\text{second}}) - (\rho^{23}_{\text{second}} - \rho^{2}_{\text{second}} - \rho^{3}_{\text{second}}) \\ &= 3\mu^3\hat{x}_1 - 3\mu^3\hat{x}_2 + \mu^3\hat{x}_3.\end{aligned} \tag{5.31}$$

By multiplying this equation by -1/3, the second order pileup correction matches the remaining terms in Equation 5.26, meaning that the pileup spectrum up to $\mathcal{O}(\mu^3)$ has been successfully generated from candidate islands. In summary, the full empirical pileup correction up to triple pileup is given by

$$\begin{aligned}\rho_{\text{pu}} &= \rho_{\text{pu, first}} + \rho_{\text{pu, second}} \\ &= \frac{1}{2}(\rho^{12}_{\text{first}} - \rho^{1}_{\text{first}} - \rho^{2}_{\text{first}}) - \frac{1}{3}(\rho^{123}_{\text{second}} - \rho^{1}_{\text{second}} - \rho^{2}_{\text{second}} - \rho^{3}_{\text{second}}) \\ &+ \frac{1}{3}(\rho^{12}_{\text{second}} - \rho^{1}_{\text{second}} - \rho^{2}_{\text{second}}) + \frac{1}{3}(\rho^{13}_{\text{second}} - \rho^{1}_{\text{second}} - \rho^{3}_{\text{second}}) + \frac{1}{3}(\rho^{23}_{\text{second}} - \rho^{2}_{\text{second}} - \rho^{3}_{\text{second}}).\end{aligned} \tag{5.32}$$

The result of applying the first and second order pileup subtraction to the $\Delta t'$ clusters can be seen in Figure 5.2 in purple. Because the $\Delta t'$ clustering algorithm has less pileup to begin with, and because the empirical pileup subtraction algorithm incorporates the spatially dependent behavior of the clustering algorithm, the subtracted spectrum is very flat in the pileup region. Overall, the combination of the new $\Delta t'$ clustering algorithm and the empirical pileup correction results in a very successful pileup correction, with fewer pileup events than any other combination of clustering and pileup subtraction algorithm. An estimation of the systematic uncertainty due to the pileup correction will be discussed in section 6.3.

5.2.3 Blinding

Multiple levels of blinding are implemented in order to avoid biasing the final result. The hardware blinding, which changes the clock frequency by an unknown amount, was discussed in Chapter 3. In addition, a software blinding is applied so that ω_a analyzers are blinded with respect to each other. The software

blinding is implemented by introducing the blinded parameter R, where R is related to ω_a by

$$\omega_a(R) = 2\pi\ 0.2291\ \text{MHz}\ \left[1 + (R - \Delta R) \times 10^6\right], \tag{5.33}$$

ΔR is an analyzer-specific unknown software offset offset. Therefore, a shift of 1 in R corresponds to a 1 ppm shift in ω_a relative to 0.2291 MHz. Removal of the software blinding allows for comparison between analyzers. This blinding has not yet been removed at the time of this thesis.

Future precession frequency results will be presented in terms of R, as given above, rather than in terms of ω_a, and will generally be in ppm.

5.2.4 Weighting

There are several ways to weight the decay positrons in order to create the histogram for fitting. Per Equation 2.11, the asymmetry A is dependent on the energy of the positron E; specifically, higher energy positrons have a higher asymmetry and carry more information about the anomalous precession frequency ω_a. Figure 5.5 (left) shows precession frequency histograms binned by energy. Generally, the asymmetry increases with energy, with higher energy positrons containing more information about ω_a. Vertical lines indicate a single ω_a period, showing that the phase inverts in lower energy bins. Figure 5.5 (right) shows the asymmetry as a function of energy bin, as extracted from fits to the data on the left. The negative asymmetry at energies below 1000 MeV are evident; the asymmetry approaches 1 as the energy approaches E_{max}.

Defining $y \equiv E/E_{\text{max}}$, where $E_{\text{max}} \approx 3.1$ GeV, the simplest weighting method is to assign each positron above a certain energy threshold a weight of one, or $w(y) = 1$. This is referred to as the threshold or "T-method." Because it is the simplest, it is sometimes used to diagnose fitting issues. However, the statistical power can be maximized by weighting the positrons by their asymmetry, or $w(y) = A(y)$, which is called the "A-method" [3, 74]. The statistical figure of merit (FOM), with which the ω_a variance inversely scales, is $N\bar{A}^2$ for the T-method, where N is the number of positrons included in the histogram and $\bar{A}$ is the average asymmetry. For the A-method, it is NA_{rms}^2, where A_{rms}^2 is the root mean square asymmetry above threshold [3]. These FOMs as a function of y were computed in a Monte Carlo simulation and are given in Figure 5.6. The simulation included basic detector acceptance, or the probability of a muon with certain characteristics being measured in the detectors. In the T-method, the numerical advantage of including low

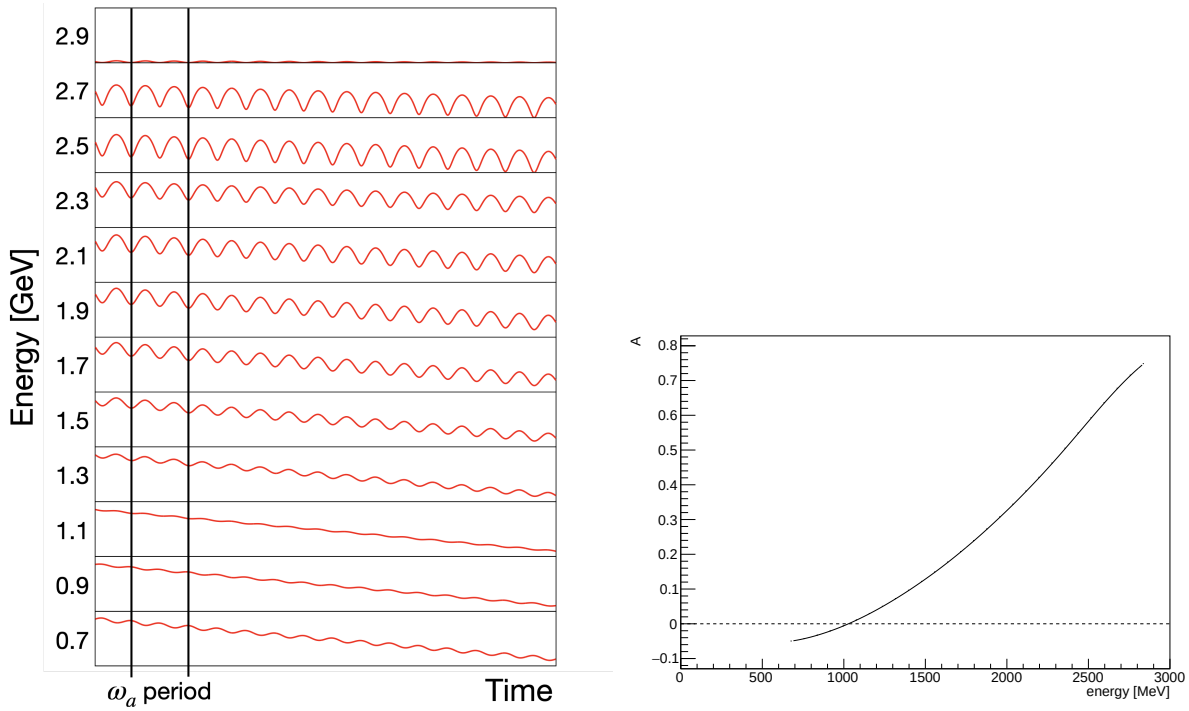


Figure 5.5: Left: histograms from 200 MeV energy bins showing increased asymmetry in higher energy bins. The vertical lines indicate a single ω_a period; the phase changes sign at around 1 GeV. Right: Sample asymmetry vs. energy curve used to weight positrons by their asymmetry for the A-weighted method.

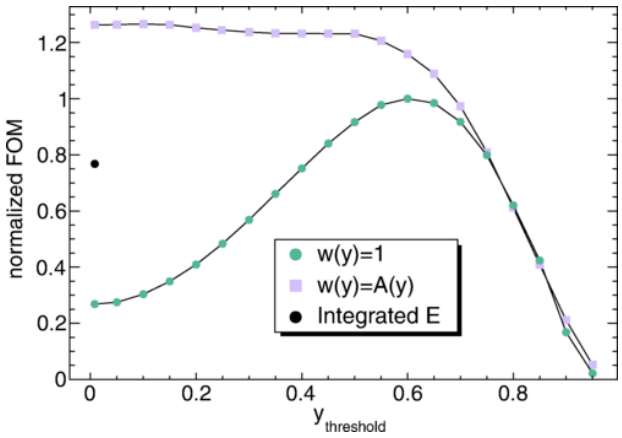


Figure 5.6: Normalized statistical figures of merit produced by Monte Carlo simulation for the threshold (in green) and asymmetry (in purple) methods described in the text as a function of $y_{\text{threshold}} \equiv E/E_{max}$, where $E_{max} \approx 3.1$ GeV, the maximum positron energy. The asymmetry weighted method gives superior statistical precision to the threshold method. In the threshold method, statistical precision is optimized in the threshold method at about 1.7 GeV. The isolated black point indicates the figure of merit for the integrated energy method (or Q-method), which is not discussed in this thesis. Reproduced from [3].

energy positrons is diluted by the low amount of information these positrons carry. Overall, weighting by asymmetry gives superior statistical precision, and therefore the asymmetry method is reported as the final result.

The data are now in the form of a one dimension histogram of decay positrons as a function of time. The oscillations of this histogram ultimately give the precession frequency.

5.3 Fitting procedure

Once the clusters have been corrected for gain and pileup effects and appropriately weighted, they are put into a time-binned histogram to be fit to extract ω_a. The fit start time is standardized among analyzers and is set to $\approx 30.2\ \mu s$, which maximizes the statistical precision while avoiding most of the time-dependent effects, such as beam motion due to scraping and gain changes due to the beam splash, at the beginning of the fill. For the T-method result, the energy threshold is scanned to maximize statistical precision of ω_a. A sample scan is shown in Figure 5.7, showing the uncertainty on ω_a as a function of energy, as extracted from a fit. An optimal threshold of 1690 MeV is extracted. For the Run-2 and Run-3 analysis, the threshold is equalized between analyzers to 1700 MeV so that the results can be more easily compared and combined.

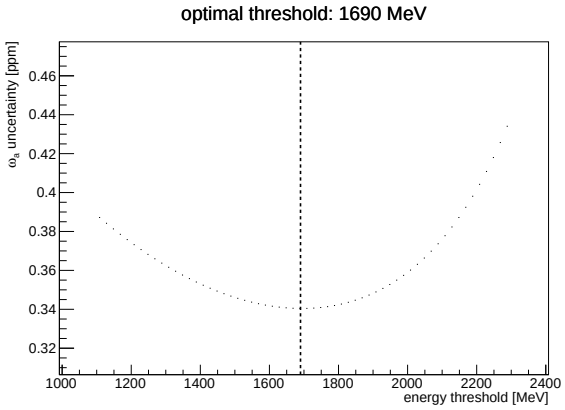


Figure 5.7: Sample scan of ω_a uncertainty as a function of energy threshold for the T-method analysis. The optimal threshold is determined to be 1690 MeV.

For the A-method result, all analyzers use clusters between 1000 and 3000 MeV. Although in principle low energy clusters would result in a slight statistical improvement for the A-method, as seen in Figure 5.6, very low energy clusters are affected by noise and other detector concerns and therefore are excluded.

Based on the calculation described in Chapter 2, the basic five parameter fit function is

$$N(t) = N_0 e^{-t/\tau} \left[1 + A\cos(\omega_a(R)t - \phi)\right], \tag{5.34}$$

where N_0 is the number of positrons at $t = 0$, τ is the time dilated muon lifetime, A is the asymmetry, and ϕ is the spin phase at injection determined by the fit. This fit function is applied to the one dimensional histogram, as seen in Figure 5.8 (left). The muon lifetime $\tau = 64.42\ \mu s$ is visible. At this early stage, the five parameter fit is instead parameterized with a sine and cosine term:

$$N(t) = N_0 e^{-t/\tau} \left[1 + A_C\cos(\omega_a(R)) + A_S\sin(\omega_a(R)\right], \tag{5.35}$$

Rather than with A and ϕ fit parameters.

After applying the 5 parameter fit function, a fast Fourier transform (FFT) is taken of the residuals, as shown in Figure 5.8 (right). Several prominent peaks remain in the FFT, indicating underlying oscillations in the positron number that have not yet been included in the fit function. The vertical dashed lines indicate expected locations for these frequencies, based on estimates coming from beam dynamics calculations. The

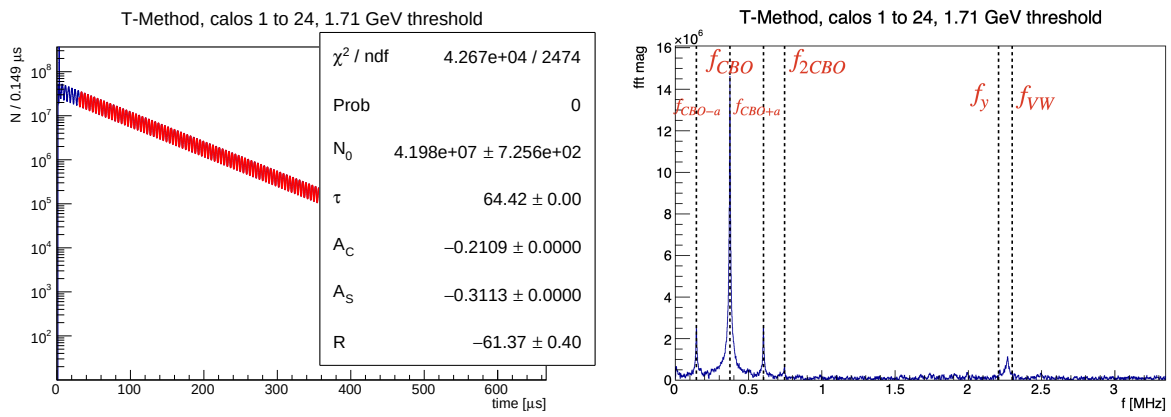


Figure 5.8: Left: five parameter T-method fit to Run-2 data, where R is the blinded precession frequency value and A_C and A_S are the coefficients of the sine and cosine terms used for initial fitting stages. The value of R is blinded and given in ppm. Right: FFT of residuals to 5 parameter fit showing peaks from beam dynamics effects. The f_{CBO} peak results from coherent betatron oscillations of the beam. The peaks f_y and f_{VW} result from vertical betatron and vertical width oscillations, respectively.

source of these peaks, and the methods for fitting them, will be discussed in the following sections.

5.3.1 Beam dynamics

The residual peaks in Figure 5.8 suggest that effects beyond the basic precession frequency oscillation are causing modulations in the positron rate measured by the calorimeters. These oscillations can be estimated using the approximation that the ESQ field is constant around the ring. The field index n is defined as [4]

$$n = \frac{R_0}{vB_0}\frac{\partial E_y}{\partial y}, \tag{5.36}$$

where R_0 is the central orbit radius, v is the muon velocity, B_0 is the magnetic field strength, and the final term is the gradient in the vertical electric field arising from the ESQ geometry and voltage. In Run-2 and Run-3, $n = 0.108$. Estimates for beam oscillations using the n value and other storage ring parameters are given in Table 5.1. These estimates are useful for identifying the source of oscillations present in the ω_a histogram.

The 24 calorimeter detectors are arranged at discrete locations around the ring. Therefore, the acceptance (probability of a decay positron being measured) is not uniform around the ring. Further, different detectors have different acceptances due to their position behind certain ring systems (e.g. ESQ or kicker plates).

Name	Frequency	Expression	Estimated value [rad/μs]
Cyclotron	ω_c	v/R_0	42.15
Horizontal betatron	ω_x	$\sqrt{1-n}\omega_c$	39.81
Vertical betatron	ω_y	$\sqrt{n}\omega_c$	13.85
Coherent horizontal betatron	ω_{CBO}	$\omega_c - \omega_x$	2.34
Vertical waist	ω_{VW}	$\omega_c - 2\omega_y$	14.45

Table 5.1: Beam oscillation frequencies and their expected values based on the continuous quadrupole approximation, with field index $n = 0.108$. Here, v is the muon velocity and R_0 is the radius of the storage ring. Reproduced from [4].

Therefore, some frequencies will be aliased at the cyclotron frequency because of discrete sampling in each calorimeter or tracking detector. Frequencies greater than the Nyquist frequency (half the cyclotron frequency $f_c/2$), such as the horizontal betatron oscillation ω_x, are aliased. Frequencies less than than the Nyquist frequency, such as the vertical betatron oscillation ω_y, will be measured at their un-aliased value.

The frequencies can also be measured in the tracking detectors allowing further identification of the different oscillations. Figure 5.9 shows the reconstructed radial position (left) and vertical position (right) of the beam as a function of time in fill. The plot on the left shows the aliased horizontal betatron oscillation, and the plot on the right shows the vertical betatron oscillation. Beam oscillations will also generally decohere during the measurement period. Because of the finite momentum spread of the beam, and because the beam oscillation frequencies are momentum dependent, any coherent beam motion will disappear over time [63]. Each of these oscillations will also be assigned a decoherence envelope in the fit function. The decoherence is visible in Figure 5.9; note that the CBO has a significantly longer lifetime than the vertical oscillation.

Coherent betatron oscillations

The largest peak visible in the FFT is at ≈ 0.37 MHz. The beam oscillates horizontally in the storage region at the horizontal betatron oscillation, ω_x, with a magnitude dependent on the strength of the magnetic kick. Up until Run-3NO, the beam was under-kicked, so the oscillation around the ideal central orbit was large. The kicker strength was improved in Run-3NO, so the amplitude was reduced; however, this oscillation is still present in all datasets.

The horizontal betatron frequency f_x is greater than half the cyclotron frequency f_c, and is therefore

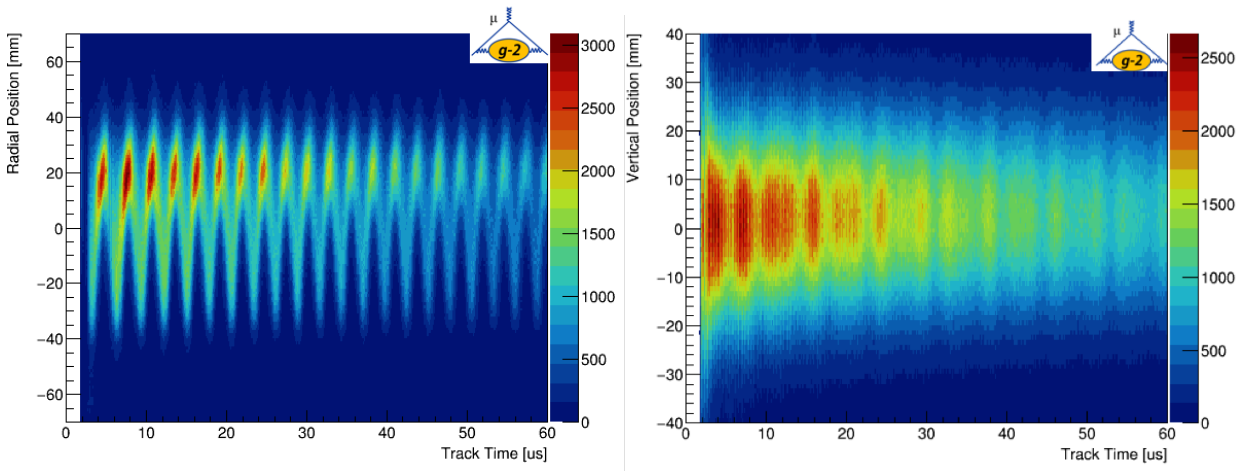


Figure 5.9: Left: Reconstructed radial position of the muon decay vs. time in fill, as measured in the tracking detectors. This is the coherent betatron oscillation (CBO) also measured in the calorimeters. Right: reconstructed vertical position of the muon decay vs. time in fill. The vertical width oscillation is visible.

aliased when measured in tracking or calorimeter detectors at discrete locations around the ring. This oscillation is known as the coherent betatron oscillation or CBO, where $\omega_{CBO} = \omega_c - \omega_x$ [4]. A schematic of this aliasing effect is shown in Figure 5.10. The cyclotron wavelengths are shown as vertical black lines, and a single calorimeter or tracking detector is labeled below. The slower CBO oscillation, shown in red, is the frequency that is actually measured by the calorimeter detectors, and therefore the frequency which must be fit for the ω_a analysis.

The CBO frequency amplitude is quite large in an individual detector. However, it can be largely cancelled by adding data from calorimeters across the ring from each other. If the acceptance of all the calorimeters were exactly equal (if particles with the same characteristics had the same probability of being measured in different calorimeters), the CBO oscillation would cancel due to the phase offset around the ring. Because detector acceptance is nonuniform, the cancellation is incomplete and a signal remains in the summed data.

The CBO oscillation is given a decoherence "envelope" E_{CBO} so that it will decrease as a function of the time in fill, as visible in Figure 5.9. The most general parameterization of the CBO oscillation is then

$$N_{CBO}(t) = 1 + A_{CBO}E_{CBO}(t, ...)\cos(\omega_{CBO}t - \phi_{CBO}), \tag{5.37}$$

where A_{CBO} is the CBO amplitude, ω_{CBO} is the CBO frequency, and ϕ_{CBO} is the CBO phase. The CBO

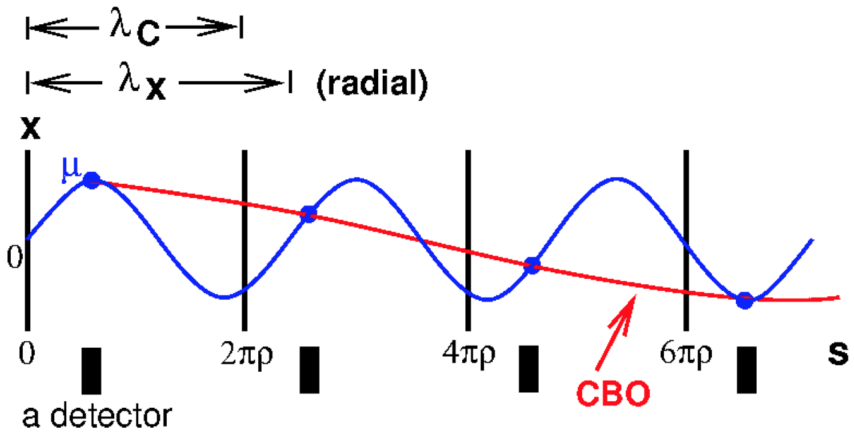


Figure 5.10: Schematic demonstrating how the horizontal betatron frequency (in blue) becomes aliased by the cyclotron frequency (vertical black lines) and is measured as the CBO frequency (in red) in discrete calorimeter detectors. Reproduced from [7].

envelope is given here as $E_{CBO}(t, ...)$ to indicate that the number of additional fit parameters will depend on the choice of envelope.

The simplest form of E_{CBO} is an exponential decay function:

$$E_{CBO}(t, \tau_{CBO}) = e^{-t/\tau_{CBO}}, \tag{5.38}$$

where τ_{CBO} is the CBO lifetime. This is the formulation used for the Run-1 analysis. This envelope is not generally known precisely, so other envelopes may be preferable. Fitting by calorimeter number increases the amplitude of the CBO oscillation, so CBO fitting details are easier to determine when fitting by calorimeter. This will be discussed further in section 5.5.

Equation 5.37 then multiplies the five parameter fit function Equation 5.62 to give:

$$N(t) = N_0 e^{-t/\tau} \left[1 + A\cos(\omega_a(R)t + \phi)\right] \;\; N_{CBO}(t). \tag{5.39}$$

If $N_{CBO}(t)$ is multiplied with the five parameter fit function, beat frequencies at $\omega_{CBO} \pm \omega_a$ will also be present. These frequencies are marked by vertical dashed lines on the FFT. After including the CBO oscillation in the fit function, the CBO peak at 0.37 is reduced. The resulting FFT of the residuals is shown in Figure 5.11. Compared to Figure 5.8, good suppression of the CBO peak at 0.37 MHz and its beat

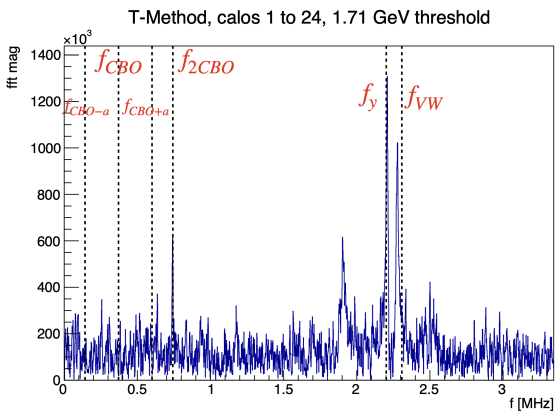


Figure 5.11: FFT of residuals to 9 parameter fit (five parameter fit with CBO oscillation) showing good suppression of the CBO peak at 0.37 MHz. Other peaks due to the vertical waist oscillation, vertial betatron oscillation, and 2*CBO oscillation are now visible.

frequencies is achieved.

In the Run-1 analysis, the CBO frequency changed significantly throughout the fill due to the damaged ESQ resistors. An additional smaller frequency change with a shorter time constant was due to the residual effects of scraping, which is when the voltage is adjusted to scrape the beam against collimators to remove muons likely be lost later. After scraping, the beam relaxes to its nominal frequency and position. In Run-1, these effects were accounted for in the fit by changing ω_{CBO} by

$$\omega_{CBO} \rightarrow \omega_{CBO} + ae^{-t\tau_a}/t + be^{-t/\tau_b}/t, \tag{5.40}$$

where a, τ_a accounted for the shorter lifetime effect due to scraping and b and τ_b accounted for the long lifetime effect caused by the damaged ESQ resistors. The lifetimes and amplitudes were measured by the tracking detectors.

Although the larger effect due to the damaged resistors is no longer relevant for Run-2/3, the smaller effect due to scraping remains. The time changing CBO measured by the trackers in Run-2 is seen in Figure 5.12. This figure shows measurements from both tracker stations: station 12 in front of calorimeter 13 on the left and station 18 in front of calorimeter 19 on the right. Although the effect has largely disappeared by the fit start time of 30 μs, a small effect is incorporated into the fit using the same parameterization as in

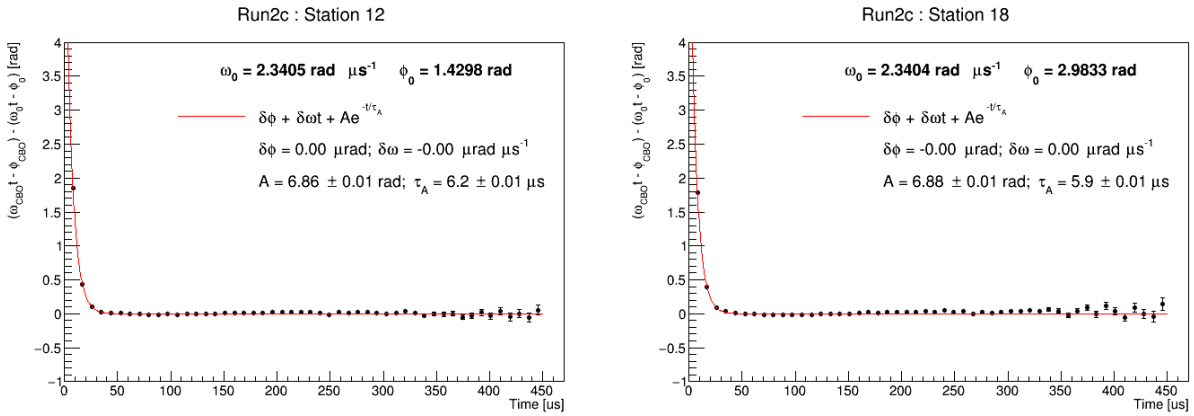


Figure 5.12: Time changing CBO frequency in Run-2 as measured by the tracking detectors. This change is due to the residual effects of scraping at the fit start time. The time constant is $\approx 6\mu s$, resulting in a small effect on the fit. Reproduced from J. Mott.

Run-1:

$$\omega_{CBO} \rightarrow \omega_{CBO} + ae^{-t\tau_a}/t, \tag{5.41}$$

with only one lifetime required. The a and τ_A values were calculated from averages between the two tracker stations and fixed in the fit.

After the main CBO oscillation of N has been incorporated into the fit, other smaller peaks are now visible, including a peak at twice the CBO frequency (0.74 MHz). The respective sizes of the CBO frequency harmonics depends on the detector acceptance; other harmonics are too small to be seen [75]. The $2\omega_{CBO}$ oscillation is incorporated into Equation 5.37 by multiplying the fit equation by

$$N_{2CBO}(t) = 1 + A_{2CBO}E_{CBO}(t, \tau_{CBO}/2)\cos(2 \cdot \omega_{CBO}t - \phi_{2CBO}), \tag{5.42}$$

using half the CBO lifetime and twice the CBO frequency, with the additional parameters A_{2CBO} and ϕ_{2CBO}.

In addition, the CBO also causes modulation of the A, and ϕ parameters in the basic five parameter fit:

$$A(t) = A \cdot A_{CBO,A}E_{CBO}(t)\cos(\omega_{CBO}t - \phi_{CBO,A}) \tag{5.43}$$

$$\phi(t) = \phi + A_{CBO,\phi}E_{CBO}(t)\cos(\omega_{CBO}t - \phi_{CBO,\phi}), \tag{5.44}$$

so that the constant asymmetry and phase are modified by time-dependent CBO oscillations. This results in an additional four CBO-related terms in the fit: $A_{CBO,A}$, $\phi_{CBO,A}$ for the CBO modulation of A and $A_{CBO,\phi}$, $\phi_{CBO,\phi}$ for the CBO modulation of ϕ. These last six smaller terms will be incorporated into the fit function for the final fit result in a later section. Note that E_{CBO} is assumed to have the same form for these oscillations as for the main CBO amplitude oscillation.

Vertical betatron oscillation

Now, the vertical betatron oscillation is visible in Figure 5.11. Because it is less than $\omega_c/2$, it is not aliased by the cyclotron frequency. Similarly to the CBO frequency, additional terms of the form

$$N_y = 1 + A_y e^{-t/\tau_y} \cos(\omega_y t - \phi_y) \tag{5.45}$$

can be included in the fit function with a standard exponential decay envelope, where new parameters A_y, τ_y, ω_y, and ϕ_y are the vertical amplitude, lifetime, frequency, and phase. However, because of the vertical oscillation's smaller amplitude and its proximity to the vertical width oscillation (discussed in the next section), it can be difficult to fit in lower statistics datasets. Specifically, in Run-2 and Run-3NO, the y oscillation lifetime τ_y could not be fit with good precision.

An alternate approach was pursued to identify τ_y. By only fitting horizontal slices of the calorimeter, the vertical oscillation could be greatly magnified, since the cancellation between the oscillation in the top and bottom of the calorimeter would not occur. This magnification of the vertical oscillation allows for τ_y to be extracted. Figure 5.13 shows the result of performing a 9 parameter fit (5 parameter + CBO) to the top two and bottom two rows of the calorimeter separately. The y oscillation peak at 2.2 MHz is so large that other peaks are not really visible. The oscillation was included in the fit using Equation 5.45. The fit results are shown in the middle and right plots in Figure 5.13. Although the χ^2 is not good and the FFT is not flat, the lifetime τ_y can be extracted with high precision. The lifetime τ_y is consistent between the fits to the top and bottom two rows of the calorimeter, with $\tau_{y,\text{top}} = 96.3 \pm 0.9\ \mu s$ and $\tau_{y,\text{ bottom}} = 95.7 \pm 0.8\ \mu s$. The lifetime is therefore fixed to the average of 96 μs in Run-2, Run-3BM, and other sub-datasets where τ_y cannot be extracted.

The larger statistics in Run-3BM allow the lifetime to be determined without fixing it, so it is allowed

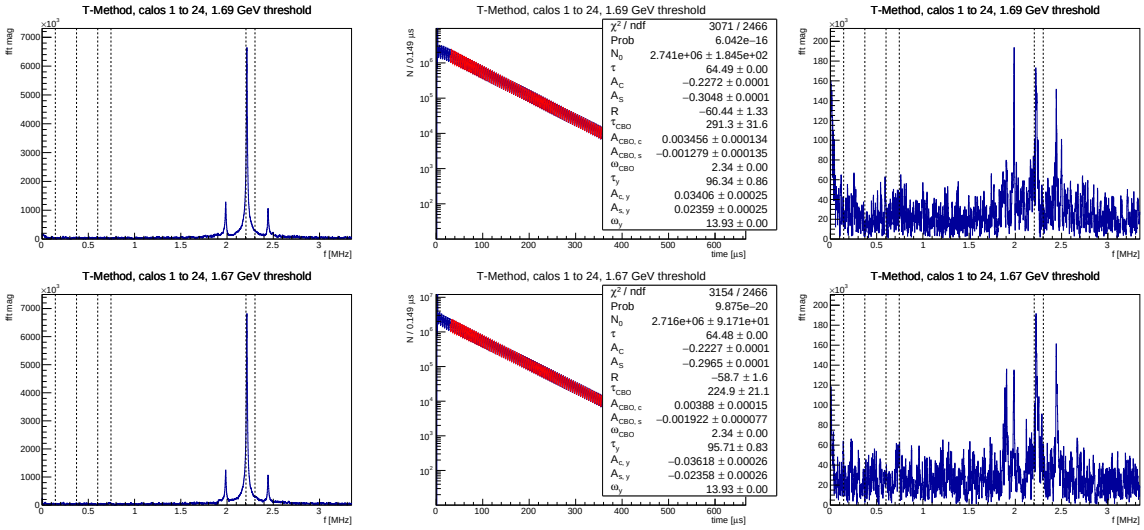


Figure 5.13: Top row: fits to Run-2 data from the top two rows of the calorimeter only. The left plot is the FFT of a 9 parameter fit, with the large vertical betatron oscillation at 2.2 MHz clearly visible. The middle plot includes terms for the vertical oscillation, and the right plot is the FFT of that fit. Bottom row: same as above, but for the two bottom rows of the calorimeter, showing that τ_y is consistent.

to float. The fitted τ_y in Run-3, $102.6 \pm 20.0\ \mu s$ is consistent with $\tau_y = 96\ \mu s$ extracted by the previously described procedure, lending confidence to the choice to fix the lifetime for some datasets.

With the y oscillation included, the fit function becomes

$$N(t) = N_0 e^{-t/\tau}\left[1 + A(t)\cos(\omega_a(R)t + \phi(t))\right]\ N_{CBO}(t)\ N_{2CBO}(t)\ N_y(t). \tag{5.46}$$

Vertical width oscillation

The peak in the FFT at ≈ 2.28 MHz is referred to as the vertical width or vertical waist oscillation. This is the aliased form of twice the vertical betatron oscillation. Because $2\omega_y > \omega_c/2$, ω_{VW} is measured in its aliased form $\omega_{VW} \approx \omega_c - 2\omega_y$. This oscillation is visible in detectors as a vertical width oscillation, as seen in Figure 5.9 (right), which shows the reconstructed vertical position of the muon decay vs. time.

In order to fit this oscillation, a term of the form

$$N_{VW} = 1 + A_{VW} e^{-t/\tau_{VW}} \cos(\omega_{VW} t - \phi_{VW}) \tag{5.47}$$

is included in the fit function. Four parameters (A_{VW}, τ_{VW}, ω_{VW}, and ϕ_{VW}) are needed to account for the vertical waist oscillation. The full fit function is now

$$N(t) = N_0 e^{-t/\tau} \left[1 + A(t)\cos(\omega_a(R)t + \phi(t))\right] \; N_{CBO}(t) \; N_{2CBO}(t) \; N_y(t) \; N_{VW}(t). \tag{5.48}$$

VW-CBO beat frequency

After applying the full fit function with all known frequencies (Equation 5.48), a prominent peak remained at around 1.9 MHz. Figure 5.14 shows the FFT for Run-2 and Run-3BM. The peak is more visible in Run-3BM, although it is present in both datasets. This frequency did not correspond to any known oscillation, nor did it seem specific to any given dataset or time period. The magnitude of the FFT peak did depend somewhat on the random seed, but it was generally present, particularly in the clean and high statistics Run-3 datasets.

Combining known frequencies together suggested that the 1.9 MHz peak could be a beat frequency of the CBO and VW oscillations at $f_{VW} - f_{CBO}$. The VW and CBO frequency terms can be expanded:

$$\begin{aligned} N_{CBO}(t) \; N_{VW}(t) = 1 &+ A_{CBO} e^{-t/\tau_{CBO}} \cos(\omega_{CBO} t - \phi_{CBO}) \\ &+ A_{VW} e^{-t/\tau_{VW}} \cos(\omega_{VW} t - \phi_{VW}) \\ &+ A_{CBO} A_{VW} e^{-t/\tau_{CBO} - t/\tau_{VW}} \cos(\omega_{CBO} - \phi_{CBO}) \cos(\omega_{VW} - \phi_{VW}). \end{aligned} \tag{5.49}$$

Using cosine identities, the final beat term becomes

$$\begin{aligned} e^{-t/\tau_{CBO} - t/\tau_{VW}} \Bigg(&\frac{A_{CBO} A_{VW}}{2} \cos\left((\omega_{CBO} + \omega_{VW})t - (\phi_{CBO} + \phi_{VW})\right) \\ &+ \frac{A_{CBO} A_{VW}}{2} \cos\left((\omega_{CBO} - \omega_{VW})t - (\phi_{CBO} - \phi_{VW})\right) \Bigg), \end{aligned} \tag{5.50}$$

so that the beat frequency amplitude is $\frac{A_{CBO} A_{VW}}{2}$. The first question is why oscillation at VW-CBO would be visible but not the oscillation at VW+CBO. This question can be answered by performing a sum across the results from different calorimeters. If each calorimeter is fit separately, then the CBO and VW phases go through approximately 2π around the ring, as seen in Figure 5.15, which shows the CBO phase ϕ_{CBO}

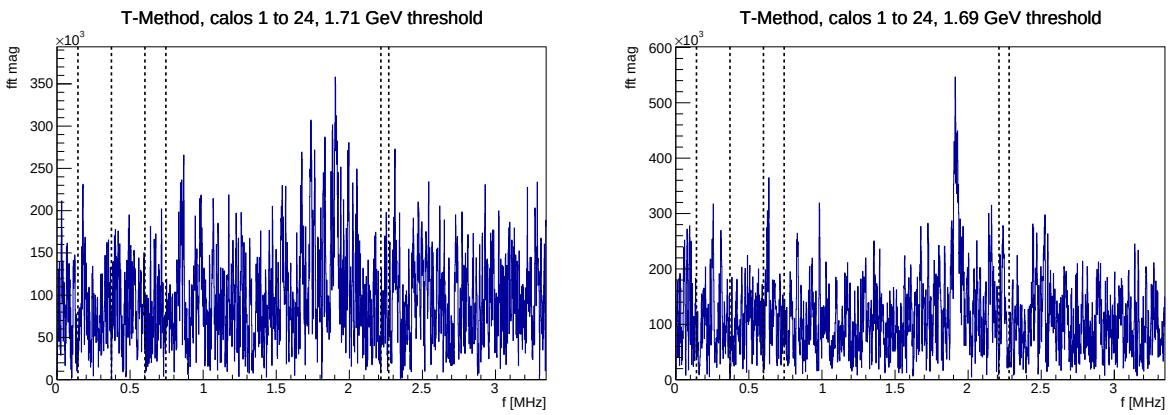


Figure 5.14: FFT of fit to Run-2 (left) and Run-3BM (right) with all standard parameters for Run-1 (CBO, VW, y, etc.) showing a peak at 1.9 MHz. Although the peak is more prominent in Run-3, it is present in both FFTs.

on the left and the VW phase ϕ_{VW} on the right extracted from the fit. Following J. Mott's reasoning [76], an approximation is made that the phase advances by $2\pi/24$ for each calorimeter numbered N_C = 0, ... 23, then the CBO and VW terms become

$$A_{CBO}e^{-t/\tau_{CBO}}\cos\big(\omega_{CBO}t-(\phi_{CBO}+2\pi N_C/24)\big), \tag{5.51}$$

$$A_{VW}e^{-t/\tau_{VW}}\cos\big(\omega_{VW}t-(\phi_{VW}+2\pi N_C/24)\big) \tag{5.52}$$

The two beat terms then become

$$\begin{aligned}&\frac{A_{CBO}A_{VW}}{2}\cos\big((\omega_{CBO}+\omega_{VW})t-(\phi_{CBO}+\phi_{VW}+4\pi N_C/24)\big)\\&+\frac{A_{CBO}A_{VW}}{2}\cos\big((\omega_{CBO}-\omega_{VW})t-(\phi_{CBO}-\phi_{VW})\big).\end{aligned} \tag{5.53}$$

When summing over N_C, and assuming perfect cancellation (no acceptance effects), the beat frequency at VW+CBO will then cancel, but the beat frequency at VW-CBO will remain. This matches what is seen in the data, where the peak at VW-CBO (1.9 MHz) is visible but the peak at VW+CBO is not.

The second issue is that the peak at 1.9 MHz is much larger than the expected beat frequency amplitude of $A_{CBO}A_{VW}/2$. Simulations were performed by K.S. Khaw and Y. Zeng [77] to determine if detector

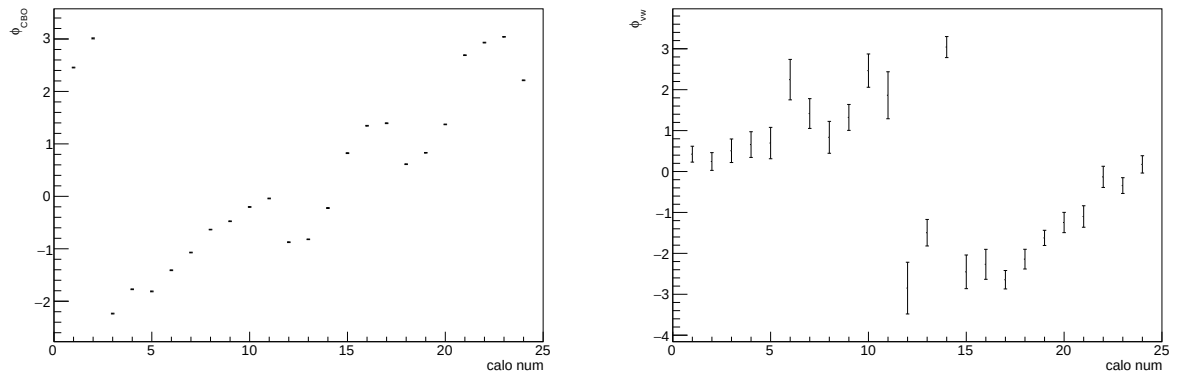


Figure 5.15: Phases of the CBO and VW oscillations as a function of calorimeter number, showing approximate rotation of 2π around the ring.

acceptance effects could exaggerate the size of this cross term. A simple model histogram was generated by

$$N_{gen} = N_0 \epsilon(x)\epsilon(y), \tag{5.54}$$

where $\epsilon(x)$ and $\epsilon(y)$ are the x and y efficiencies of measuring the positron taken from detailed simulations. The time dependence was incorporated into the equation by modeling the x and y betatron oscillations as $x(t) = A_x \cos(\omega_x t - \phi_x)$ and $y(t) = A_y \cos(\omega_y t - \phi_y)$, with standard parameters used for ω_x, ϕ_x, ω_y, and ϕ_y. The previous equation then becomes

$$N_{gen}(t) = N_0 \epsilon(x(t))\epsilon(y(t)). \tag{5.55}$$

The detector acceptance can be decomposed into correlated and uncorrelated components. That is, if the 2D acceptance map is fit to the polynomial

$$\epsilon_{fit}(x, y) = a_0 + a_1 x + a_2 y^2 + a_3 y^4 + a_4 x y^2 + a_5 x y^4, \tag{5.56}$$

then the fit function can be separated into the correlated component $\epsilon_C(x, y)$ and uncorrelated component $\epsilon_U(x, y)$, where

$$\epsilon_U(x, y) = a_0 + a_1 x + a_2 y^2 + a_3 y^4 \tag{5.57}$$

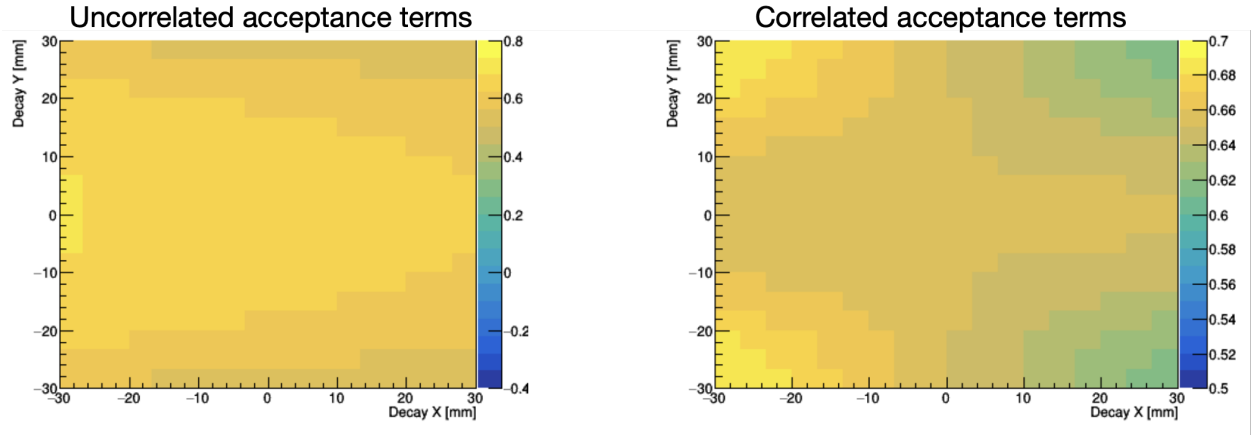


Figure 5.16: Left: x-y uncorrelated components of the simulated detector acceptance ϵ. Right: x-y correlated components. Reproduced from [77].

$$\epsilon_C(x, y) = a_0 + a_3 xy^2 + a_5 xy^4. \tag{5.58}$$

These two components are shown in the two plots in Figure 5.16, with the uncorrelated components on the left and the correlated components on the right. Figure 5.17 shows the difference between generating the histogram with the different components. The red line shows the FFT of the Monte Carlo data, the blue lines the FFT of the fit function, and the green lines the residual FFT after applying the fit function. When only the uncorrelated acceptance is used to generate MC data, the residual peak at 1.9 MHz is small. However, when the correlated acceptance is used to generate data, the 1.9 MHz peak is large. This suggests that the correlated components of the detection efficiency map couple the horizontal CBO oscillation and vertical VW oscillation together and increase the amplitude of their beat frequency beyond what it would normally be without acceptance effects. Further studies with the model demonstrated that simply allowing the beat frequency amplitude to float resulted in a good fit to the Monte Carlo data.

Following this simulation work, in order to account for the 1.9 MHz peak in the real data, the VW-CBO cross term is expanded and the VW-CBO amplitude allowed to float. Although the VW+CBO term may be present because of unequal acceptance between calorimeters, it is too small to be visible in any of the data

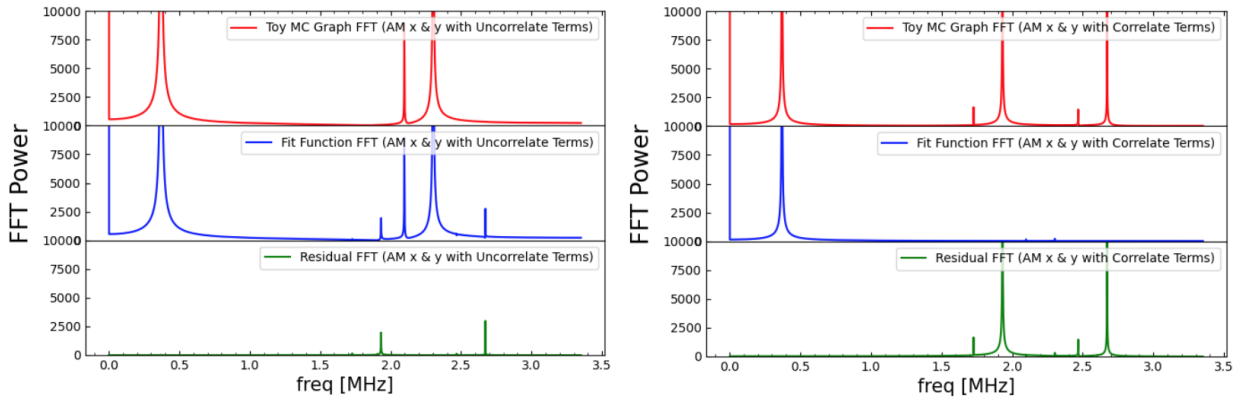


Figure 5.17: Figures showing the generation of the 1.9 MHz peak using a toy Monte Carlo model. The left figure uses only uncorrelated acceptance terms, and the right figure uses only correlated acceptance terms. The red lines are the Monte Carlo FFT, the blue lines the fit function FFT, and the green lines the FFT of the fit residual. In the left plot, the 1.9 MHz peak is small in the residual FFT, whereas in the right plot the 1.9 MHz peak is large. This suggests that correlated x-y acceptance terms are responsible for the 1.9 MHz peak. Reproduced from [77].

fit in this thesis. In the fit function Equation 5.48, $N_{CBO}(t)\ N_{VW}(t)$ is replaced by $N_{CBO,VW}(t)$, where

$$\begin{aligned} N_{CBO,VW}(t) = 1 &+ A_{CBO}e^{-t/\tau_{CBO}}\cos(\omega_{CBO}t - \phi_{CBO}) \\ &+ A_{VW}e^{-t/\tau_{VW}}\cos(\omega_{VW}t - \phi_{VW}) \\ &+ e^{-t/\tau_{CBO}-t/\tau_{VW}}A_{VW-CBO}\ \cos\left((\omega_{VW} - \omega_{CBO})t - \phi_{VW-CBO}\right). \end{aligned} \tag{5.59}$$

The final term is the beat frequency. The lifetime and frequency of this beat term is fixed from the CBO and VW oscillations, but the amplitude and phase are allowed to float. Therefore two additional terms are included in the fit, A_{VW-CBO} and ϕ_{VW-CBO}.

5.3.2 Lost muons

Muons can be lost from the storage ring before decaying into positrons. When muons are lost, the population of stored muons is depleted faster than the muon lifetime, resulting in a slow (low frequency) term in the FFT if left uncorrected, as visible in Figure 5.8 (right). Therefore, the lost muons must be incorporated into the fit function. Lost muons have been discussed in previous sections such as subsection 3.2.5, but to summarize, their time spectrum can be measured by searching for their distinctive signature in the calorimeters. Muons lost from the storage ring can pass through several successive calorimeters, acting as a minimum ionizing

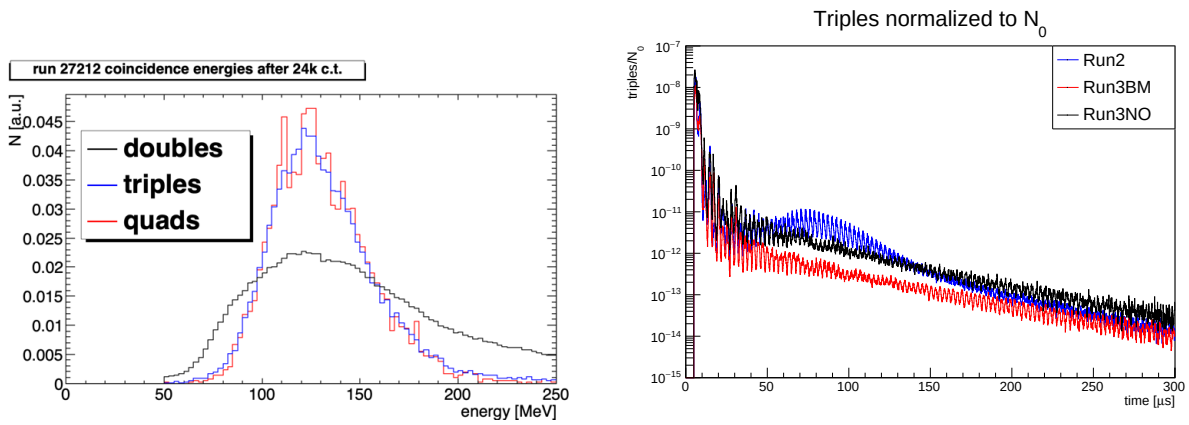


Figure 5.18: Left: Energy distribution of double, triple, and quadruple coincidences in adjacent calorimeters showing the MIP peak between 50 and 250 MeV. The double coincidence spectrum is contaminated by accidental coincidences, whereas the triple and quadruple coincidences are similar in shape. Reproduced from A. Fienberg. Right: Triples spectrum normalized to N_0 for Run-2, Run-3BM, and Run-3NO.

particle (MIP) and depositing a small amount of energy in each detector, as seen in Figure 3.7. Further, adjacent hits will have a distinctive time difference of ≈ 6.4 ns due to the flight time of the muon between calorimeters.

Figure 5.18 shows the energy distribution of double, triple, and quadruple coincidences in adjacent calorimeters. The double coincidence spectrum is contaminated by accidental coincidences, whereas the triple and quadruple coincidence spectra are very similar in shape, indicating that the triples spectrum is a relatively pure sample of true coincidences; therefore, the triples spectrum is generally used. Energies between around 170 MeV are characteristic of the MIP peak, as discussed in section 4.5. A lost muon spectrum is therefore generated using triple coincidences with energy 100 MeV $< E <$ 250 MeV and time difference 5.5 ns $< \Delta t <$ 7 ns.

These triple coincidences are plotted as a function of time in fill for the three datasets in Figure 5.18. They are normalized to the fit parameter N_0 for comparison between the three datasets. The two Run-3 datasets have a flat structure, with the losses occurring at the muon lifetime. This indicates that a constant number of muons are being lost per number of muons stored. In Run-2, however, there is a distinctive "bump" at 50-150 μs where the number of losses increases and then decreases again.

The lost muon bump changed significantly over the course of Run-2, as shown for two sample time periods in Figure 5.19. In the left plot, the triples spectrum is relatively flat and the number of losses low,

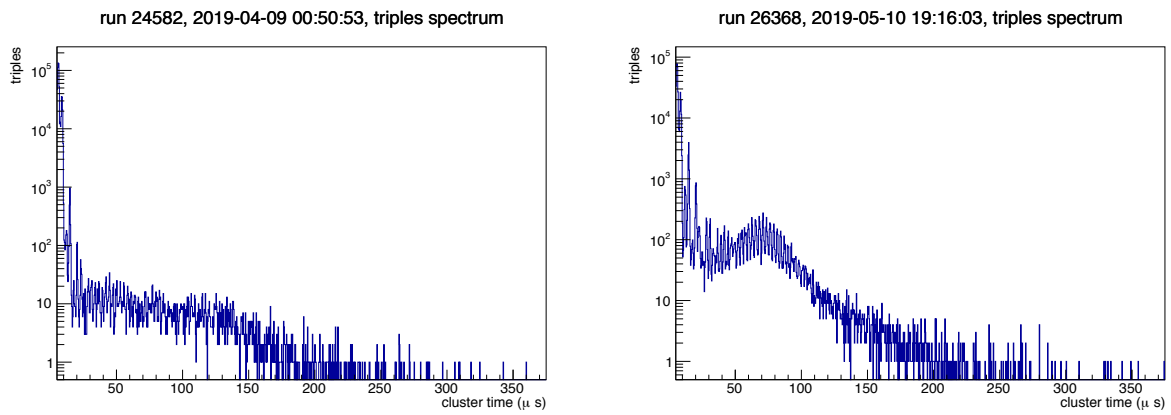


Figure 5.19: Run-2 triples spectra showing the lost muon bump appearing and disappearing at different periods during Run-2.

whereas in the right plot, there is a distinctive bump between 50 and 100 μs. In order to understand the origin of this feature in the muon loss spectrum, the size of the bump was measured by taking the integral of the spectrum from 50 to 150 μs and normalizing it to the integral of the spectrum at early times. The size of the bump was correlated with various magnet parameters, including the magnet yoke temperature and the magnet feedback current, which is used to reduce the drift in the magnetic field. These correlations can be seen in Figure 5.20. The plots are color coded by date ranges associated with different magnet set points, which generally also corresponds to the sub-datasets described earlier. In general, as the temperature increased into spring and summer, the feedback current would be changed until it reached its limit, and then the magnet set point would be changed. The size of the bump seems to be related to both the temperature and the feedback current, suggesting that a temperature-induced magnet change is causing the lost muon bump.

The bottom two plots are the most conclusive. The left plot shows the average hit position of positrons in the calorimeters, and the right plot shows the tracker measurement of the vertical beam position. Both show a strong correlation with the size of the bump; as the vertical position increases, the size of the bump increases. This suggests that the bump is related to changes in the radial field, which are indirectly suggested by the other field metrics. This study also implies that changes of the beam position of 0.5mm or less are causing significant changes in the behavior of the lost muons, leading to a coherent increase late in time. This behavior remains unexplained by simulation. Regardless, all the Run-2 datasets can still be fit together using a summed triples spectrum.

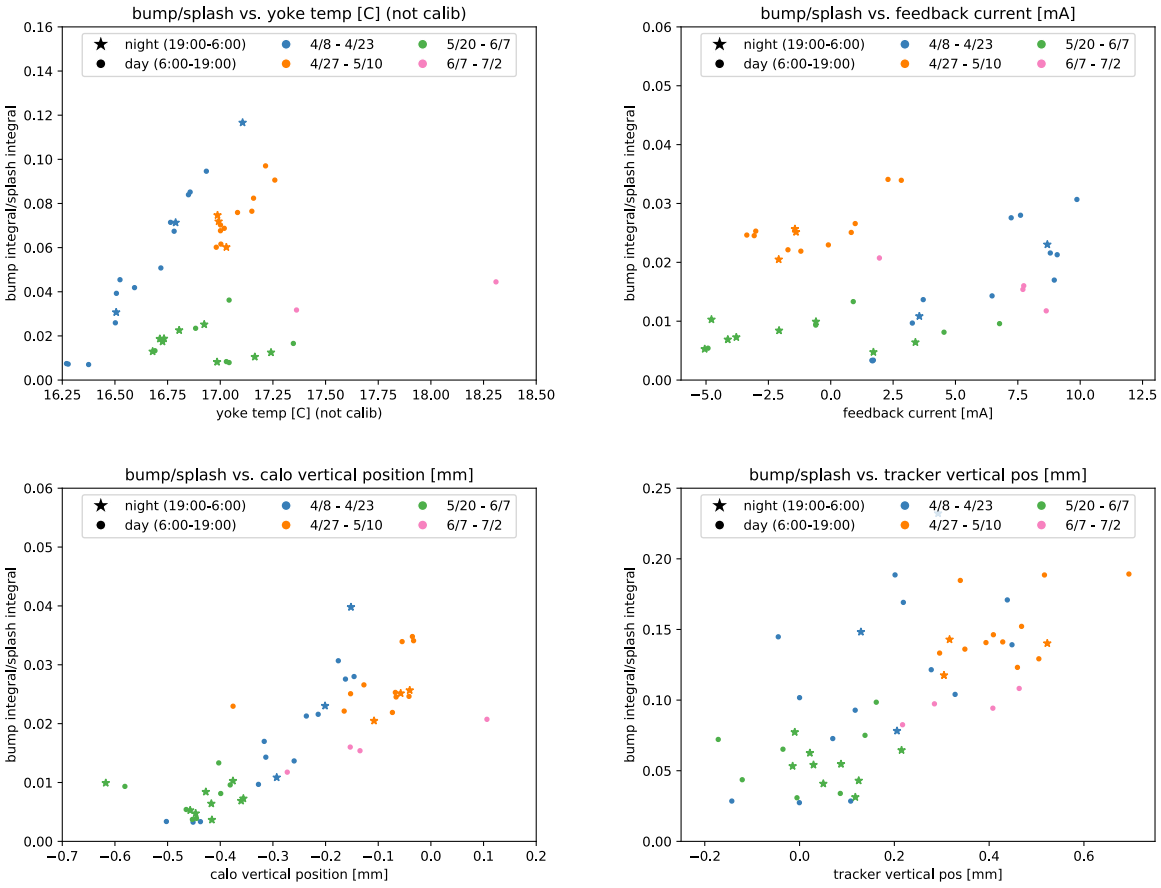


Figure 5.20: Correlation between lost muon bump size (parameterized by integral of muon bump region to early time "splash" region) and magnet and beam parameters. The correlations with magnet yoke temperature (top left) and feedback current (top right) suggest a relationship between temperature-dependent magnet changes and the size of the bump. The bottom plots shows a strong correlation with average hit position of positrons in the calorimeters (left) and the vertical beam position measured by the tracker (right), suggesting that a temperature-dependent radial field change is likely to have caused the lost muon bump.

The muon losses are incorporated into the fit function using the triples spectrum, which provides the time dependent behavior of the losses. The efficiency of measuring a lost muon is unknown, so a lost muon scaling factor is provided by the fit. The correction factor for the lost muons is [3, 63]

$$\Lambda(t) = 1 - K_{loss} \int_0^t e^{t/\tau} L(t') dt', \tag{5.60}$$

where $L(t)$ is the triples spectrum shown in Figure 5.18 (right) and K_{loss} is the scaling factor. K_{loss} is related to the efficiency of measuring a lost muon ϵ by

$$K_{loss} \equiv \frac{1}{\epsilon N_0}. \tag{5.61}$$

K_{loss} is therefore dependent on N_0 and also on ϵ. This correction multiplies N_0:

$$N_0 \rightarrow N_0 \Lambda(t), \tag{5.62}$$

so that the muon loss correction factor reduces the number of stored muons relative to N_0 without losses. K_{loss} should always be positive so that stored muon number is decreasing instead of increasing.

This approach to the muon loss fitting depends on the ability of the fit to identify the correct scaling factor for the loss function. This was generally the case in Run-1, when the losses were large due to the damaged quadrupole resistors, and in Run-2, when the characteristic bump allowed the fit to determine the scaling. In Run-3, however, the losses were occurring at the muon lifetime, making K_{loss} and τ indistinguishable in the fit, because two terms are being fit to a single lifetime. This correlation introduces significant instability into the fit parameters, sometimes resulting in the uncertainty on R being incorrect. While the lifetime τ can be calculated from the momentum distribution of muons in the storage region, this distribution itself has an associated uncertainty which would need to be assigned to the lifetime, and the high correlation of K_{loss} and τ remains an issue.

In addition to the high degree of correlation between τ and K_{loss} in Run-3, the fit consistently prefers a negative value for K_{loss}, which is non-physical. This may be related either to the correlation or to a residual uncorrected slow term which can be absorbed by K_{loss}, as discussed in subsection 5.6.1. There are several

possible approaches to this problem. One is to fix K_{loss} to a value derived from the Run-2 data, which is positive and therefore physical. However, since the beam dynamics in Run-2 was clearly different than in Run-3, as evidenced by the large coherent bump in the Run-2 spectrum, it is not necessarily true that the efficiency of measuring a lost muon would be the same in both runs. For this analysis, the K_{loss} parameter was removed entirely (set to 0) for Run-3BM and Run-3NO. The effect on χ^2 is minimal, and the over-fitting problem when both parameters are allowed to float is avoided. The systematic uncertainty due to this correlation will be discussed further in subsection 6.7.1.

With this additional parameter, the final fit function is

$$N(t) = N_0 e^{-t/\tau} \Lambda(t) \left[1 + A(t) \cos(\omega_a(R) t + \phi(t))\right] \; N_{CBO,VW}(t) \; N_{2CBO}(t) \; N_y(t). \tag{5.63}$$

There are 26 fit parameters if all parameters are allowed to float.

5.4 Final fit results

The full fit function, Equation 5.63, is now applied to the histogram. The FFTs of fit residuals for the three datasets (Run-2, Run-3BM, Run-3NO) are shown in Figure 5.21. All datasets have flat FFTs, indicating that all visible oscillations have been accounted for. No peaks are visible at the vertical dashed lines where peaks were expected to be present, indicating that the fit has successfully removed them. Further, a good χ^2 is achieved for all datasets, as listed in Table 5.3.

The fit choices are kept as similar as possible between datasets, but there are several differences required by the statistics or unique conditions of each. Firstly, τ_y is fixed to 96 μs in Run-2 and Run-3NO, which had lower statistics than Run-3BM, where τ_y was allowed to float. The value of 96 μs was determined using the procedure described in section 5.3.1 and is consistent with the free floating τ_y parameter from Run-3BM. Additionally, K_{loss} was fixed to 0 in Run-3BM and Run-3NO. The distribution of several parameters and uncertainties across an ensemble of 100 random seeds is shown in Figure 5.22. The red line corresponds to the case when K_{loss} is allowed to float in the fit, whereas the blue line corresponds to the case when K_{loss} is fixed to 0. A Gaussian function is fit to the histograms when relevant, and the mean of the Gaussian distribution is given. First, R is largely unaffected by the choice to fix K_{loss}, with a difference in the average

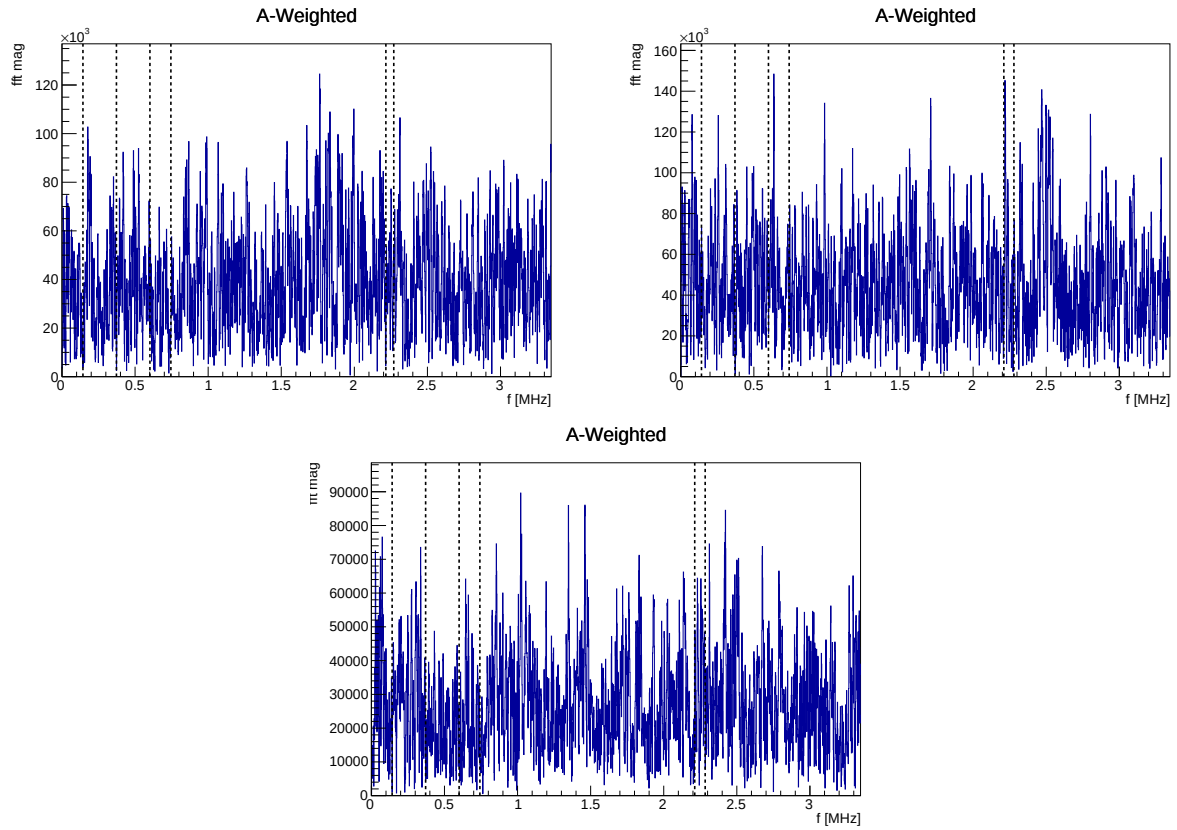


Figure 5.21: Final A-weighted fit fast Fourier transform of residuals for the three datasets (Run-2, Run-3BM, Run-3NO). A flat FFT has been achieved, showing that all oscillations are accounted for in the fit.

value of only 3 ppb. The error on R becomes much more stable when K_{loss} is fixed, whereas when K_{loss} is allowed to float, the error varies significantly, which is a nonphysical result. Parameters correlated with K_{loss} such as τ also become much better determined. Finally, the reduced χ^2 is actually improved when K_{loss} is fixed due the improved stability of the fit. These plots justify the choice to fix K_{loss} to zero for the final results.

The fit parameters are defined in Table 5.2, and their values for the three datasets are given in Table 5.3. The parameters indicated with a star (χ^2/ndf and R) are taken from distributions of 100 random seeds, like those shown in Figure 5.22, while the rest of the parameters are taken from a single random seed. Further, all scans, sub-dataset fits, and crosscheck plots are taken from a single random seed.

Several conclusions can be made from the table of fit results. First of all, a high fit quality was achieved for all three datasets, with reduced χ^2 of 4238/4130, 4142/4130, and 4158/4132, respectively. Run-3BM was

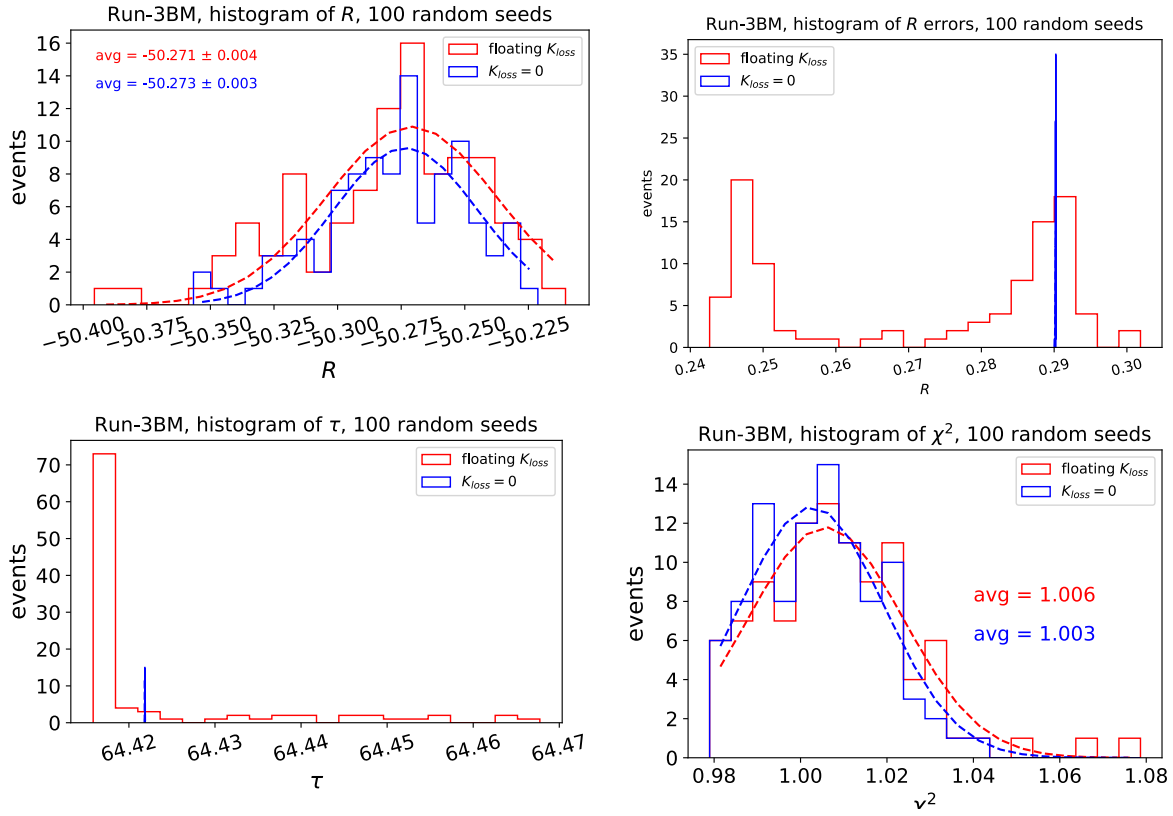


Figure 5.22: Histograms of Run-3BM A-weighted fit results from an ensemble of 100 random seeds. A Gaussian function is fit to the results where relevant in order to extract the mean. The results with a floating K_{loss} parameter and K_{loss} fixed to zero are compared. Top left: histogram of R, in units of ppm, showing a difference of 3 ppb between the two distributions. The final value of R is extracted from this plot. Top right: error on R, showing that the fit determines a much more stable value of the error when K_{loss} is fixed. Bottom left: the muon lifetime τ, showing that parameters correlated with K_{loss} such as τ are also more stable when fixing K_{loss}. Bottom right: reduced χ^2 distributions, showing that the χ^2 is slightly improved when fixing K_{loss}.

Fit parameter	Definition	Notes
N_0	Number at $t = 0$	
τ	Muon lifetime	
A	Precession frequency asymmetry	
ϕ	Precession frequency phase	
R	Blinded precession frequency	
τ_{CBO}	CBO decoherence lifetime	Changes significantly if constant offset C_{CBO} is added.
A_{CBO}	CBO amplitude	
ϕ_{CBO}	CBO phase	
ω_{CBO}	CBO frequency	Modified by $A_{CBO,ch}$ and $\tau_{CBO,ch}$ in per-calo fits. Modified by tracker time changing CBO parameters in full fit.
τ_{VW}	VW decoherence lifetime	
A_{VW}	VW amplitude	
ϕ_{WW}	VW phase	
ω_{VW}	VW frequency	
K_{loss}	Muon loss normalization	Dependent on the application of the residual gain correction. Negative when allowed to float in Run-3. Set to 0 with limited effect on χ^2.
$A_{CBO,A}$	CBO A modulation	
$\phi_{CBO,A}$	Phase of CBO A modulation	
$A_{CBO,\phi}$	CBO ϕ modulation	
$\phi_{CBO,\phi}$	Phase of CBO ϕ modulation	
A_{2CBO}	Amplitude of 2*CBO	
ϕ_{2CBO}	Phase of 2*CBO	
τ_y	y decoherence lifetime	Cannot be fit accurately in Run-2 and Run-3NO, fixed to 96 μs.
A_y	y oscillation amplitude	
ϕ_y	y oscillation phase	
ω_y	y oscillation frequency	
A_{VW-CBO}	Amplitude of VW-CBO beat frequency	
ϕ_{VW-CBO}	Phase of VW-CBO beat frequency	
C_{CBO}	Constant offset for CBO envelope	Only used for per-calorimeter fits.
$A_{CBO,ch}$	Amplitude of CBO frequency change	Only used for per-calorimeter fits.
$\tau_{CBO,ch}$	Lifetime of CBO frequency change	Only used for per-calorimeter fits.

Table 5.2: Table of standard fit parameters for the Run-2/3 analysis.

Fit parameter	Run-2	Run-3BM	Run-3NO
χ^2/ndf*	4238/4130	4142/4130	4158/4132
N_0	(2.019 $\pm$ 0.002) $\times 10^7$	(2.67262 $\pm$ 0.00003) $\times 10^7$	(0.95554 $\pm$ 0.00002) $\times 10^7$
τ	64.431 $\pm$ 0.001	64.4219 $\pm$ 0.0005	64.3994 $\pm$ 0.0008
A	0.36123 $\pm$ 0.00001	0.362999 $\pm$ 0.000009	0.36167 $\pm$ 0.00002
ϕ	-2.16547 $\pm$ 0.00006	-2.16232 $\pm$ 0.00005	-2.17516 $\pm$ 0.00008
R [ppm]*	-59.520 $\pm$ 0.337	-50.273 $\pm$ 0.290	-49.326 $\pm$ 0.487
τ_{CBO}	261 $\pm$ 7	247 $\pm$ 6	234 $\pm$ 19
A_{CBO}	0.00386 $\pm$ 0.00003	0.00318 $\pm$ 0.00003	0.00172 $\pm$ 0.00005
ϕ_{CBO}	-0.319 $\pm$ 0.008	-0.157 $\pm$ 0.009	-0.19 $\pm$ 0.03
ω_{CBO}	2.3407 $\pm$ 0.0001	2.3291 $\pm$ 0.0001	2.3319 $\pm$ 0.0003
τ_{VW}	23 $\pm$ 3	31 $\pm$ 4	25 $\pm$ 6
A_{VW}	0.0023 $\pm$ 0.0005	0.0010 $\pm$ 0.0002	0.0008 $\pm$ 0.0004
ϕ_{VW}	2.0 $\pm$ 0.2	0.6 $\pm$ 0.2	2.3 $\pm$ 0.5
ω_{VW}	14.262 $\pm$ 0.005	14.320 $\pm$ 0.004	14.35 $\pm$ 0.01
K_{loss}	9.0 $\pm$ 0.6	0 (fixed)	0 (fixed)
$A_{CBO,A}$	0.00035 $\pm$ 0.00005	0.00013 $\pm$ 0.00005	0.00012 $\pm$ 0.00009
$\phi_{CBO,A}$	-0.15 $\pm$ 0.16	0.3 $\pm$ 0.4	-1.1 $\pm$ 0.7
$A_{CBO,\phi}$	0.00009 $\pm$ 0.00006	0.00012 $\pm$ 0.00005	0.00023 $\pm$ 0.00009
$\phi_{CBO,\phi}$	1.0 $\pm$ 0.6	1.0 $\pm$ 0.4	0.9 $\pm$ 0.4
A_{2CBO}	0.00017 $\pm$ 0.00002	0.00014 $\pm$ 0.00002	0 (fixed)
ϕ_{2CBO}	3.0 $\pm$ 0.1	-2.7 $\pm$ 0.1	0 (fixed)
τ_y	96 (fixed)	104 $\pm$ 19	96 (fixed)
A_y	0.00022 $\pm$ 0.00003	0.00038 $\pm$ 0.00004	0.00034 $\pm$ 0.00004
ϕ_y	0.2 $\pm$ 0.2	0.3 $\pm$ 0.1	0.4 $\pm$ 0.3
ω_y	13.922 $\pm$ 0.004	13.887 $\pm$ 0.002	13.899 $\pm$ 0.004
A_{VW-CBO}	0.0009 $\pm$ 0.0003	0.0009 $\pm$ 0.0002	0.0016 $\pm$ 0.0006
ϕ_{VW-CBO}	3.0 $\pm$ 0.2	1.8 $\pm$ 0.2	2.5 $\pm$ 0.5

Table 5.3: A-weighted fit parameters for Run-2, Run-3BM, and Run-3NO. * indicates that the value is taken from a distribution of 100 random seeds. All other parameters are taken from a representative seed.

the highest statistics dataset, with a precision of 290 ppb. Run-2 had a precision of 337 ppb and Run-3NO had a precision of 483 ppb, the lowest statistics dataset, meaning that the three datasets will have a combined statistical precision of 200 ppb. Additionally, A_{CBO} was about 2 times lower in Run-3NO compared to Run-2 and Run-3BM, as expected due to the improved kick strength. All parameters that remained free floating in the fits are well defined.

The full covariance matrix for Run-3BM is given in Figure 5.23, showing that most parameters are largely uncorrelated with R, which allows for it to be measured precisely. Only the phase ϕ is highly correlated, as discussed in subsection 3.2.1. Otherwise, CBO parameters are the most correlated with R, including τ_{CBO} and A_{CBO}, because the CBO causes the largest modulation of the positron rate and because it lasts late into the fill. Otherwise, the correlation is under 0.02. Full covariance matrices for the other two datasets are given in Appendix B.

5.5 Calorimeter binned analysis

Effects such as the CBO are largely cancelled by summing data from all calorimeters together, because calorimeters across the ring from each other have opposite CBO phases. In order to ensure that beam dynamics effects are properly understood, the oscillation is exaggerated by fitting data from a single calorimeter only. The fit parameters are then plotted as a function of calorimeter number and checked for consistency.

The necessity of randomization at the cyclotron period, as discussed in subsection 5.2.1, was first discovered when fitting data from single calorimeters. The residual cyclotron signal was present as an "R-wave" or R oscillation as a function of calorimeter number. This R wave is shown in Figure 5.24 in blue. A sine function has been fit to the data, with the wavelength fixed to be one rotation around the ring. The result after cluster times have been randomized at the cyclotron period is shown in red. The amplitude of the oscillation has been reduced by a factor of ≈ 2.5. Note that since the amplitude of the sine wave cannot be negative, a nonzero amplitude will always be extracted from the data. The reduction in the amplitude of the R-wave justifies the choice to randomize.

When fitting Run-2 and Run-3 data by calorimeter, a fitting issue that was not apparent in Run-1 arose. In some calorimeters, a residual peak remained at the CBO frequency even after applying the full fit function validated on the full dataset, as seen in Figure 5.25 (top). The size of this residual peak varied significantly

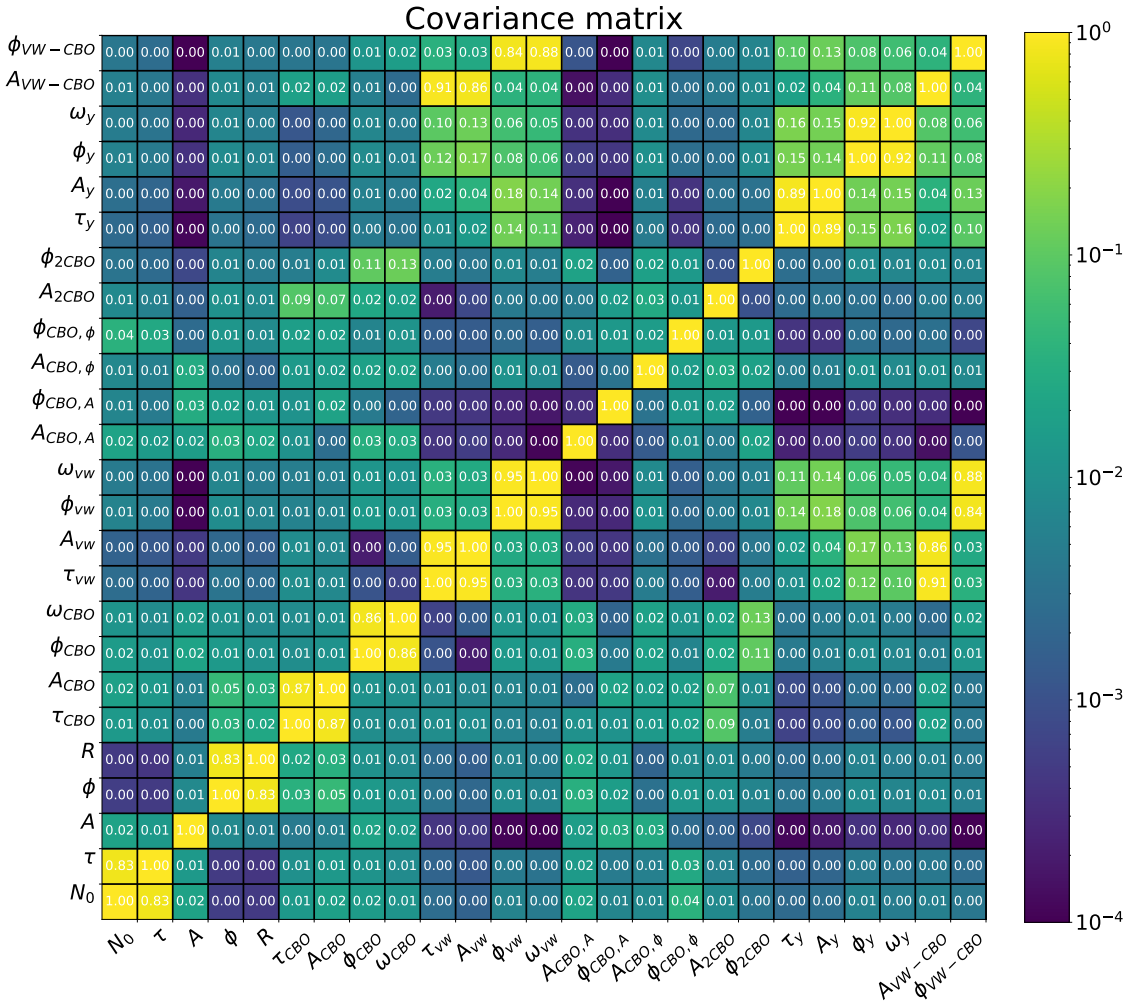


Figure 5.23: Run-3BM full fit covariance matrix.

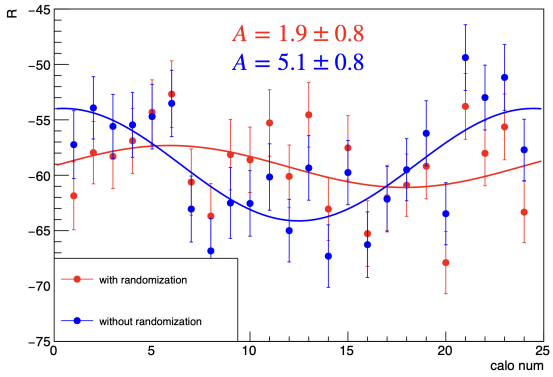


Figure 5.24: The value of R, in ppm, vs. calorimeter number in Run-2C with and without randomization at the cyclotron period. A sine fit has been applied. The oscillation has been reduced by a factor of ≈ 2.5 by applying the randomization procedure.

across calorimeter; for example, calorimeter 3 (left) saw such a peak and calorimeter 18 (right) did not.

A remaining peak at the CBO frequency indicates that either the CBO envelope has not been correctly modeled or that the CBO frequency is changing as a function of time. First, different functional forms of the CBO envelope were used. The default form of the envelope is $E_{CBO}(t) = e^{-t/\tau_{CBO}}$, but this form is not necessarily the most ideal. By introducing a constant offset, $E_{CBO}(t) = e^{-t/\tau} + C_{CBO}$, the remaining CBO peak was reduced slightly, as seen in Figure 5.27, but the problem remained.

A residual peak could also be caused by a time changing frequency. Because the behavior is calorimeter-dependent, it must be acceptance-related. Each calorimeter has a different likelihood of measuring a positron as a function of radial position, which affects the measured CBO frequency. As the CBO oscillation decoheres, each detector measures a different portion of the beam, resulting in unique time-dependent behavior in different calorimeters. Some of this behavior has been replicated in simulation.

In order to determine whether there was evidence of a time changing ω_{CBO}, fits were performed on narrow time windows for each calorimeter, with τ_{CBO} fixed to a high number. Each window was $4T_a$ wide, where $T_a \approx 4.37$ is the ω_a period. This window width allowed for fits to be successfully performed while still preserving time dependent behavior. Because ω_{CBO} and ϕ_{CBO} are highly correlated, $\omega_{CBO}t - \phi_{CBO}$ was calculated for each window. Then, a linear function was fit to the data and iteratively subtracted off to remove the linear behavior which would occur if ω_{CBO} was constant. Finally, only the nonlinear behavior remained. Figure 5.25 (bottom) shows the $\omega_{CBO}t - \phi_{CBO}$ residuals corresponding to the above FFTs.

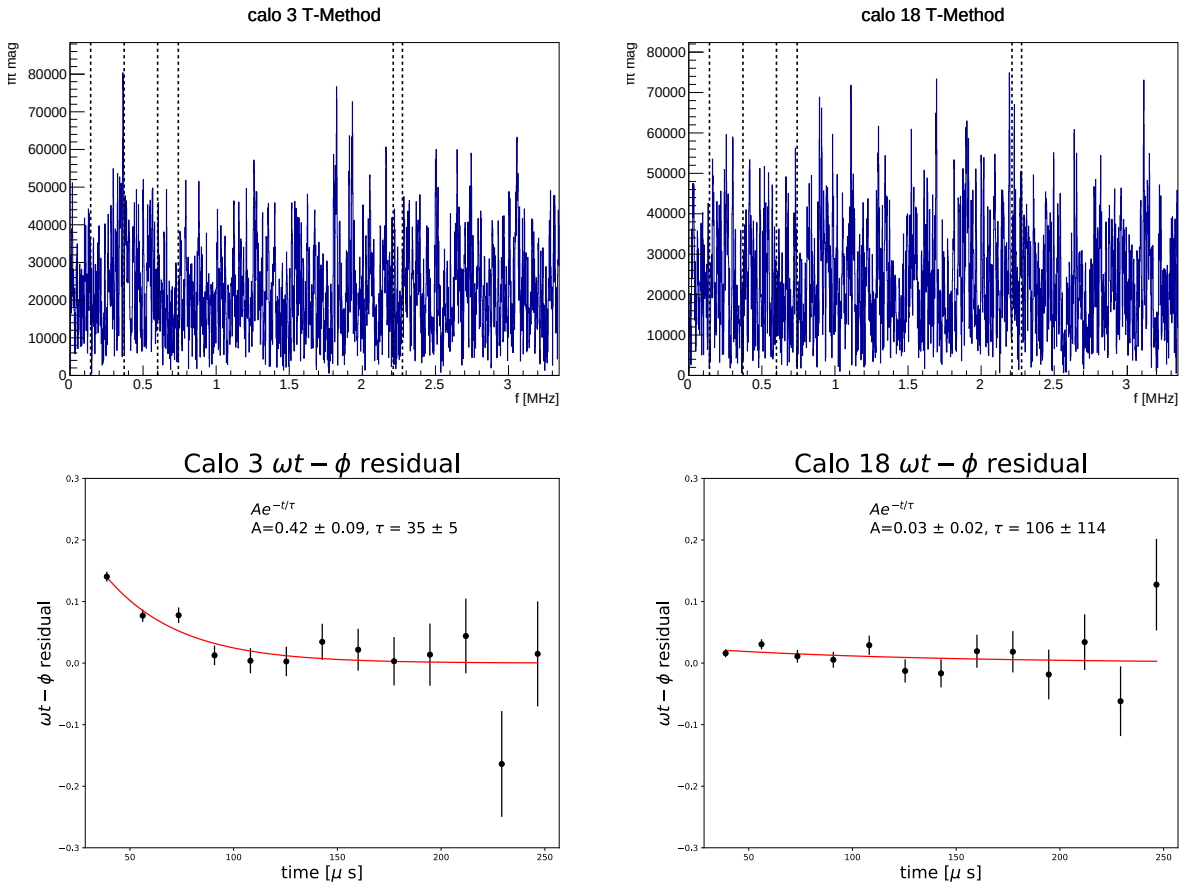


Figure 5.25: Top: Representative FFTs after a full ω_a fit has been applied to a single calorimeter. Some calorimeters, like calorimeter 3 (left) show a large residual peak at the CBO frequency ($\approx$ 0.37 MHz), whereas others, such as calorimeter 18 (right) show no peak at the CBO frequency. Bottom: $\omega_{CBO}t - \phi_{CBO}$ is calculated for each calorimeter from fits over short time windows. The linear behavior is subtracted off. The residual is then fit to an exponential decay to study the degree of nonlinear behavior. Calorimeter 3, which has a larger residual CBO peak, has a higher degree of nonlinear behavior.

Calorimeter 3, which has a large residual CBO peak, has a high degree of nonlinear behavior, whereas calorimeter 18, which has no visible residual CBO peak, has limited nonlinear behavior. The residual was then fit to an exponential decay function of the form $Ae^{-t/\tau}$, with the results given in the figure.

In order to incorporate this information into the fit, two free time changing CBO parameters were introduced in the form

$$\omega_{CBO} \rightarrow \omega_{CBO} + A_{CBO,ch} e^{-t/\tau_{CBO,ch}}, \tag{5.64}$$

where $A_{CBO,ch}$ and $\tau_{CBO,ch}$ are the time changing CBO amplitude and lifetime, respectively. This also follows the form of the time changing CBO frequency due to scraping measured by the trackers, as discussed in section 5.3.1 (Equation 5.41). Introducing these terms suppresses the residual peak, as shown in Figure 5.27, so that with the constant envelope offset C_{CBO} and these time changing parameters combined, the peak is fully suppressed. Because the parameters are ill defined in some calorimeters with lower residual CBO peaks and so that the amplitudes can be compared, $\tau_{CBO,ch}$ is fixed to 30 μs. These parameters are plotted as a function of calorimeter in Figure 5.26. Because the CBO frequency changing behavior is different across calorimeters, and because the detectors are all sampling the same beam, the differences must be due to the acceptance functions in the detectors. Figure 5.28 shows the CBO peak normalized to a flat region of the distribution as a function of detector number, with the figure shaded based on whether the detector is behind a quadrupole plate, a kicker plate, a tracking detector or free space, since the material in front of the calorimeter affects its acceptance function. The size of the peak is associated with the type of material in front of it, with the ESQ regions of the ring having especially large residual peaks. This plot also suggests that the scale of the effect is very similar in Run-2 vs. Run-3BM, meaning that an effect unique to either run (such as the lost muon bump, for example), is unlikely to be the source of the problem. Run-3NO has a lower CBO amplitude and also lower statistics, so the peak is much less visible.

Because the CBO oscillation is slightly correlated with R, the CBO fitting issue also causes an R wave, although one smaller in magnitude than the cyclotron-induced one. The difference between the R wave with and without the new CBO parameters is shown in Figure 5.27 (right). The amplitude of the oscillation is reduced by a factor of 1.5 by including the additional CBO terms. Although these additional terms are required for the per-calo fits, the ideal parameterization and the exact physics underlying it is unknown, and so a systematic uncertainty will be associated with the CBO frequency fitting in section 6.4. The additional

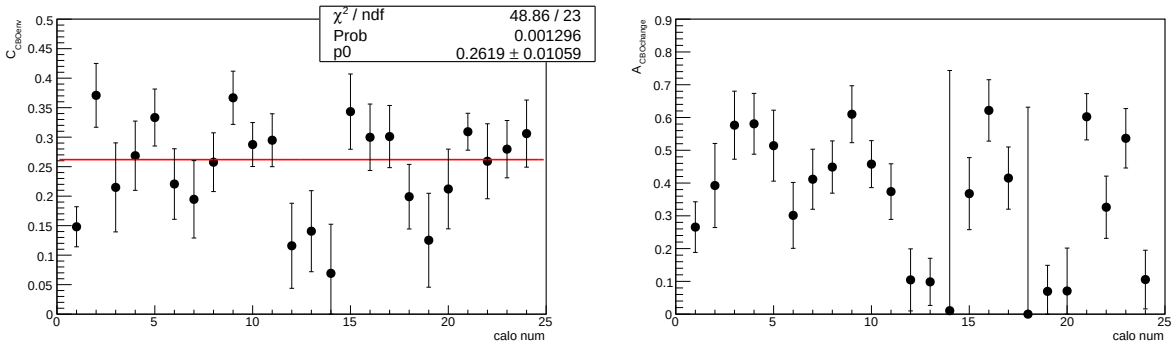


Figure 5.26: Additional CBO parameters for per-calorimeter fitting: C_{CBO}, the constant offset for the CBO envelope, and $A_{CBO,ch}$, the amplitude of the time changing CBO, as a function of calorimeter number. τ_{CBO} has been set to 30 μs so that $A_{CBO,ch}$ can be compared.

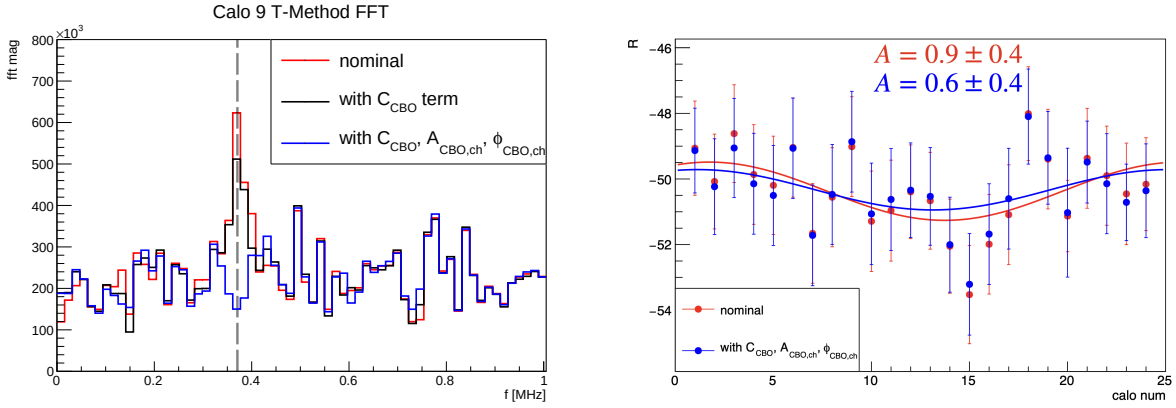


Figure 5.27: Left: Sample FFT from Run-3BM showing that the residual CBO peak is suppressed by the addition of a constant offset in the CBO envelope, C_{CBO}, and the time changing CBO parameters $A_{CBO,ch}$ and $\phi_{CBO,ch}$. The vertical dashed line indicates the CBO frequency. Right: reduction in R oscillation vs. calorimeter number after inclusion of additional CBO parameters. The value of R is given in ppm.

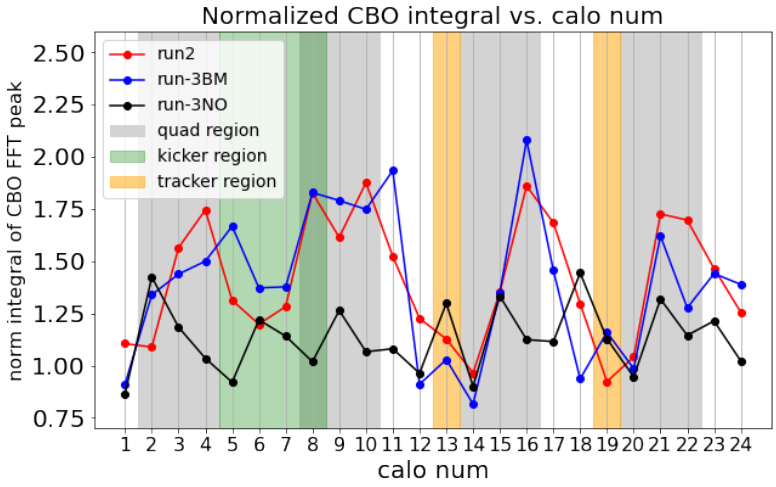


Figure 5.28: Residual CBO peak after the standard fitting procedure as a function of calorimeter number for the three datasets. The integral of the CBO peak has been normalized to a flat region of the distribution. The plot has been shaded based on whether that calorimeter is behind a kicker plate, quadrupole plate, tracker, or in free space. The association between regions of the ring and the size of the CBO peak indicates an acceptance-related effect.

parameters C_{CBO} and $A_{CBO,ch}$ are not necessary when the data from all calorimeters are added together, so they are not included in the final fit results table.

5.6 Energy binned analysis

Fits are also performed as a function of energy bin. This allows for consistency checking as a function of energy, as well as to generate the A vs. E curve required for the A-weighted analysis. Parameters are also checked for consistency as a function of energy. Figure 5.29 shows R and χ^2 as a function of energy for the Run-3BM dataset. Good fit quality is achieved across all datasets. The reduced precision on R around 1000 MeV, when the asymmetry is almost zero, is evident.

5.6.1 Slow term

In Run-1, the muon loss scaling factor K_{loss} was found to decrease in high energy bins. This remains true for the Run-2/3 analysis, as shown in Figure 5.30 for Run-2 (left) and Run-3BM (right), in red. This is a nonphysical result, as muon losses are unrelated to the positron energy. This behavior was found to be

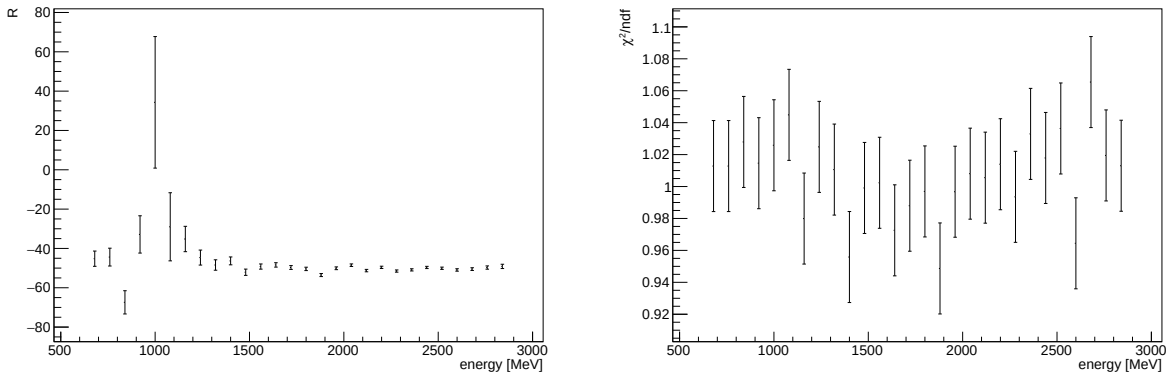


Figure 5.29: R and χ^2/ndf as a function of energy for Run-3BM, showing good fit quality across different energy bins.

consistent with an uncorrected gain sag. Because there are so few events in the highest energy bins, small gain changes affect high energy bins more than low energy bins; here, K_{loss} is absorbing other slow (low frequency) effects. For Run-1, the "gain sag" that would give rise to the K_{loss} behavior was $\approx 0.5 \times 10^{-3}$, around the level that could be undetectable by the laser system. In order to correct this behavior, a gain-like correction was applied to the energy-time precession frequency histogram of the form

$$N_g e^{-t/\tau}\big(1 + A_g \cos(\omega_a t - \phi)\big), \tag{5.65}$$

where N_g is the gain amplitude and A_g is the gain asymmetry. The values of ω_a and τ are taken from a fit performed on the data without the correction, thereby replicating a rate-dependent gain effect oscillating at the ω_a frequency. The parameters N_g and A_g were adjusted until K_{loss} was constant as a function of energy. The source of this gain-like change was unknown at the time of the Run-1 publication, and a systematic uncertainty was assigned to it [3].

For Run-2/3 and beyond, the source of this effect was extensively studied. The slow term is much smaller in analyses that use the Recon East reconstruction, as seen in Figure 5.31. The blue points correspond to a RW analysis, and the red points correspond to a RE analysis. This difference was leveraged by S. Foster to determine the pathology of Recon West that was not present in Recon East [78]. In summary, Recon West uses a "local" fitting approach where all crystal hits are fit separately and then combined into clusters. Each crystal hit is subject to its own fitting thresholds. On the other hand, Recon East uses a "global" fitting

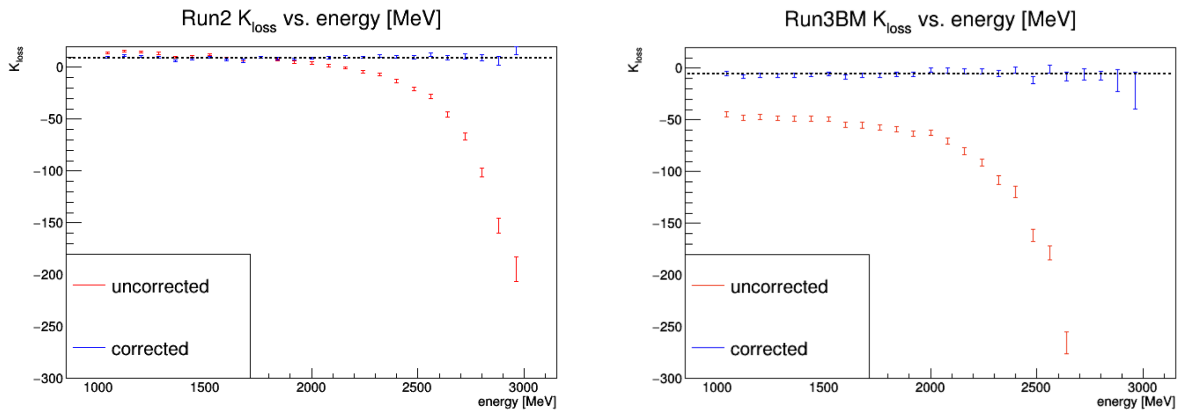


Figure 5.30: K_{loss} vs. energy uncorrected for the slow term (in red) and corrected for the slow term (in blue). A line is fit to the corrected points to demonstrate that the effect has been fully corrected. In Run-2 (left), K_{loss} is positive after being corrected. In Run-3BM (right), K_{loss} is negative after being corrected.

approach, where if one crystal hit exceeds an energy threshold, all crystals in an 3×3 square around it are fit, and crystal hit energies are allowed to be negative.

Recon West has both a primary and a secondary fitting threshold. An initial fit is performed if the trace amplitude is above a primary threshold. The residual is then taken of the fit. If the residual amplitude is greater than a secondary threshold, a second fit is performed. This allows the fitting algorithm to find multiple hits in the same island. Studies performed by S. Foster [78] demonstrated that the Recon West algorithm was missing hits in multi-cluster islands, i.e. islands from which multiple clusters were created. Because the secondary threshold was greater than the primary threshold ($\approx$ 50 MeV compared to $\approx$ 20 MeV), some pulses would be fit in single-cluster islands but not be fit in multi-cluster islands. A sample event display for a multi-cluster island is shown in Figure 5.32. The two clusters are outlined in red and blue. In Recon West, the two central crystals were not fit because they did not pass the secondary threshold. In Recon East, they were fit automatically because of the global fitting procedure.

This induces an early-to-late effect similar to a gain sag. At early times, the positron rate is higher, and a secondary pulse is more likely to be missed. At later times, nearly all islands are single-cluster islands, and the effect vanishes. Therefore the clusters at early times are generally missing a small amount of energy relative to similar clusters at late times. One possible solution to this issue is to equalize the thresholds and match them across all the crystals. However, since this update would require reprocessing all the data, the decision was made not to incorporate this change for Run-2/3. Reducing the threshold by too much can

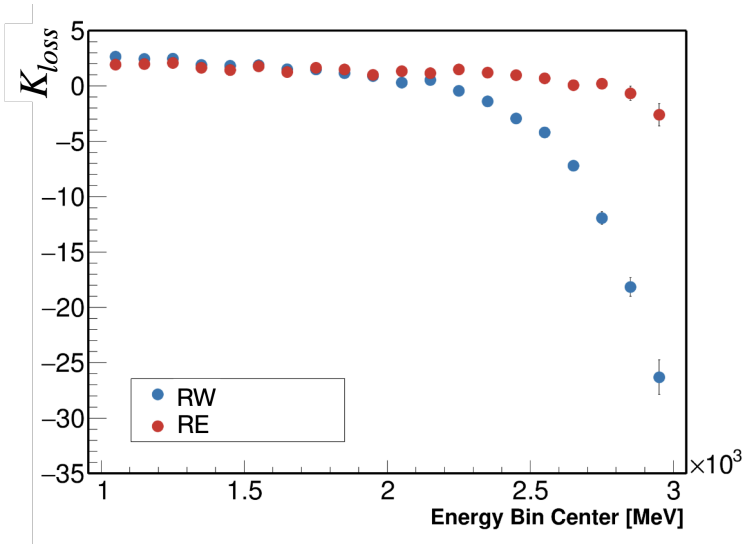


Figure 5.31: K_{loss} vs. energy in Recon East vs. Recon West, showing that the slow term causing the decrease in K_{loss} is significantly lower in Recon East. Reproduced from [78].

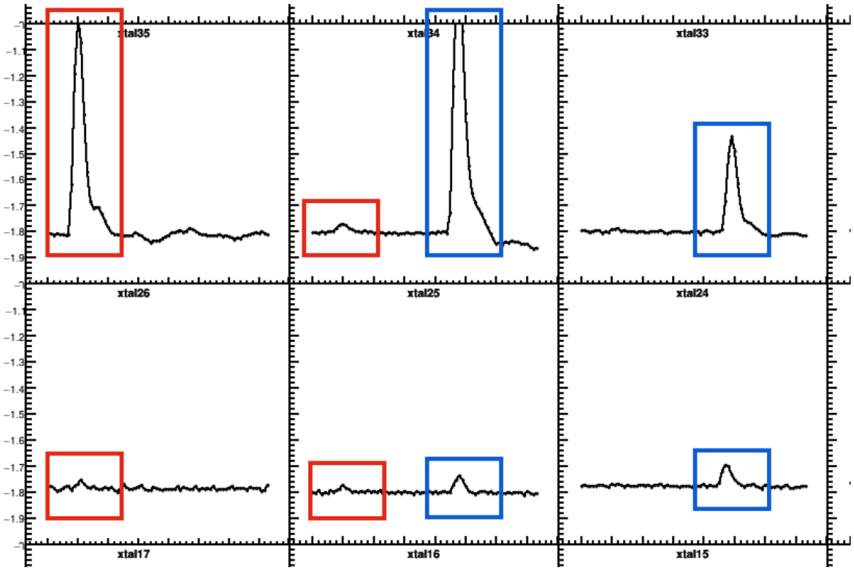


Figure 5.32: Event display showing a multi-cluster island, with the two clusters highlighted in red and blue. In Recon West, the two small hits in crystal (xtal) 16 and 25 would be below the secondary threshold and would not be fit. Reproduced from [78].

also cause problems when generating empirical pileup islands, since the amount of noise can become high relative to the signal. The slow term is likely to be an issue in Recon West going forward.

This new information about the source of the problem does justify the gain-like approach to the correction. The functional form in Equation 5.65 is consistent with the expected lifetime and behavior of the fitting error, in that the energy change is expected to be rate-dependent. Therefore, for this analysis, a gain-like correction is applied with the same procedure as the Run-1 analysis, and the correction parameters are varied to reduce the dependence of K_{loss} on energy. The parameters for the Run-2 correction are $N_g = 0.8 \times 10^{-3}$ for Run-2 and $N_g = 0.62 \times 10^{-3}$ for Run-3; the asymmetry is $A_g = 0.13$ for all datasets. The result of K_{loss} vs. energy after correction is shown in Figure 5.30 in blue, showing that K_{loss} is now flat as a function of energy. The χ^2 for the fits is also significantly improved after applying this correction. Note that K_{loss} is negative in Run-3BM even after correction; this may be due to an uncorrected slow effect. All fit results discussed in this chapter were performed after applying the residual gain correction, including the final fit results. The uncertainty assigned to this correction will be discussed further in section 6.6.

5.7 Start time, stop time, and sliding window scans

An important check of fit quality and consistency is performing the fit using different start times. If the fit parameters drift as a function of start time, this indicates that behavior at early start times may not be well-described. Scans of the asymmetry may reveal issues with the pileup or gain correction. Most importantly, R must be constant as a function of start time.

Start-time scans are performed with all but nine parameters (five basic parameters plus four CBO parameters) fixed. Because the scans extend until about 140 μs, oscillations with short lifetimes, such as the vertical betatron and vertical waist oscillations, are fixed to the values of the full fit. Otherwise, at later times, they cannot be adequately fit and can affect the uncertainties of other parameters. The start-time scans for the basic five parameters for Run-2 are shown in Figure 5.33. The black bands indicate the 1σ uncertainty on the parameter at each fit point. The blue lines indicate the allowed 1σ variation from the starting value, based only on the statistical uncertainty of the fit at each point; as the fit becomes less precise, the allowed variation of the parameter is greater. If no time-dependent behavior is present, the value is expected to be within the blue band 67% of the time. Note that some parameters are highly correlated with each other (for

example, R with the phase ϕ). All five of these parameters stay within the blue bands, indicating that the fit is adequately describing the behavior in the histogram over various start times.

Figure 5.34 shows the four CBO parameters τ_{CBO}, A_{CBO}, ϕ_{CBO}, and ω_{CBO} as a function of start time for Run-2. Although τ_{CBO} and A_{CBO} are within the bands, ϕ_{CBO} and ω_{CBO} drift outside them from about 50-120 μs. Phases and frequencies are highly correlated so the start time scans are evidence of the same underlying behavior. There is some evidence that the CBO frequency may be changing as a function of time, as discussed previously in section 5.5. When including the new parameters to account for these changes, the start time scan of ω_{CBO} is better behaved, as seen in Figure 5.36. Here $\tau_{CBO,ch}$ has been fixed to 25 μs, which is the average value of parameter for certain per-calorimeter fits where it is well-determined, and$A_{CBO,ch}$ is fixed to the value of the full fit for the start time scan. The improved stability of ω_{CBO} with these additional parameters suggests that they are warranted to describe the beam dynamics behavior.

Start-time scans are also performed for Run-3BM and Run-3NO. The five parameter scans are seen in Figure 5.35 and Figure 5.37. These scans are generally well behaved, with the possible exception of the start time scan of A in Run-3BM. This could indicate a pileup or gain related issue, although the pileup correction seems well-behaved in this dataset. Most importantly, R remains consistent over the start time scan.

The scans of CBO parameters for Run-3BM and Run-3NO can be seen in Figure 5.38 and Figure 5.39. Run-3BM has similar time changing CBO behavior to Run-2, which can be reduced by the inclusion of additional time changing CBO parameters. In Run-3NO, τ_{CBO} is fixed because the reduced CBO amplitude and lower statistics of that dataset make it hard to fit properly at later times. The parameters ϕ_{CBO} and ω_{CBO} seem better behaved, possibly because the impact of the CBO frequency is lower in Run-3NO due to the improved kick. Overall, the start time scans for all three datasets indicate good stability across start times, especially for R, which is the most important parameter.

Other scans can also be performed on the data to ensure fit stability. The start time scan is most commonly utilized because the beginning of the fill has both the highest statistics and the most unusual beam dynamics behavior. However, stop time scans (fits with different fit stop times) and sliding window scans (fits over narrow time windows across the fill) were also performed. A scan of R vs. stop time can be seen for all three datasets in Figure 5.40, showing good consistency as a function of stop time. For the sliding

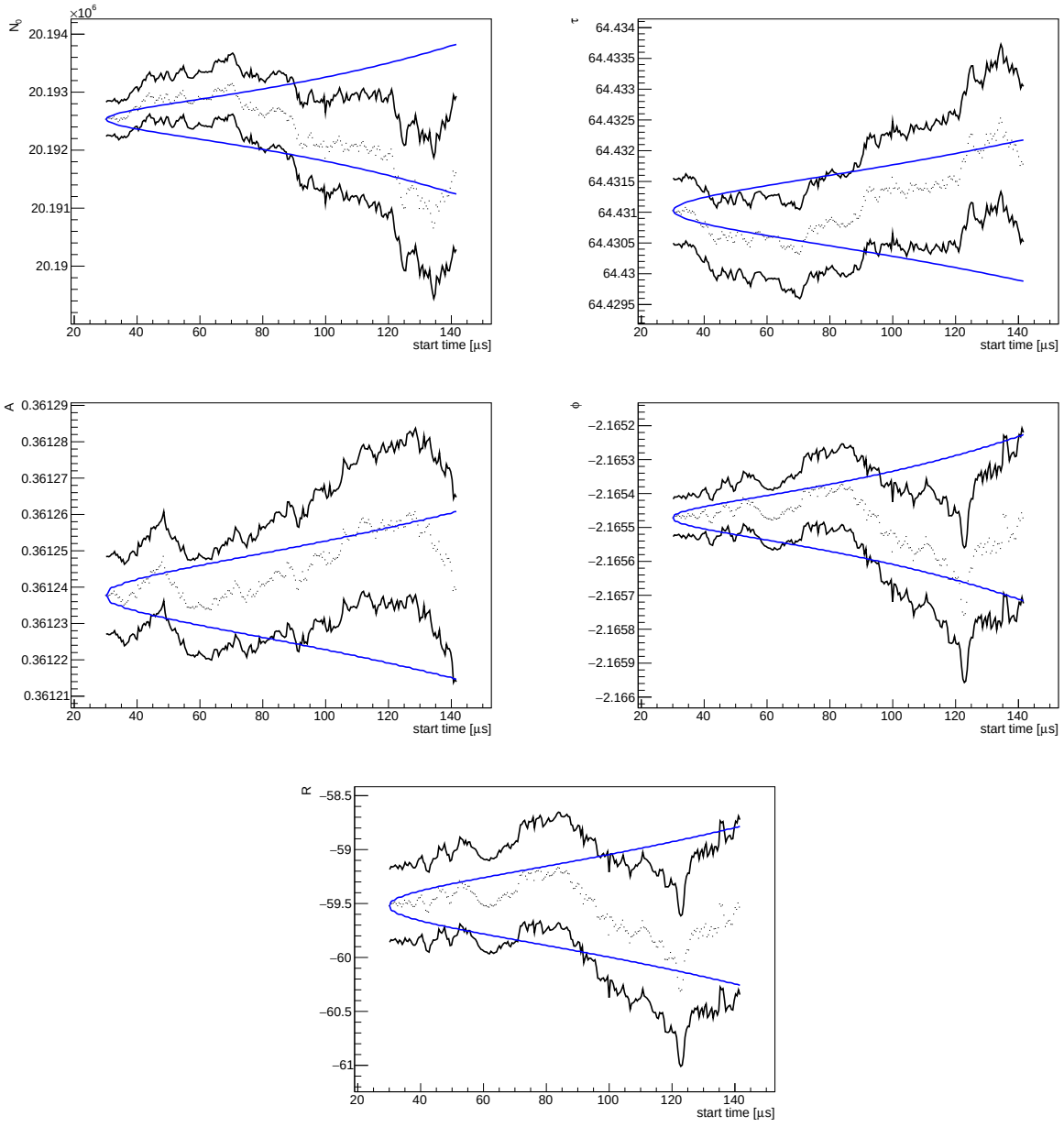


Figure 5.33: Run-2 A-weighted start time scans of the basic five parameters: N_0, τ, A, ϕ, and R. The blue lines indicate the allowed 1σ variation from the starting value. The black curves indicate the 1σ uncertainty at each fit point. All five parameters are generally consistent across the start time.

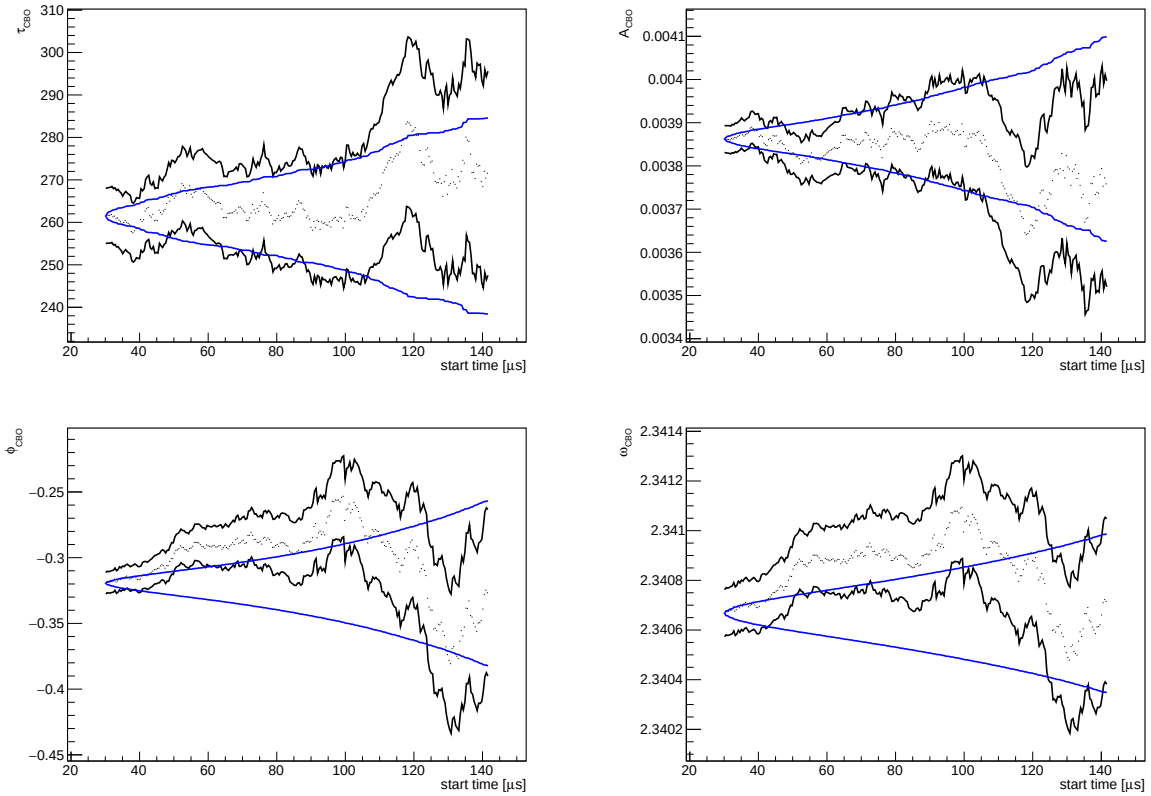


Figure 5.34: Run-2 A-weighted start time scans of the four CBO parameters: τ_{CBO}, A_{CBO}, ϕ_{CBO}, and ω_{CBO}. The blue lines indicate the allowed 1σ variation from the starting value. The black curves indicate the 1σ uncertainty at each fit point. The parameters ω_{CBO} and ϕ_{CBO}, which are correlated, drift slightly outside the blue bands, likely because of an uncorrected time changing CBO effect.

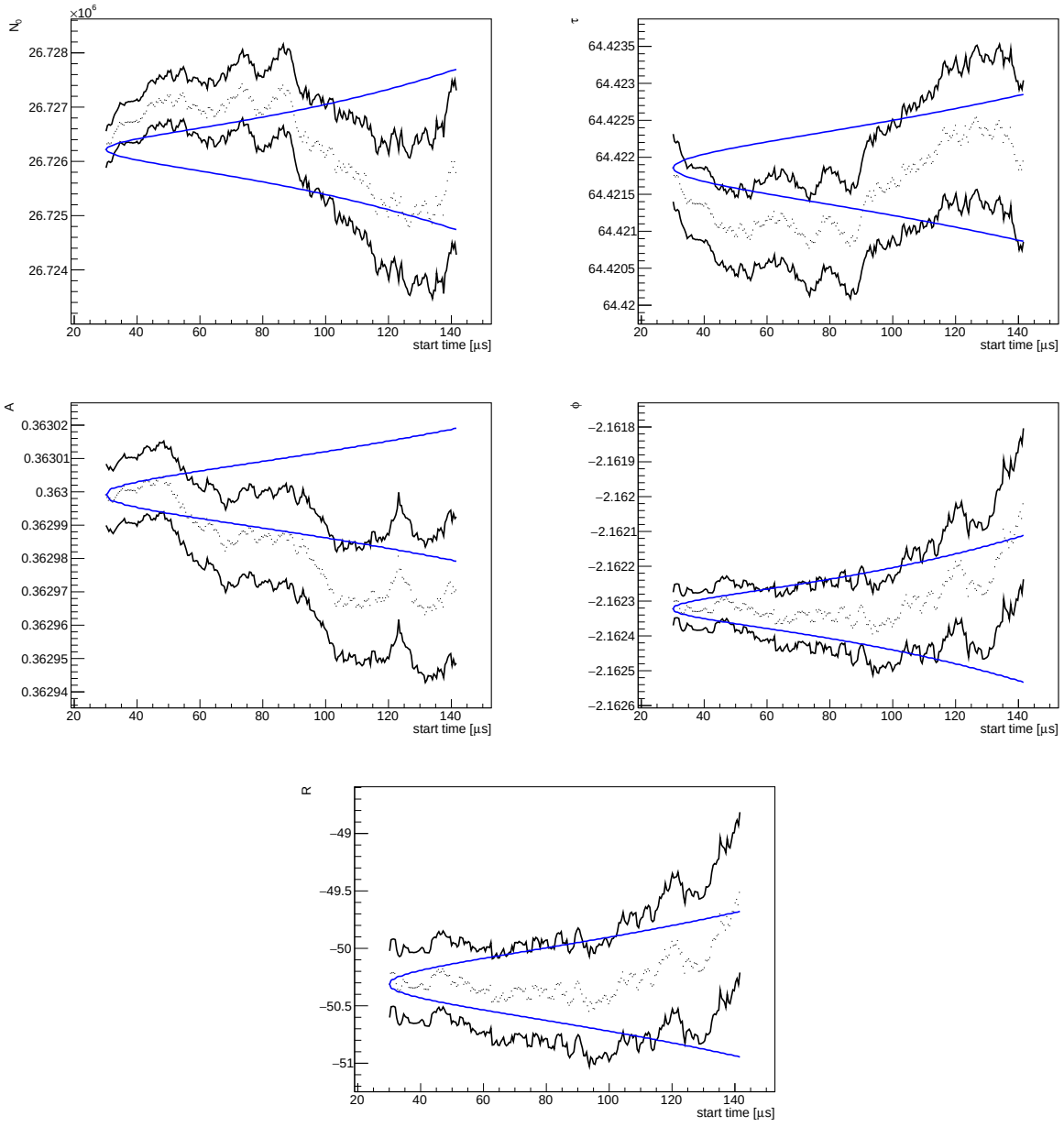


Figure 5.35: Run-3BM A-weighted start time scans of the basic five parameters: N_0, τ, A, ϕ, and R. The blue lines indicate the allowed 1σ variation from the starting value. The black curves indicate the 1σ uncertainty at each fit point. All five parameters are generally consistent across the start time.

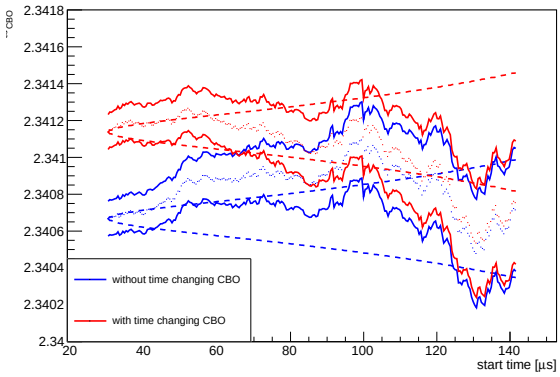


Figure 5.36: Comparison of start time scans with and without time changing CBO parameters. The parameter $\tau_{CBO,ch}$ has been fixed to 25 μs for ease of fitting. With the time changing CBO parameters, ω_{CBO} is more stable as a function of start time, indicating that the CBO behavior is better described. Here the dashed lines indicate the 1σ allowed statistical variation with respect to the starting parameter. The solid lines indicate the 1σ uncertainty on the value at each point.

window scans, the windows had a width of $4T_a$, where T_a is the ω_a period, or about 17 μs. The steps were a single ω_a period, or about 4 μs. The windows therefore overlap and are correlated with each other. The sliding window scans for R can be seen in Figure 5.41, also showing good consistency for R as a function of the time of the window.

Finally, Run-3 may be fit together to increase the statistics and determine whether any new effects emerge. Figure 5.42 shows a fit to Run-3BM and Run-3NO added together. Good fit quality is achieved and the fit parameters are well behaved. The combined precision on the A-weighted fit for the entirety of Run-3 is 250 ppb.

5.8 Fits per production dataset

In addition to performing fits on the larger datasets (Run-2, Run-3BM, and Run-3NO), fits are performed on the smaller production sub-datasets in order to cross check ω_a values against field values, check for fit quality, and examine beam dynamics trends over long periods of time. It is important to note that while Run-2 and Run-3 datasets have the same software blinding, they have different hardware blinding and the R values cannot be compared. Further, each sub-dataset fit is from a single random seed, and χ^2 and other parameter values can vary slightly with random seeds.

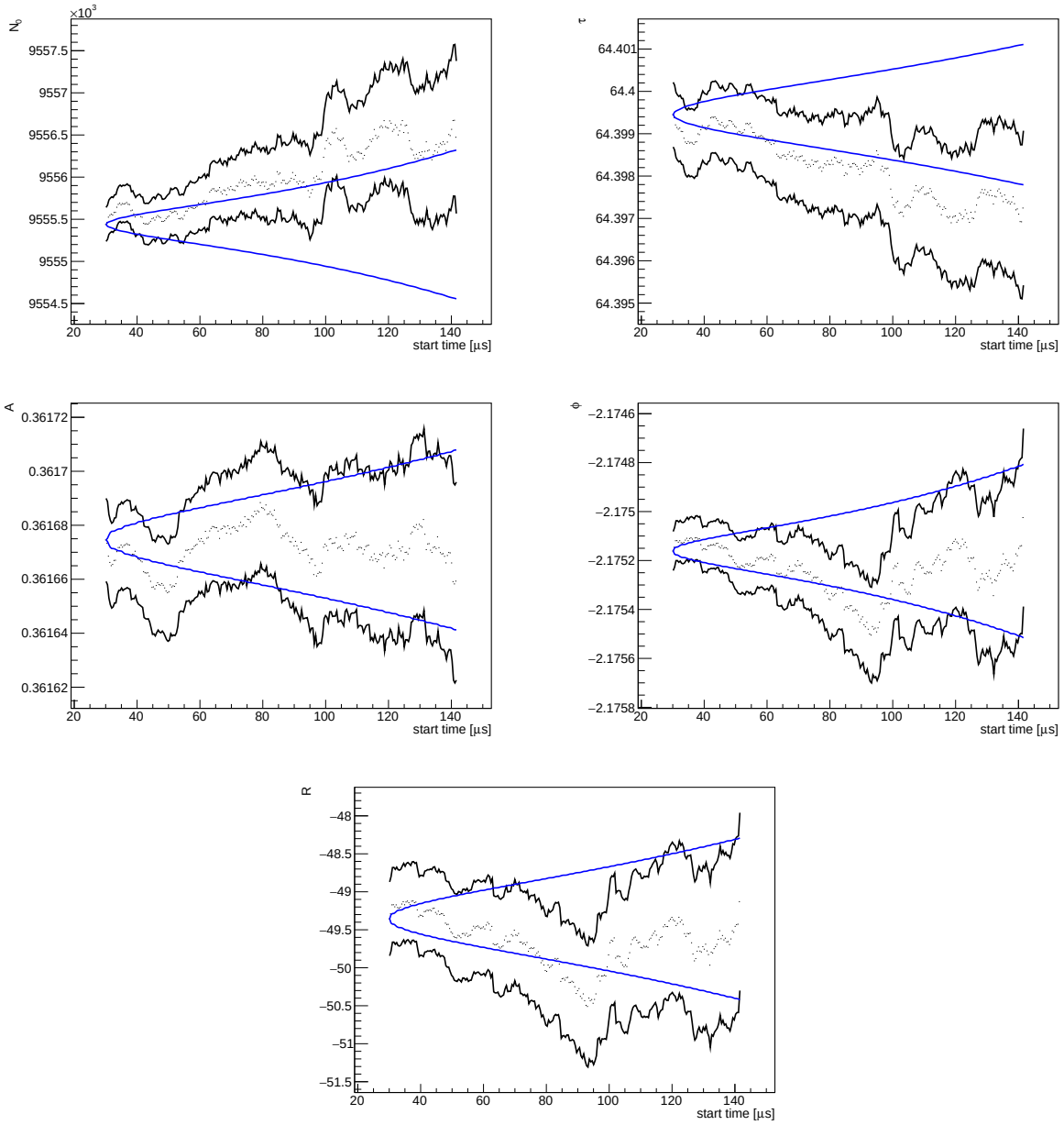


Figure 5.37: Run-3NO A-weighted start time scans of the basic five parameters: N_0, τ, A, ϕ, and R. The blue lines indicate the allowed 1σ variation from the starting value. The black curves indicate the 1σ uncertainty at each fit point. All five parameters are generally consistent across the start time.

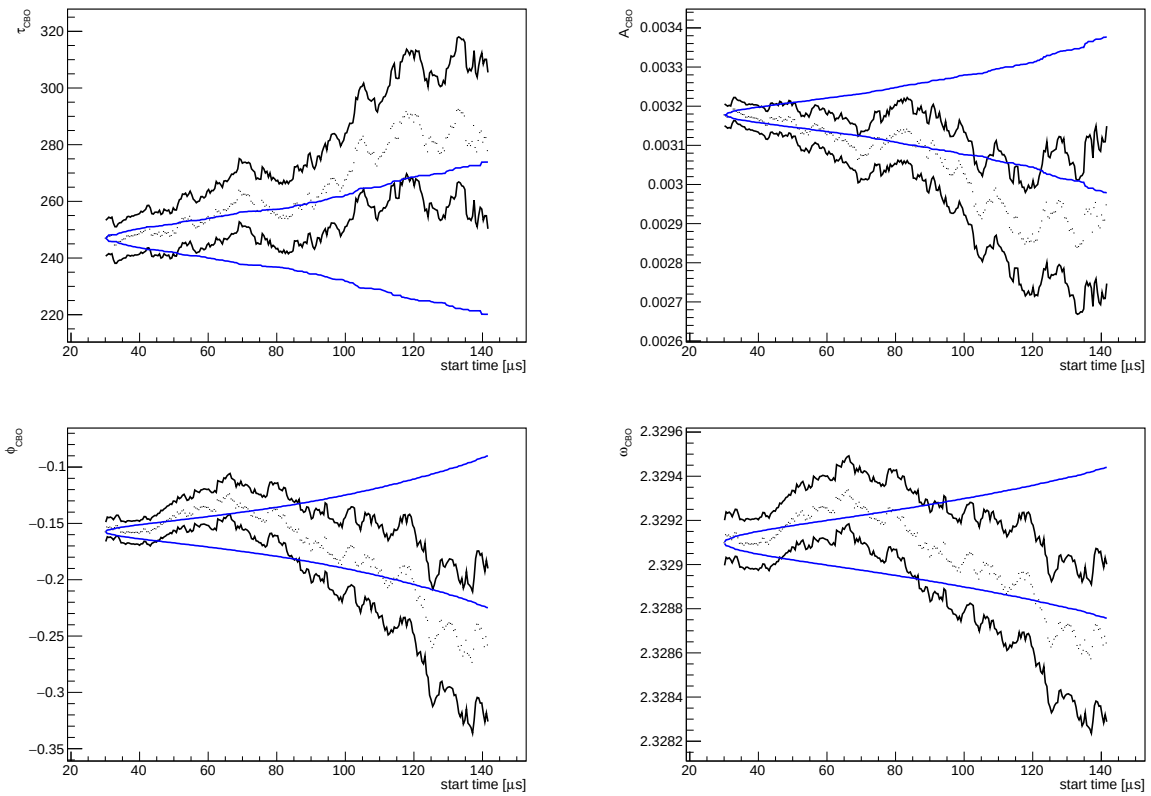


Figure 5.38: Run-3BM A-weighted start time scans of the four CBO parameters: τ_{CBO}, A_{CBO}, ϕ_{CBO}, and ω_{CBO}. The blue lines indicate the allowed 1σ variation from the starting value. The black curves indicate the 1σ uncertainty at each fit point. ω_{CBO} and ϕ_{CBO}, which are correlated, drift slightly outside the blue bands, likely because of an uncorrected time changing CBO effect.

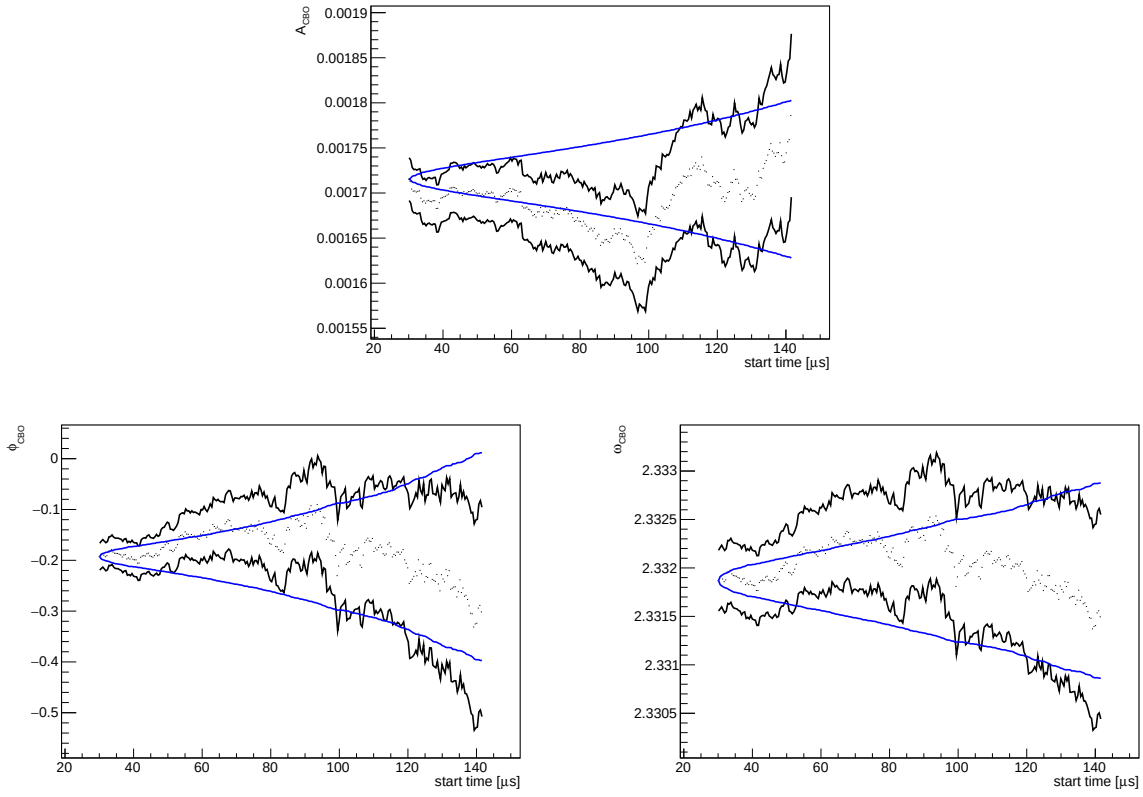


Figure 5.39: Run-3NO A-weighted start time scans of three CBO parameters: A_{CBO}, ϕ_{CBO}, and ω_{CBO}. τ_{CBO} has been fixed because of fitting difficulties with the lower CBO amplitude in Run-3NO. The blue lines indicate the allowed 1σ variation from the starting value. The black curves indicate the 1σ uncertainty at each fit point. ω_{CBO} and ϕ_{CBO}, which are correlated, drift slightly outside the blue bands, likely because of an uncorrected time changing CBO effect.

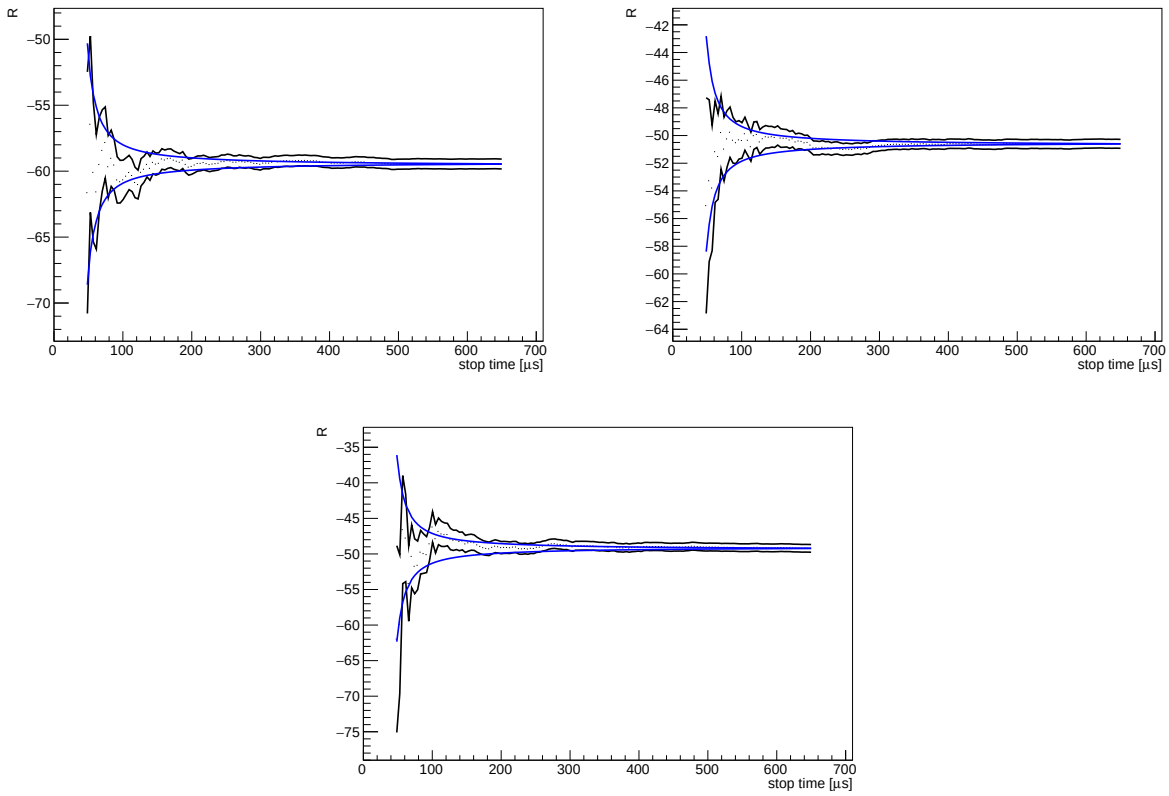


Figure 5.40: Stop time scans of R for the three datasets showing good consistency across stop times.

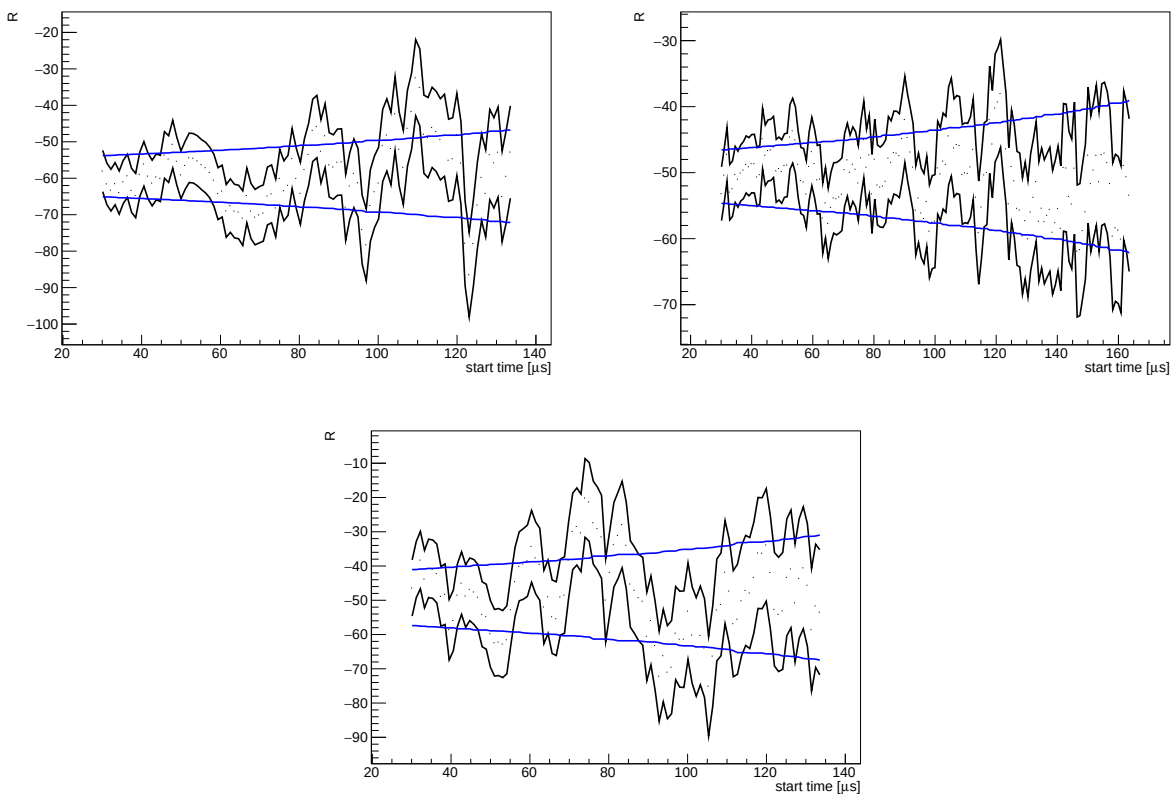


Figure 5.41: Fits performed over windows of 4 times the ω_a period with steps of one ω_a period for Run-2 (upper left), Run-3BM (upper right), and Run-3NO (bottom). The value of R is consistent as a function of the window time.

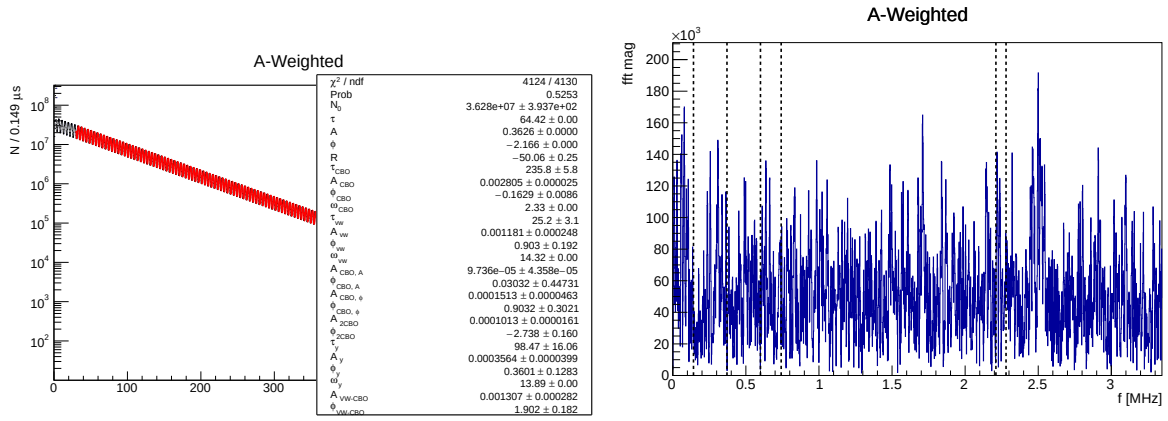


Figure 5.42: Fit and fast Fourier transform to data from Run-3BM and Run-3NO combined, showing good fit quality.

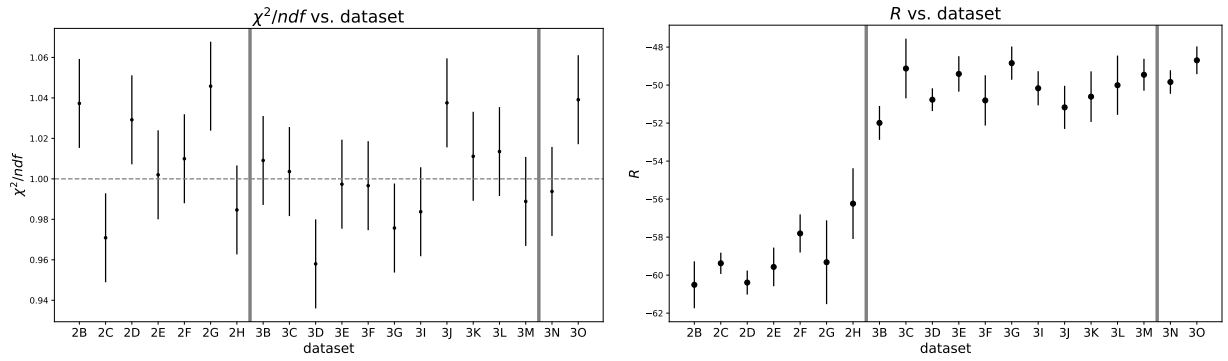


Figure 5.43: Left: χ^2/ndf vs. sub-dataset, showing χ^2/ndf around 1, shown with a dashed line. Right: R vs. sub-dataset. Note that Run-2 and Run-3 datasets have a different hardware blinding so their R values are not comparable.

Because the sub-datasets have lower statistics, certain fit parameters are not well-determined. Decisions to fix parameters to 0 or to the value for the larger dataset are made on a per-dataset basis.

Figure 5.43 shows the χ^2 / ndf and R values for the Run-2 and Run-3 sub-datasets. Vertical lines indicate the borders between larger datasets. The overall fit quality is good with a distribution around a reduced χ^2 of 1. The value of R is relatively consistent between Run-2 datasets and between Run-3 datasets; the values cannot be compared between Run-2 and Run-3. Figure 5.44 shows the CBO and vertical oscillation amplitudes vs. sub-dataset. The kicker hardware was gradually improved over Run-3BM, with a full kick finally achieved for Run-3NO. These improvements are reflected in the decreasing CBO amplitude over the entirety of Run-3. The different CBO behavior also justifies dividing Run-3 into two datasets. The vertical betatron amplitude A_y varies slightly across Run-3 as well (y parameters are fixed to 0 for several Run-2 datasets with limited statistics). Although the improved kick reduced the CBO amplitude, it slightly increased the y amplitude in Run-3NO relative to preceding datasets, but it was still in line with A_y in Run-3 as a whole.

Preliminary blinded crosschecks can be performed with the field measurements. For each production dataset, R is plotted vs. B, where B is the blinded field value in ppm (Figure 5.45). No uncertainty is associated with the B value in this plot. The parameter R is defined with respect to a reference frequency, as given in Equation 5.33. The blind offset ΔR is small (10s of ppb) so is taken to be 0 when calculating the expected slope of the line. Similarly, B is defined with respect to a reference frequency as follows:

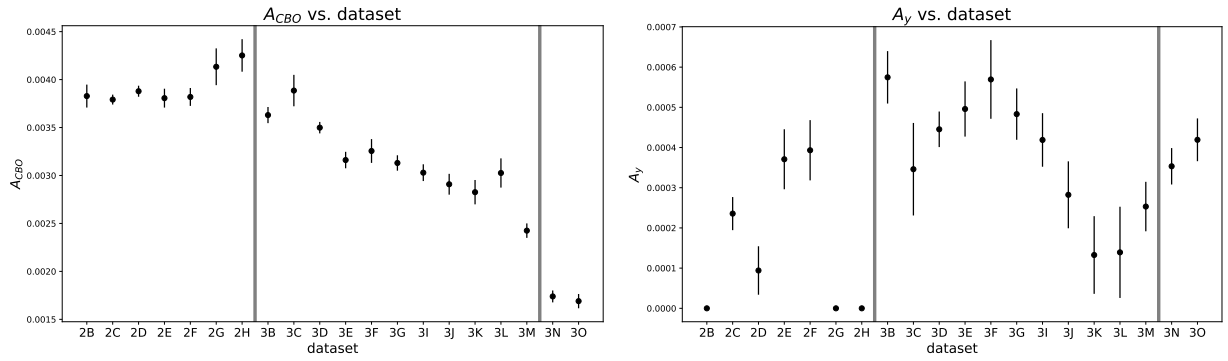


Figure 5.44: Left: A_{CBO} vs. sub-dataset showing a gradually decreasing A_{CBO} during Run-3 as the kicker was improved, resulting in the smallest CBO amplitude in Run-3NO. Right: A_y vs. sub-dataset, showing a slight increase in Run-3NO compared to previous sub-datasets. The parameter A_y was fixed to 0 in certain Run-2 datasets because of lack of statistics.

$$\omega_p(B)\ [\text{MHz}] = 61.79\ \text{MHz} \times 10^{-6}\ B + 61.74\ \text{MHz}. \tag{5.66}$$

Unblinded, $\omega_a/\tilde{\omega}_p \approx 0.0037$ [2]. With the definitions of the blinded R and B, the slope becomes

$$R/B \approx 0.0037\ \frac{61.79\ \text{MHz} \times 10^{-6}}{-2\pi \times 0.2291\ \text{MHz} \times 10^{-6}} \tag{5.67}$$

$$\approx 0.16. \tag{5.68}$$

The linear fit to the data in Figure 5.45 is required to have this slope. The data do not contain much information about the slope as the values of B are very similar across each of the runs and the uncertainty is dominated by high ω_a uncertainty from small statistics datasets. Points may also be affected by small differences in ω_a corrections such as the electric field correction. However, an acceptable p-value of 0.20 and 0.33 is achieved for Run-2 and Run-3, respectively. Run-2 and Run-3 cannot be compared at this time because they have a relative hardware blinding.

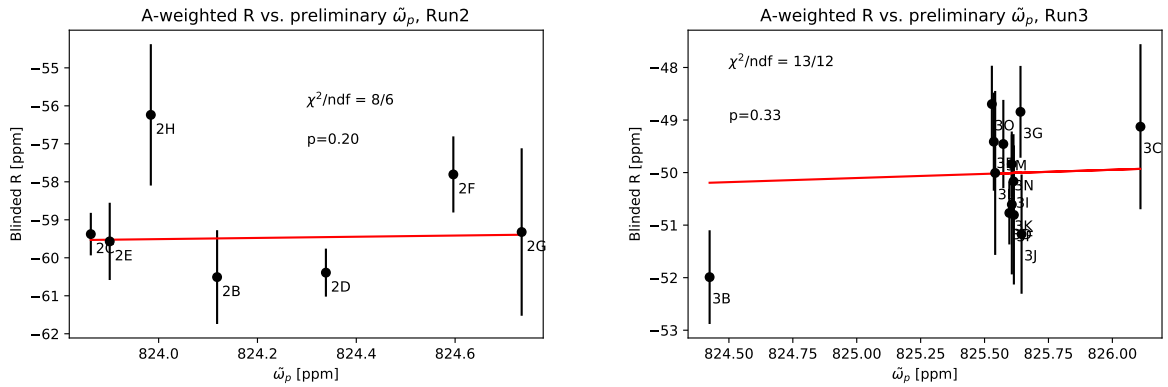


Figure 5.45: Blinded R values vs. preliminary $\tilde{\omega}_p$ for Run-2 and Run-3. The slope has been fixed to -0.16, as described in the text. No uncertainties are given to the values of B.

Chapter 6

Sources of systematic uncertainty in the precession frequency analysis

This chapter discusses the sources of systematic uncertainty associated with extracting ω_a from a fit to the precession frequency histogram. Chapter 3 discussed the components of the ω_a and $\tilde{\omega}_p$ analyses. Although all the corrections in the numerator of Equation 3.1 are considered ω_a systematic errors, this section discusses systematic uncertainties associated with the ω_a fitting procedure itself, as described in the previous chapter. These effects are generally granted an uncertainty rather than a correction. Therefore, there is both an uncertainty on the lost muon correction ($\sigma_{C_{ml}}$, discussed in subsection 3.2.5), and an an uncertainty associated with the lost muon fitting procedure in the precession frequency histogram (discussed in this chapter). Systematic uncertainties on ω_a are related to reconstruction issues and detector effects, lack of knowledge about beam behaviors, and the randomization procedure used to remove the cyclotron frequency from the data. Each source of systematic uncertainty will be estimated, although some uncertainty values are still under study. Ultimately, the Run-2/3 analysis is still dominated by the statistical uncertainty.

6.1 Run-1 ω_a systematic uncertainties

The full uncertainty table for the Run-1 ω_a analysis is shown in Table 6.1 for the four Run-1 datasets; the results are published in [3]. In Run-1, the ω_a uncertainty was highly statistics-dominated, with a 434 ppb

Uncertainty [ppb]	Run-1a	Run-1b	Run-1c	Run-1d
Statistical	1207	1022	823	675
Gain	12	9	9	5
Pileup	39	42	35	31
CBO	42	49	32	35
Randomization	15	12	9	7
Slow term	21	21	22	10
Total systematic	64	70	54	49
Total uncertainty	1209	1025	825	676

Table 6.1: Run-1 ω_a systematic uncertainties, given in ppb, for the four Run-1 datasets. The total uncertainty was highly statistics-dominated. Reproduced from [3].

statistical uncertainty compared to a 56 ppb systematic uncertainty. The contributions to the systematic uncertainty for Run-1 uncertainties form the framework for discussing systematic uncertainty in Run-2 and beyond. All of these systematic uncertainties are still relevant for Run-2/3 and will be discussed in order below, although some procedures for evaluating them have changed. For example, the new pileup correction requires a new approach to the pileup uncertainty. Additionally, the unique behavior of the muon losses in Run-3 compared to Run-1 and Run-2 will be discussed and granted an uncertainty. The Run-2/3 systematic uncertainties are given at the end of the chapter in Table 6.9 for comparison.

The official Run-1 uncertainties come from different analysis procedures, some of which had smaller systematic uncertainties than A. Fienberg's analysis for Run-1 [63], from which this analysis inherits many of its fitting procedures. For example, the "slow term" uncertainty is taken from an analysis that used the Recon East reconstruction algorithm, which has a smaller slow term than the Recon West reconstruction algorithm used by this analysis, as mentioned in the previous chapter. Therefore, the Run-2/3 uncertainties will also be discussed with an eye towards the Run-1 UW analysis. In Run-2/3, the combined statistical uncertainty is expected to be 200 ppb, which will still be much larger than the combined systematic uncertainty. The TDR goal for the ω_a systematic uncertainty, including uncertainties on corrections, is 70 ppb [7], so systematic uncertainties of that order should be reduced. Systematic uncertainties below 10 ppb are considered irrelevant.

6.2 Gain

As discussed in section 4.6, three types of gain corrections are applied to the data in order to maintain consistent detector gain across the measurement period. The out-of-fill (OOF) gain correction is not relevant for the ω_a systematic uncertainty calculation, because it is not associated with an early-to-late effect. However, the in-fill gain (IFG) and short-term double pulse (STDP) gain corrections are relevant, with the IFG having the largest effect. If these gain drops are not fully corrected, the energy scale will change during the fill, resulting in different positrons being included in the histogram at the beginning of the fill compared to the end. To assess the systematic uncertainty due to gain, a histogram is generated without the IFG correction. Then, R can be scanned as a function of IFG multiplier, where a multiplier of 0 means that no IFG correction has been applied and a multiplier of 1 indicates that the nominal IFG correction has been applied. This scan implicitly assumes the worst-case scenario that all channels have a coherent error on the multiplier. A sample result for the Run-2 A-method can be seen in Figure 6.1. The sensitivity of R to the IFG multiplier is 39 ppb per IFG multiplier. The IFG amplitude is generated from fits to in-fill laser pulses during data taking, and its uncertainty comes from fitting uncertainties. If an uncertainty of 10% is allowed on the multiplier, in line with the Run-1 values, then the uncertainty due to the IFG amplitude is 4 ppb. There is also an uncertainty on the IFG lifetime, which is highly correlated with the amplitude in the fits. This issue is being investigated by other groups, but the lifetime uncertainty is expected to be small and comparable to the amplitude uncertainty. Scans of the IFG multiplier for Run-3BM and Run-3NO have shown uncertainties of 4 and 12 ppb, respectively, as summarized in Table 6.2. These values are in line with Run-1 uncertainties and generally small overall.

The effect of the short-term double pulse correction was about 1 ppb in Run-1, so it is not re-evaluated here. However, with the new clustering algorithm that decreases the cluster dead time to < 1 ns, very short time scale double pulse effects may be relevant. Simulations are underway to examine this effect, but they have not been finished at the time of this thesis. They are expected to be significantly less than 10 ppb, so no uncertainty is assigned to STDP-related effects.

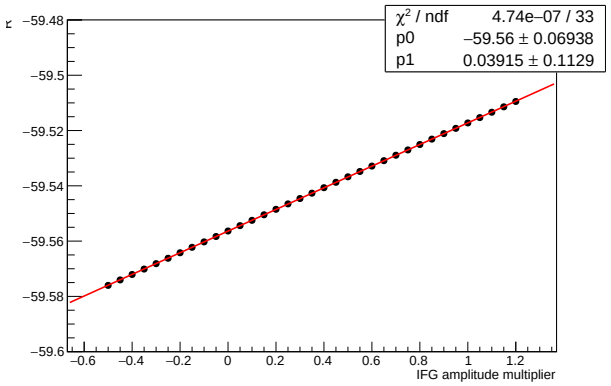


Figure 6.1: Scan of R as a function of in-fill gain (IFG) multiplier for the A-Weighted method in Run-2. A linear fit has been applied. The slope is 39 ppb per IFG multiplier. For clarity, the error bars are not shown.

Uncertainty	Run-2	Run-3BM	Run-3NO
In-fill gain	4	4	12

Table 6.2: Uncertainty in ppb due to the in-fill gain correction (IFG) based on a scan of R as a function of IFG multiplier and assuming a 10% uncertainty on the amplitude.

6.3 Pileup

In Run-1, the UW analysis used the energy density pileup correction method described in subsection 5.2.2. This correction required a scaling factor in order to find the correct amplitude for the pileup spectrum. In order to assess the systematic uncertainty due to the pileup amplitude, fits were performed with different pileup multipliers to extract the sensitivity of R to pileup. The fit χ^2 was also plotted as a function of the pileup multiplier. The uncertainty on the pileup amplitude was taken as the allowed 1σ difference in χ^2 from the pileup amplitude with the minimum value of χ^2. The uncertainty due to the pileup amplitude was about 15 ppb [63].

A larger uncertainty arose from the pileup phase, or the shift in time of the pileup spectrum. A similar scan was performed to find the sensitivity of R as a function of pileup phase, but the χ^2 was not very sensitive to the phase. Therefore, the cluster dead time of 2.5 ns was used to calculate an uncertainty on the pileup time of 1.4 ns. Based on this calculation, the pileup phase was assigned an uncertainty of 38 ppb in the A-method, resulting in a relatively large combined uncertainty of 51 ppb.

One of the goals of the UW analysis for Run-2/3 was to decrease the pileup uncertainty, which should be

less than 40 ppb for the Fermilab TDR goal [7]. The new $\Delta t'$ clustering, which is discussed in section 4.7, was one improvement aimed at decreasing this uncertainty. From the framework of the energy density pileup correction, the $\Delta t'$ clustering would decrease the systematic uncertainty in two ways: it reduced the dead time from 2.5 ns to about 0.75 ns, which would decreases the phase uncertainty calculated above. Secondly, since pileup was about 4 times lower in the $\Delta t'$ energy spectrum, the sensitivity of the fit to the pileup multiplier or phase should be reduced.

The Run-1 pileup uncertainty analysis is repeated on Run-2 data as an illustrative example. Scans of χ^2 and R as a function of pileup multiplier are shown for the A-weighted fits in Figure 6.2. The slope of the linear fit to R is -0.07 per pileup multiplier. For reference, the sensitivity of R in Run-1a was -0.22 [63], meaning that the sensitivity has decreased by about a factor of 3. Following the Run-1 calculations, the uncertainty due to the pileup amplitude is 3 ppb. For the pileup phase, the smaller dead time and decreased sensitivity reduced the uncertainty to 6 ppb in the A-method. Added linearly, the pileup uncertainty would be 9 ppb, much lower than the 51 ppb uncertainty in Run-1.

However, because the lower dead time induced a time dependence to the reconstruction algorithm, the empirical pileup subtraction method was used for this analysis. The empirical pileup spectrum is determined through combinatorics on the data itself, as described in subsection 5.2.2, and therefore there is no inherent scaling factor required to apply the correction. Similarly, because the pileup spectrum is not binned in time and because the time reconstruction of the pileup events is the same as the time reconstruction in the data, there is no inherent pileup phase uncertainty. New sources of uncertainty due to the empirical pileup method are discussed instead.

Although the empirical pileup correction does not have the same type of amplitude and phase uncertainty as the energy density pileup correction, similar amplitude and timing errors may arise. The empirical pileup candidates are generated by linearly adding the traces of the component islands. However, due to detector effects, this linear assumption may not hold for real pileup events. For example, the pixel depletion in the SiPM sensor caused by the first pulse could distort the shape of the second pulse. Although the amplitudes of the pulses are corrected using the short-term double pulse (STDP) correction, effects at very short time scales (less than a nanosecond) or nonlinear effects may not be accounted for. This topic is currently under study in simulation. As a very conservative estimate, the same dead time calculation used for the empirical

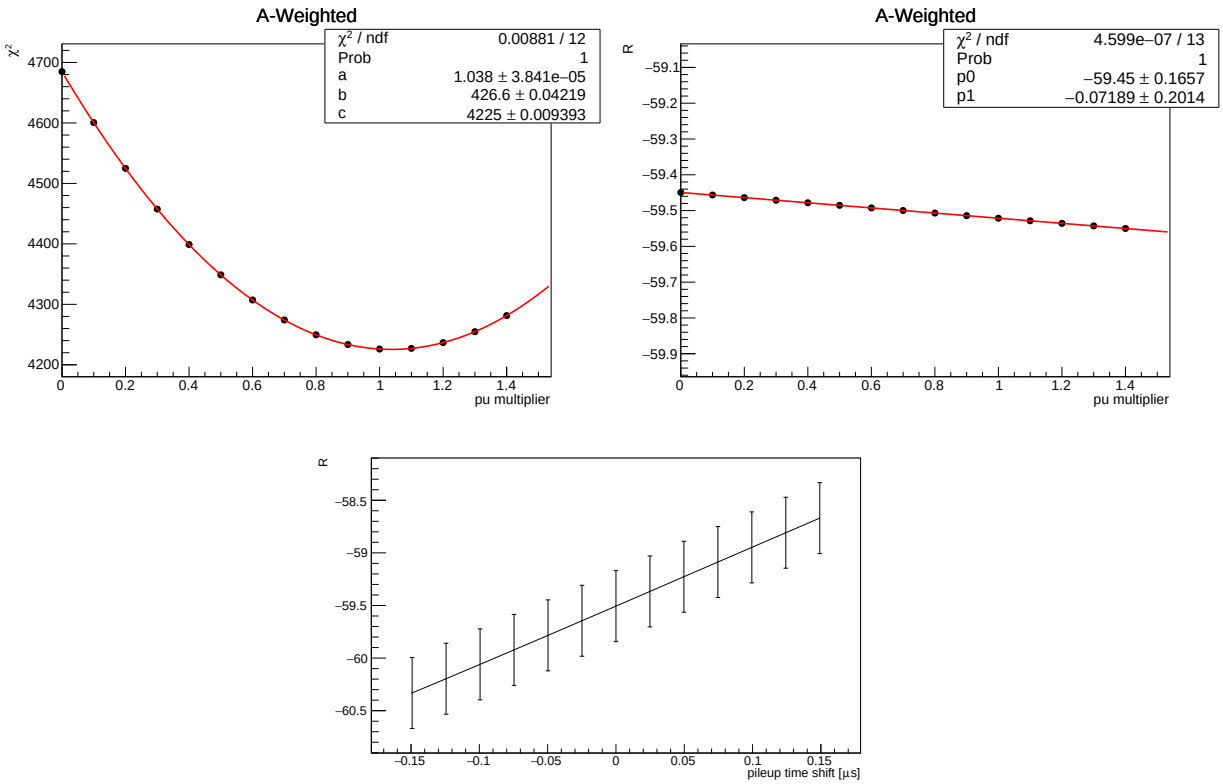


Figure 6.2: Top: χ^2 and R as a function of pileup multiplier for the empirical pileup correction. Plots are shown for the A-method for Run-2. The sensitivity of R to the pileup multiplier is -70 ppb per pileup multiplier, which is about 3 times lower than it was for the UW analysis for Run-1. For clarity, the error bars are not shown. Bottom: pileup phase scan as a function of time in μs.

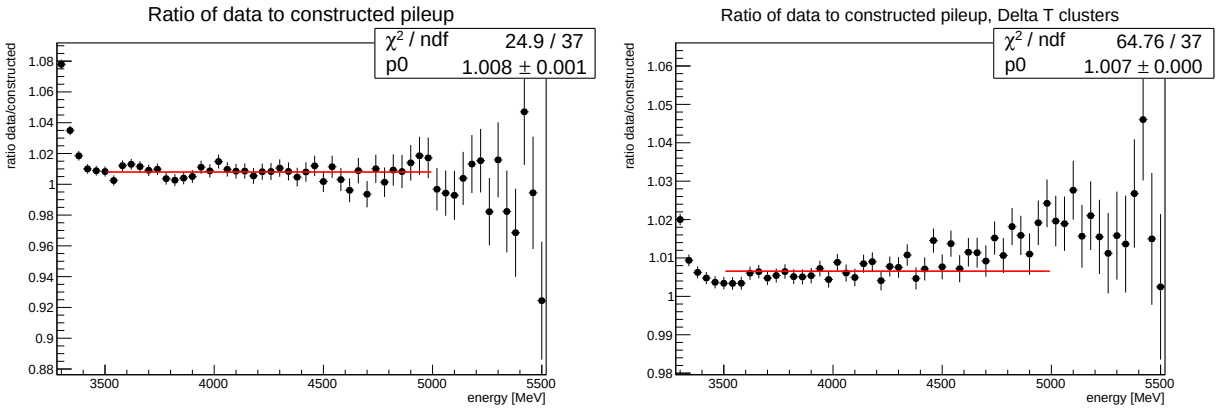


Figure 6.3: Run-2 ratio of the data to the constructed empirical pileup spectrum in the pileup region. The ratio has been fit to a constant function, and the fit result is given. Left: the updated $\Delta t'$ clusters. Right: the old Δt clusters, which contain about 4 times more pileup.

pileup is used to estimate the uncertainty due to the pileup timing. This results in values of around 4-6 ppb, which are expected to decrease with future work.

Secondly, the accuracy of the pileup scaling can be checked by taking the ratio of the uncorrected energy spectrum to the pileup spectrum, as seen for Run-2 in Figure 6.3 (left). A constant fit is applied to the data between 3500 and 5000 MeV in the pileup region. The value of this fit is 1.008 ± 0.001, meaning that the pileup correction is incorrect by 0.8%. Multiplying this by the sensitivity of R to the pileup multiplier gives less than 1 ppb, indicating that the pileup correction is doing a good job of subtracting the pileup. The ratio for Run-3BM and Run-3NO is also around 0.8%, meaning that the uncertainty is less than 1 ppb for all datasets. For comparison, the ratio is also shown for the Δt clusters, which have about 4 times as much pileup, in Figure 6.3 (right). Although the data are less linear, the factor of 1.008 remains the same, suggesting it is a small error inherent in the pileup correction. An uncertainty of 1 ppb is assigned to the pileup amplitude for all datasets.

Another source of systematic uncertainty in the empirical pileup correction is due to the rate error incurred by assuming that the trigger and shadow window have the same positron rate $r(t)$. As discussed in subsection 5.2.2, the cyclotron period $T_C = 149.2$ ns is chosen as the gap time between the trigger and shadow windows. This time allows the rate to be similar between the windows without the trigger window perturbing the shadow window. However, the positron rate in the shadow window is not exactly equal to the rate in the trigger window, because of the ω_a and other oscillations. The rate error distort the shape of the

Source of uncertainty	Run-2	Run-3BM	Run-3NO
Pileup amplitude	1	1	1
Pileup time reconstruction	6	4	5
Rate error	5	5	5
Total	12	10	11

Table 6.3: Estimates of the pileup uncertainty in ppb for the three datasets. Contributions from different sources of pileup uncertainty are added linearly.

pileup spectrum, potentially causing a greater error than simply scaling or shifting the pileup correction in time. Early simulations of this error have suggested that the effect may be less than 5 ppb, so preliminarily the uncertainty assigned to this effect is 5 ppb.

Estimates of the uncertainty due to the pileup are summarized in Table 6.3. The uncertainties are assumed to be correlated and added linearly, resulting in total pileup uncertainties of 10-12 ppb. This is significantly below the $\approx$ 50 ppb pileup uncertainty for the UW analysis in Run-1, and also below the 40 ppb TDR goal for the Fermilab experiment. Further work is underway on the pileup uncertainty, but it is expected to be lower than other sources of uncertainty in Run-2/3.

6.4 Beam motion

6.4.1 CBO envelope

The CBO oscillation is more strongly correlated to R than any of the other beam dynamics oscillations, as shown in Figure 5.23. Further, it persists late in the fill, with a lifetime of around 250 μs, meaning it will affect a large portion of the measurement period. Therefore it is important to determine the sensitivity of R to the modeling of the CBO. The analytic form of the CBO envelope, which is nominally $E_{CBO}(t) = e^{-t/\tau_{CBO}}$, is not specifically supported by tracker or calorimeter data, although a good fit quality is achieved for the combined fit when using it. There is evidence that $E_{CBO}(t)$ may not have the right form for the per-calo fits, where CBO effects are exaggerated. Residual peaks remain in the FFT at the CBO frequency when performing fits separately for each calorimeter, as discussed in section 5.5. The peaks were accounted for in that section by a combination of changing the envelope and including time changing CBO parameters. There is also further evidence that the fitting issues may be due to the wrong functional form of the envelope.

Figure 6.4 shows the CBO amplitude A_{CBO} as a function of time. It is extracted by performing a fit over narrow time windows of $4T_C$, where T_C is the ω_a period, with the CBO lifetime τ_{CBO} fixed to be very long. The red line shows the nominal form of the CBO envelope with parameters extracted from a fit over all times. The figures on the right show the residual from the red line. Calorimeter 3 and 18 are used as examples of bad and good per-calorimeter CBO fits, respectively; calorimeter 3 had a large residual peak in the FFT at the CBO frequency as seen in Figure 5.25. Compared to calorimeter 18, calorimeter 3 has a larger residual from the form of the CBO envelope, especially between 50 and 150 μs. The residual is correlated with the size of the remaining CBO peak across all calorimeters. This gives more credence to the possibility of CBO envelope mis-fitting.

Therefore, alternative forms of the envelope were tested to determine the sensitivity of R to the choice of CBO envelope. These alternate envelopes were applied to the full fit. In addition to the nominal exponential envelope, three other analytical forms were used:

- $E_{CBO}(t) = e^{-(t-\mu)^2/\sigma^2} + C$: Gaussian envelope with constant offset, with $\mu = 30\mu s$
- $E_{CBO}(t) = e^{-t/\tau_{CBO}} + C$: Exponential with constant offset
- $E_{CBO}(t) = \frac{1}{1+\alpha t^2} + C$

The Gaussian envelope was found to not result in good fit quality, and so was discarded. The other two choices of envelope have similar fit quality to the nominal envelope, and therefore were used to evaluate the sensitivity of R to the choice of envelope. The amplitude $A_{CBO}(t)$ determined from a fit to Run-3BM A-weighted data is shown in Figure 6.5 for the three choices of envelope. The behavior matches closely until 300 μs, although the third envelope turns over at early times, a characteristic which has been predicted by preliminary beam simulations. At late times, the functions diverge, with the nominal envelope decaying away to 0, as opposed to the other two envelopes, which have constant offsets so that the CBO never fully decoheres. The p-value and ΔR from the nominal R value are shown for each dataset for each envelope in Table 6.4. The third envelope generally has a slightly worse p-value, with the exception of Run-3NO, which is expected to have somewhat different CBO behavior due to the increased magnetic kick.

These tests show that R is sensitive to the choice of CBO envelope, with variations in R of up to 50 ppb depending on the choice of envelope. At this time, the uncertainty due to the envelope is taken

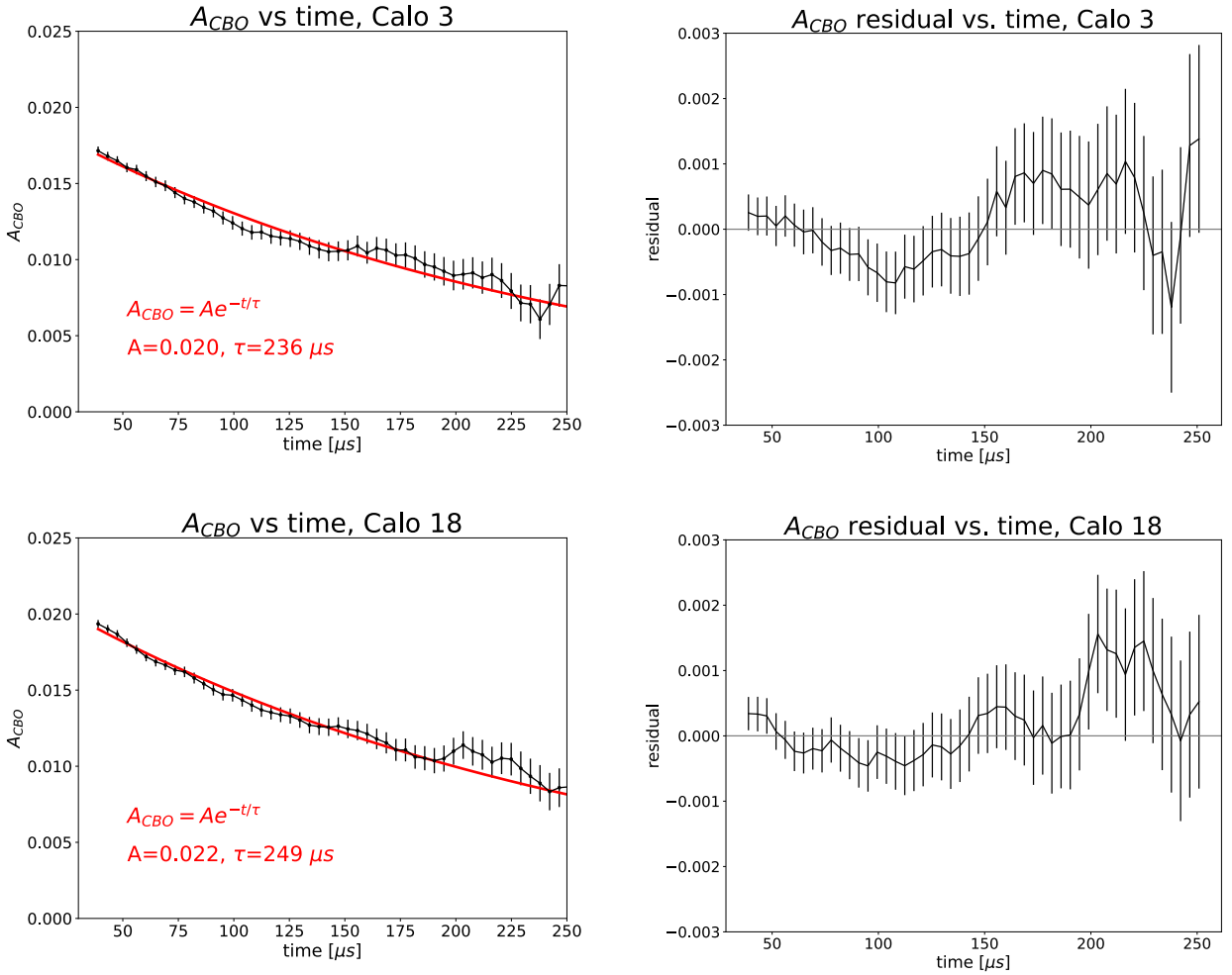


Figure 6.4: The CBO amplitude A_{CBO} vs. time, determined by performing fits over time windows of 4 times the ω_a period. The CBO lifetime τ_{CBO} has been fixed to be large so that A_{CBO} can be extracted at each time. The red line shown is the CBO envelope extracted from the full fit, with the parameters shown. The figures on the right shows the residuals with respect to the red line. For calorimeter 3, which has a larger CBO peak after the full fit function is applied, the envelope is a worse fit, especially between 50 and 150 μs. Calorimeter 18, which has no visible CBO peak in the full fit FFT, is shown for comparison.

Envelope		Run-2	Run-3BM	Run-3NO
$E_{CBO}(t) = e^{-t/\tau_{CBO}}$	p-value	0.15	0.27	0.40
	ΔR [ppb]	0	0	0
$E_{CBO}(t) = e^{-t/\tau_{CBO}} + C$	p-value	0.15	0.27	0.45
	ΔR [ppb]	-6	-10	-38
$E_{CBO}(t) = \frac{1}{1+\alpha t^2} + C$	p-value	0.11	0.24	0.45
	ΔR [ppb]	45	27	-27

Table 6.4: Variation in p-value and R based on the choice of analytical form of the CBO envelope E_{CBO}. The exponentially decaying envelope is taken to be the nominal choice.

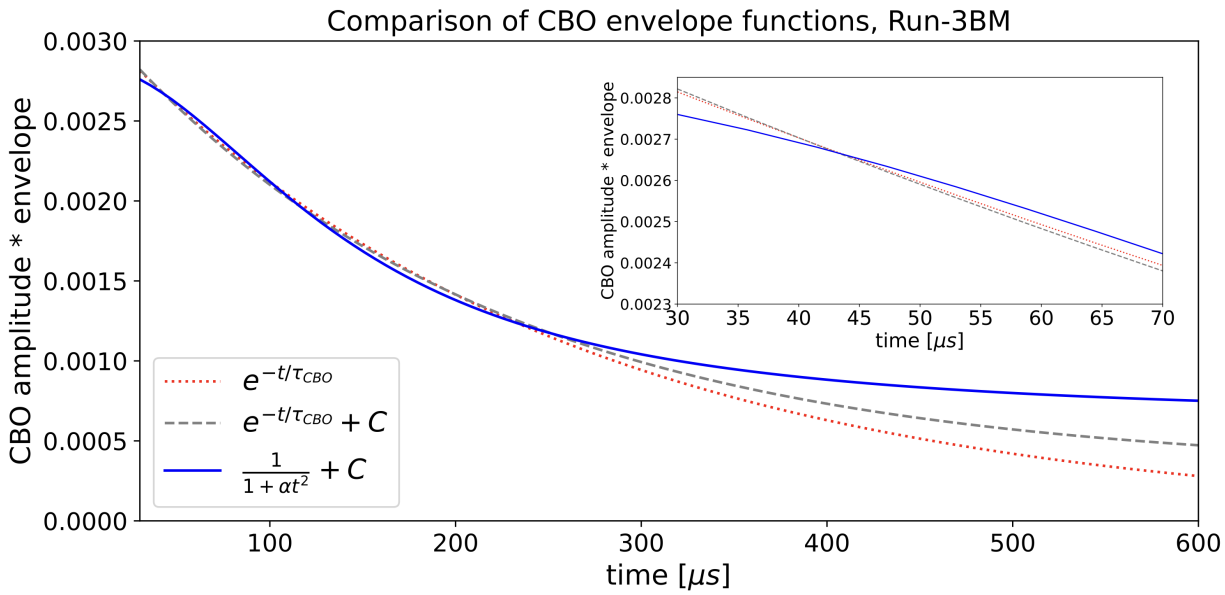


Figure 6.5: Comparison of CBO envelope functional forms from Run-3BM. The vertical axis is the CBO amplitude from the fit multiplied by the functional form of the envelope. Inset: The envelopes at early times, showing the different behavior of the third envelope in blue.

Uncertainty	Run-2	Run-3BM	Run-3NO
CBO envelope	51	37	38

Table 6.5: Table of uncertainties in ppb due to beam motion in ppb.

to be the difference in the minimum and maximum R values measured with different envelopes, as seen in Table 6.5. Simulations are currently in progress to determine the ideal form of E_{CBO}, which may reduce this uncertainty. Just as in Run-1, the CBO uncertainty is the largest contribution to the systematics uncertainty table in Run-2/3.

6.4.2 1.9 MHz peak

Although the new peak at 1.9 MHz, as discussed in section 5.3.1, is clearly present in the data and is justified by simulation, the R value was compared with and without the additional fit parameters A_{VW-CBO} and ϕ_{VW-CBO}. In the A-Method, ΔR was 17 ppb for Run-2, 23 ppb for Run-3BM, and 46 ppb for Run-3NO. For Run-2, the χ^2/ndf improved from 4271 / 4132 (p=0.06) to 4226 / 4130 (p=0.14). For Run-3BM, it improved from 4336 / 4131 (p=0.01) to 4183 / 4129 (p=0.27). For Run-3NO, it improved from 4233 /

Uncertainty	Run-2	Run-3BM	Run-3NO
Randomization	3	3	6

Table 6.6: Table of uncertainties in ppb due to running over an ensemble of 100 random seeds following randomization to remove cyclotron oscillations.

4134 (p=0.14) to 4152/4132 (p=0.41). Because of large size of the peak, the clear improvement in χ^2 for all datasets, and the physics- and simulation-motivated parameterization of the VW-CBO cross terms, no systematic uncertainty is assigned to the inclusion in the fit of these two additional terms.

6.5 Randomization

When the precession frequency histogram is filled, the cluster times are randomized by a number in $[-T_C/2, T_C/2]$, as described in subsection 5.2.1. Fits are performed over 100 random seeds, and the final R value is extracted from a Gaussian fit to the histogram of 100 R values, as shown in Figure 5.22. A systematic uncertainty is given to the uncertainty on the mean. These uncertainties are shown in Table 6.6. The randomization uncertainty of 6 ppb or less justifies not running additional random seeds.

6.6 Slow term

Since Run-1, an additional gain-like "slow term" has been evident, particularly in Recon West-based analyses like the one in this thesis. The scale of the effect, and the approach to its correction, can be studied by plotting the muon loss scaling factor K_{loss} as a function of energy. K_{loss} absorbs the reconstruction-induced slow term and decreases as a function of energy, a non-physical result. A correction is applied to the clusters in order to remove the dependence of K_{loss} on energy, as seen in Figure 5.30.

The effect is significantly lower in Recon East (RE) vs. Recon West (RW), as seen in Figure 5.31. This difference was leveraged by S. Foster to understand the source of the larger effect in RW: due to different primary and secondary energy thresholds, small pulses were less likely to be fit if another pulse was adjacent. This is contrasted with RE, where a 3×3 square around the large pulse is always fit. This knowledge justifies the Run-1 approach to the residual gain correction, where a gain-like correction was applied at the muon lifetime; the rate-dependent behavior of the slow term is similar to a gain sag. However, part of this effect

Uncertainty	Run-2	Run-3BM	Run-3NO
Residual gain correction parameters	8	5	12

Table 6.7: Table of uncertainties in ppb due to the slow term, which is corrected by the residual gain correction.

remains unexplained. Even when equalizing the energy thresholds used for reconstruction, there is still a slow term. Additionally, there is a small slow term in Recon East.

In order to determine the uncertainty on the choice of residual gain correction parameters, χ^2 and R were scanned as a function of the residual gain amplitude and residual gain asymmetry. T-method scans are shown for Run-2 and Run-3BM in Figure 6.6. One, two, and three σ bands in χ^2 (that is, χ^2 +1, +4, and +9) are marked on both plots in black, gray, and light gray respectively. The nominal value of the residual gain correction based on Run-1 values and scanning the amplitude to flatten K_{loss} vs. energy is shown as a red cross. The plot on the right shows ΔR with respect to the nominal R value, showing the variation of R across the χ^2 well as a function of the correction parameters. The range of R values based on a choice of χ^2 variation is shown in the legend. The 2σ well is taken to be the uncertainty on R due to the choice of residual gain parameters.

Uncertainty on the residual gain correction parameters may be an underestimate of the total uncertainty due to the slow term. Because the source of the slow term is only partially understood, more subtle effects may exist which are not being properly corrected by the application of a gain-like correction. However, for this thesis, the uncertainty due to the choice of correction parameters is the only uncertainty assigned to the slow term. These uncertainties are summarized in Table 6.7.

6.7 Lost muons

6.7.1 τ-K_{loss} correlation

The muon loss scaling factor K_{loss} is described in subsection 5.3.2. Because of the definition of K_{loss}, it must be positive, as a negative K_{loss} would indicate that the number of stored muons is increasing over the fill. Therefore, the fact that K_{loss} is negative in Run-3BM and Run-3NO even after the residual gain correction is a nonphysical result, and a systematic uncertainty must be assigned to it. The lost muon

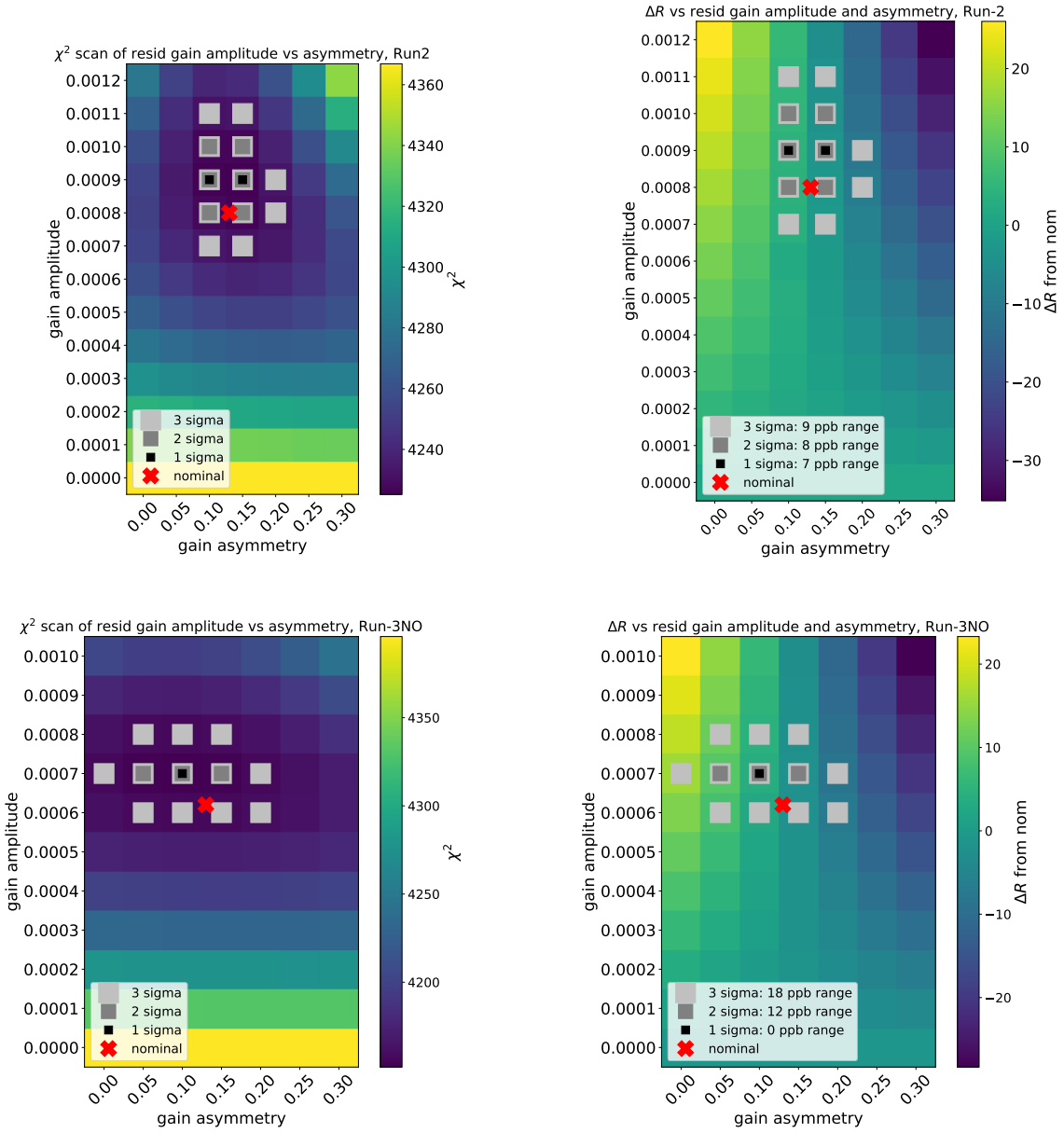


Figure 6.6: A-method χ^2 (left) and ΔR (right) scans as a function of the residual gain amplitude and asymmetry for Run-2 (top) and Run-3NO (bottom). The scan of χ^2 shows a good match with the nominal parameters, marked with a red x. The gray squares mark the 1, 2, and 3σ allowed variation based on the χ^2. The figures on the right show ΔR compared to the nominal R value. The legend shows the variation in R over the different χ^2 bands.

spectrum in Run-3 was exponential at the muon lifetime, meaning that the fractional loss rate was consistent during the fill. As a result, K_{loss} is largely inseparable from τ in the fit. This correlation is explored by plotting χ^2 and R as a function of τ and K_{loss}, where most other parameters are allowed to float (Figure 6.7). The central value of the plot is determined by the default K_{loss} and τ parameters for the full fit. The dashed lines indicate the upper and lower bounds of the values of τ calculated from the fast rotation distribution, as determined by K. Labe. These dashed lines should suggest physically reasonable values of τ. The gray and black squares indicate the 1, 2, and 3σ χ^2 bands.

For the Run2 analysis, both τ and K_{loss} are well-determined, and K_{loss} has a physical (positive) value. Therefore, the uncertainty resulting from the K_{loss}-τ correlation is minimal. Additionally, the fit converges on a physical value of τ, as determined from the fast rotation distribution.

However, in the Run3 analysis, K_{loss} and τ are highly correlated. Both parameters have a much wider allowed range, and K_{loss} generally converges on a negative value. The negative K_{loss} value may be related to the Recon West slow term, which was discussed in the previous section; Figure 5.30 shows that K_{loss} is increased in Run-3, including at low energies, when applying the residual gain correction. If the slow term is not being perfectly corrected, K_{loss} may still be absorbing some effects of the slow term. Further, the allowed χ^2 range is much wider, corresponding to a wide range of possible τ and K_{loss} values. If the 2σ χ^2 band is used to quantify the allowed R range, the uncertainty due to this correlation is 0 ppb in Run-2, 11 ppb in Run-3BM, and 9 ppb in Run3-NO.

Alternatively, K_{loss} can be viewed as an unnecessary parameter in Run-3NO, since τ and K_{loss} are both describing the same behavior. By setting K_{loss} to 0 in the fit and examining the change to χ^2 and R, the necessity of the parameter can be studied. Figure 5.22 shows the distribution of R values with and without K_{loss} fixed to 0. In this case, there is only a change of 3 ppb, which is similar in the other two datasets. Because these shifts in R with and without K_{loss} set to 0 are manifestations of the same effect described in the τ-K_{loss} scans, the uncertainty can conservatively be taken to be 0 ppb in Run-2, 11 ppb in Run-3BM, and 9 ppb in Run3-NO. K_{loss} can comfortably be set to 0 in fits to avoid over-fitting, as long as τ is not also restricted. Notably, τ will absorb the lost muon effects, shifting the value of τ. For example, in Run3-BM, τ shifts from 64.416 μs to 64.422 μs when K_{loss} is fixed to 0. The uncertainties due to the muon losses in the fit are summarized in Table 6.8.

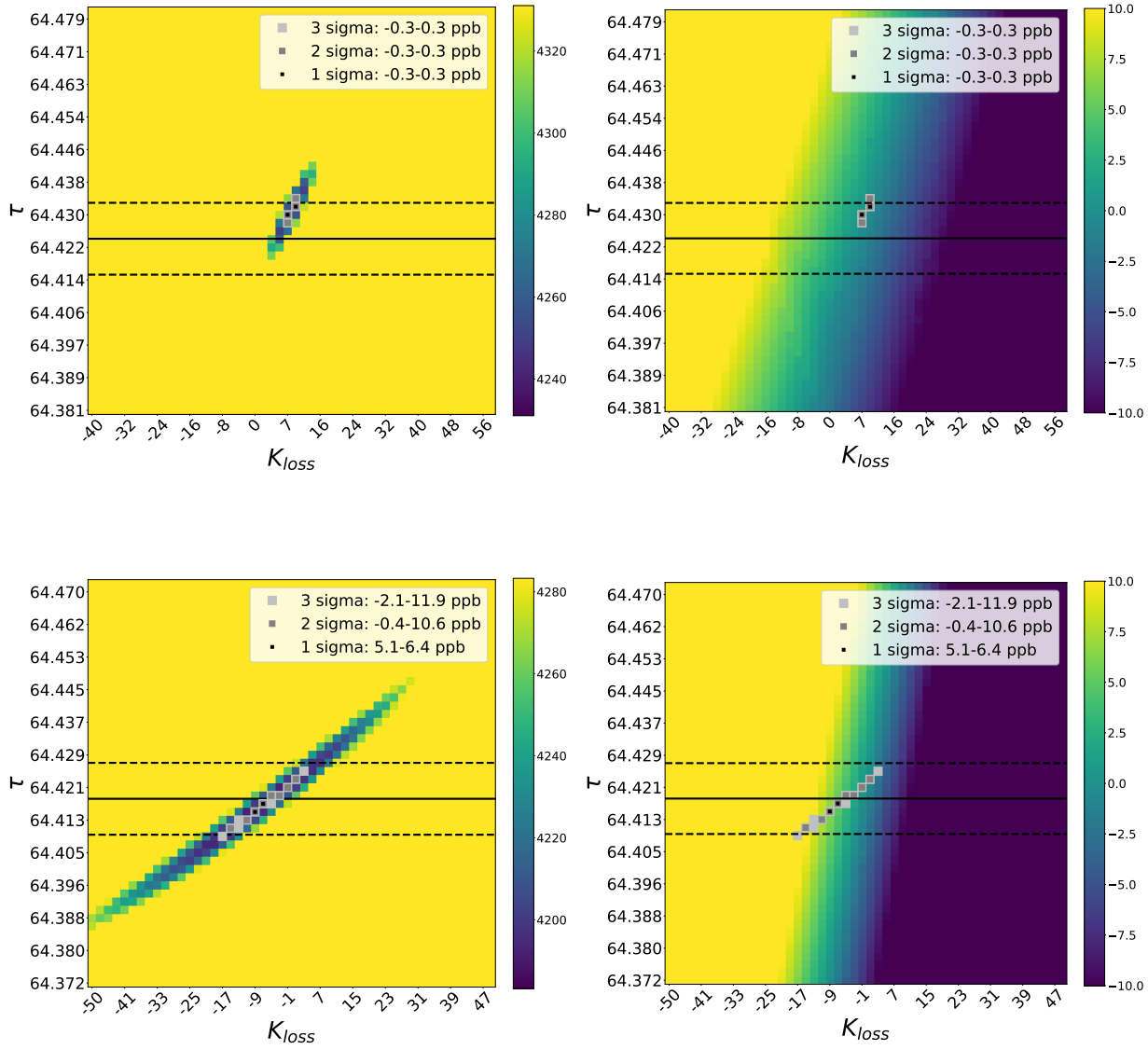


Figure 6.7: Scans of χ^2 and R as a function of τ and K_{loss} for Run2 (top) and Run3BM (bottom). The parameters τ and K_{loss} are much better determined in Run2; in Run3, K_{loss} converges on a nonphysical negative value and the χ^2 well is much wider. The solid black line indicates the value of τ determined from the fast rotation distribution, as determined by K. Labe. The dashed lines indicate the uncertainty on that determination. The legend shows the variation of R over the allowed χ^2 regions.

Uncertainty	Run-2	Run-3BM	Run-3NO
τ-K_{loss} correlation	0	11	9

Table 6.8: Table of uncertainties in ppb due to muon losses. No uncertainty is attributed to the choice of muon loss cuts; the uncertainty comes purely from the correlation between τ and K_{loss} in the Run-3 fits.

6.7.2 Other muon loss systematic uncertainties

Other muon loss systematic uncertainties may arise from incorrect modeling of the loss spectrum. In Run-1, the loss spectrum was generated with different time and energy cuts or by using quadruple coincidences instead of triples, for example. All of these uncertainties were on the order of a few ppb, and no uncertainty was quoted in the final Run-1 publication. Given that K_{loss} has been removed as a parameter in Run-3BM and Run-3NO, the modeling of the loss spectrum is irrelevant. Additionally, the Run-2 losses were still an order of magnitude lower than then Run-1 losses. Therefore, no additional uncertainties are granted to other muon loss effects.

6.8 Comparisons of old and new reconstructions

Both the positron reconstruction and the pileup correction were altered from the Run-1 analysis described in [63]. However, histograms using the original reconstruction and pileup correction were produced alongside the updated version so that a comparison between the results could be made. Since the original reconstruction was extensively validated for the Run-1 publication, a comparison provides additional evidence for the stability of the new reconstruction. Figure 6.8 compares the Δt (old) reconstruction with the $\Delta t'$ (new) reconstruction for the A and T-weighted analyses for the three datasets. Fits for the new reconstruction were performed over an ensemble of 100 random seeds, but only a single seed was fit for the old reconstruction. Therefore, the error bars on the $\Delta t'$ R values shows the range of the R values from all random seeds. The "nominal" R for the full analysis, which is the $\Delta t'$ A-weighted fit for each dataset, has been subtracted off from all R values. Generally, the Δt fits are consistent with the $\Delta t'$ R values, sometimes having a lower and sometimes a higher R value than the central $\Delta t'$ number. Overall, this comparison provides a good cross check with the analysis procedure published in Run-1.

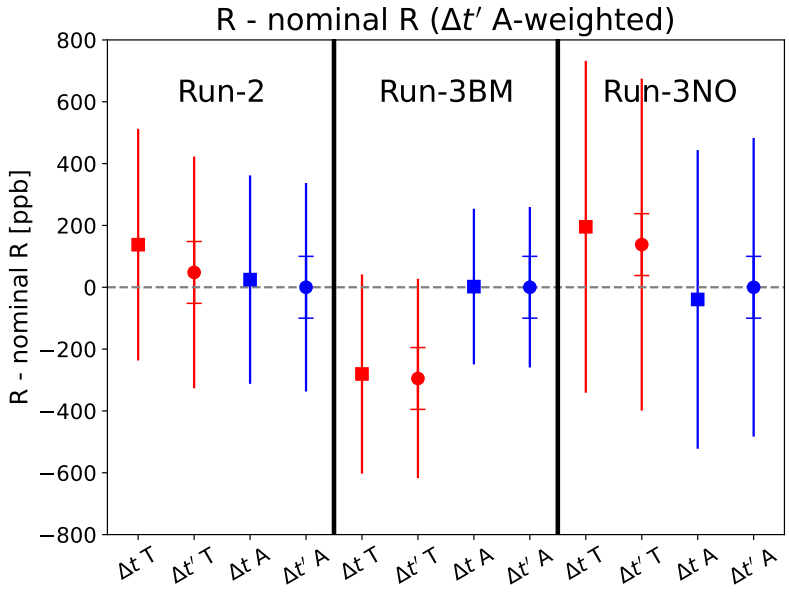


Figure 6.8: Values of R for the old analysis (Δt reconstruction and energy weighted pileup method) and new analysis ($\Delta t'$ reconstruction and empirical pileup method) for T- and A-method fits for the 3 larger datasets. The "nominal" R for this analysis, the A-weighted R value for the new analysis, is subtracted off for comparisons between datasets. For the new analysis, the central value is shown with the error bars indicating the range of random seed fit values. For the old analysis, only one random seed is shown.

Uncertainty	Run-2	Run-3BM	Run-3NO	Combination
Statistical	337	290	487	200
Gain	4	4	12	5
Pileup	12*	10*	11*	11*
CBO	51*	37*	38*	42*
Randomization	3	3	6	4
Slow term	8	5	12	7
Lost muons	0	11	9	7
Total systematic	53*	40*	44*	45*
Total uncertainty	341*	293*	489*	205*

Table 6.9: Full ω_a uncertainty table for Run-2/3 given in ppb. * indicates an estimate.

6.9 Conclusions

Table 6.9 summarizes all of the precession frequency uncertainties for the three Run-2/3 datasets. An asterisk indicates an estimate, subject to change upon further study. The various sources of uncertainty are treated as uncorrelated and combined in quadrature. All datasets are statistics-dominated, with the largest source of systematic uncertainty coming from the CBO envelope modeling. The slow term and pileup uncertainties are significantly reduced relative to the UW analysis values for Run-1. The greatest possibility for improvement relates to the modeling of the CBO envelope. If beam dynamics models and future analysis can motivate a specific form of the CBO envelope, this uncertainty should be reduced. The reduction in the CBO oscillation due to the application of radio frequency (RF) pulses may reduce this uncertainty for parts of Run-4 and Run-5. Overall, the Run-2/3 analysis is still statistics dominated, with the total systematic uncertainty being about 45 ppb compared to the statistical uncertainty of 200 ppb.

Chapter 7

Conclusion

This thesis presents a blinded precession frequency analysis of Run-2 and Run-3 of the Fermilab Muon $g-2$ experiment. The Run-1 publication confirmed the result of the Brookhaven Muon $g-2$ experiment, and the averaged results stand in tension with the official Standard Model calculation to 4.2σ. The Run-2/3 precession frequency analysis will have a combined statistical precision of $\approx$200 ppb, which is more than a factor of two improvement over the statistical precision in Run-1. Further, the Run-2/3 dataset had improved operating conditions, with a lower phase-acceptance correction, better hall cooling in Run-3, and reduced beam oscillations in part of Run-3. The systematic uncertainty in this precession frequency analysis is estimated to be 45 ppb, an improvement over the Run-1 uncertainty.

This analysis leaves several venues for future investigation. The cause of the remaining slow effect in certain reconstruction algorithms was extensively investigated in preparation for the Run-2/3 analysis. An improvement to the reconstruction has already been implemented for the Run-4+ analysis in order to reduce this effect, but the problem will remain to a certain extent. Further improvements to the reconstruction or to the estimate of the systematic uncertainty due to this effect may be good areas for future investigation. The uncertainty due to the CBO decoherence envelope is the largest source of systematic uncertainty in Run-2/3 as of this writing. There are several possible approaches to reducing this systematic uncertainty. The amplitude of the CBO oscillation may be exaggerated by fitting calorimeters individually or selecting vertical slices of the calorimeter in order to improve the sensitivity to the envelope shape. An empirical envelope may also be generated from these studies, which could be used in the fit function so that an analytical enve-

lope is not required, similarly to how the muon losses are incorporated into the fit. Additionally, simulation of the beam dynamics may give further insight into the correct form of the envelope. The reduction in CBO amplitude because of the RF system in Run-4 and beyond will hopefully decrease this systematic uncertainty but may make the oscillation more difficult to fit. Lastly, Run-3 already posed difficulties for fitting the muon losses due to a very uniform spectrum with an equal number of muons lost per stored muon across time in the fill. A different approach to lost muons may be required for future runs, such as removing them from the fit entirely or fixing the scaling parameter K_{loss} based on previous runs. However, the latter would require assuming that the efficiency of measuring a lost muon is the same across runs with demonstrably different beam dynamics, so it may not be the ideal option.

However, as in Run-1, other systematic uncertainties, such as those due to the phase-acceptance correction, the phase-momentum correlation in the electric field correction, and the kicker transient, are likely to be larger than the ω_a systematic. Major efforts are also being made to reduce these sources of systematic uncertainty and to ensure that the ω_a systematic remains small. In the next year, this analysis will be unblinded with respect to the other precession frequency analyses, and finally the full result will be unblinded and published.

Given the improvements that have been made in hardware, software, and analysis since the Run-1 result, the experiment expects to meet its final uncertainty goal of 140 ppb for the full experimental run. Future Muon $g-2$ publications may determine whether the tension with the Standard Model reaches the 5σ threshold for a discovery of new physics. Crosschecks between theory groups are also underway in order to resolve tension between data-driven and lattice QCD calculations of the hadronic vacuum polarization. Given the current ambiguous state of the theory calculation with the new lattice results, it is more important than ever to ensure that the experimental result is as precise as possible.

7.1 Future experiments

The Fermilab Muon $g-2$ experiment is scheduled to finish running in 2023. However, future experiments offer new venues to search for beyond the Standard Model physics. The Muon $g-2$ experiment at J-PARC in Japan [79] will not be as precise as the Fermilab experiment, with an estimated statistical uncertainty of 450 ppb. However, it utilizes a different method than the BNL and FNAL experiments, avoiding the need for

a large electric field correction C_e, which was 489 ppb in Run-1, greater than the combined uncertainty on the measurement. The systematic uncertainty for the J-PARC experiment is expected to be less than 70 ppb. At J-PARC, the need for an electric field to provide focusing is eliminated by using a beam with $approx1000$ times lower emittance, which is produced by re-accelerating a thermal muon beam. The beam momentum is therefore not restricted to have the "magic momentum" of 3.094 GeV/c; instead, it will be only 300 MeV/c. The muons are stored in a ring 20 times smaller than the BNL/FNAL ring. This experiment also has the ability to measure the muon's electric dipole moment (EDM).

Although the J-PARC experiment is the only future experiment planned to measure muon $g-2$, there are many other related muon, charged lepton flavor violation, and lepton flavor universality experiments currently proposed. Only a few are described here. The PIONEER experiment [80, 81], which has major contributions from the University of Washington, will take place at the Paul Scherrer Institute (PSI). It will measure rare pion decay branching ratios in order to test for CKM unitarity and lepton flavor universality. The branching ratio of pions decaying into electrons vs. muons is sensitive to a variety of new physics. Currently, the theory value for this branching ratio is much more precise than the experimental value, so the PIONEER experiment hopes to improve on its predecessor experiments in order to approach the theory precision.

The Mu2e experiment [82], which measures the neutrinoless conversion of the muon into electrons in the presence of a nucleus, will succeed the Muon $g-2$ experiment at the muon campus at Fermilab. This process is lepton flavor violating and has never been observed. Lepton flavor violation was long thought to be forbidden in the SM, but neutrino oscillations are evidence of LFV, although the SM LFV is small. cLFV (lepton flavor violation in charged particles) has not been observed; measuring this process would be evidence of new physics. The Mu3e experiment [83, 84] is also searching for cLFV with the decay $\mu^+ \to e^+e^-e^+$ at PSI. The MEG-II experiment [85] is searching for the cLFV decay $\mu^+ \to e^+\gamma$. MEG-II is an upgrade to the MEG experiment, which finished data taking in 2013, achieving an upper limit on the branching ratio for the decay of 5.3×10^{-13}. MEG-II hopes to improve the sensitivity to 6×10^{-14} and has just started data taking.

With stronger evidence of new physics from the Fermilab Run-1 result, there is even more motivation to search for BSM physics with leptons. These future experiments, which will search for muon EDMs,

lepton flavor universality, and charged lepton flavor violations, will hopefully shed light on scenarios for new physics beyond the Standard Model.

Bibliography

[1] G.W. Bennett et al. Final Report of the Muon E821 Anomalous Magnetic Moment Measurement at BNL. *Phys.Rev.*, D73:072003, 2006. doi:10.1103/PhysRevD.73.072003.

[2] B. Abi, T. Albahri, S. Al-Kilani, D. Allspach, L.P. Alonzi, A. Anastasi, A. Anisenkov, F. Azfar, K. Badgley, S. Baeßler, et al. Measurement of the Positive Muon Anomalous Magnetic Moment to 0.46 ppm. *Phys. Rev. Lett.*, 126:141801, Apr 2021. doi:10.1103/PhysRevLett.126.141801.

[3] T. Albahri, A. Anastasi, A. Anisenkov, K. Badgley, S. Baeßler, I. Bailey, V. A. Baranov, E. Barlas-Yucel, T. Barrett, A. Basti, et al. Measurement of the anomalous precession frequency of the muon in the Fermilab Muon $g-2$ Experiment. *Phys. Rev. D*, 103:072002, Apr 2021. doi:10.1103/PhysRevD.103.072002.

[4] T. Albahri, A. Anastasi, K. Badgley, S. Baeßler, I. Bailey, V. A. Baranov, E. Barlas-Yucel, T. Barrett, F. Bedeschi, M. Berz, et al. Beam dynamics corrections to the Run-1 measurement of the muon anomalous magnetic moment at Fermilab. *Phys. Rev. Accel. Beams*, 24:044002, Apr 2021. doi:10.1103/PhysRevAccelBeams.24.044002.

[5] T. Albahri, A. Anastasi, K. Badgley, et al. Magnetic-field measurement and analysis for the Muon $g-2$ Experiment at Fermilab. *Phys. Rev. A*, 103:042208, Apr 2021. doi:10.1103/PhysRevA.103.042208.

[6] T. Aoyama et al. The anomalous magnetic moment of the muon in the Standard Model. *Phys. Rept.*, 887:1–166, 2020. doi:10.1016/j.physrep.2020.07.006.

[7] J. Grange et al. Muon (g-2) Technical Design Report. 1 2015. doi:10.48550/ARXIV.1501.06858.

[8] B. L. Roberts and E. Swanson. E989 note 253: Are e989 and e821 independent experiments? E989 DocDB 23927, 2022.

[9] John David Jackson. *Classical Electrodynamics*. Wiley, New York, NY, 3rd edition, 1999.

[10] P. A. M. Dirac. The Quantum theory of electron. Part II. *Proc. Roy. Soc. Lond.*, A118:351, 1928.

[11] Matthew D. Schwartz. *Quantum Field Theory and the Standard Model*. Cambridge University Press, 3 2014. ISBN 978-1-107-03473-0, 978-1-107-03473-0.

[12] D. E. Nagel, R. S. Julian, and J. R. Zacharias. The Hyperfine Structure of Atomic Hydrogen and Deuterium. *Phys. Rev.*, 72:971, 1947. doi:10.1103/PhysRev.72.971.

[13] J. E. Nafe, E. B. Nelson, and I. I. Rabi. The Hyperfine Structure of Atomic Hydrogen and Deuterium. *Phys. Rev.*, 71:914–915, 1947. doi:10.1103/PhysRev.71.914.

[14] Julian S. Schwinger. On Quantum electrodynamics and the magnetic moment of the electron. *Phys. Rev.*, 73:416–417, 1948. doi:10.1103/PhysRev.73.416.

[15] M. Tanabashi et al. Review of Particle Physics. *Phys. Rev. D*, 98(3):030001, 2018. doi:10.1103/PhysRevD.98.030001.

[16] A. Czarnecki and W. J. Marciano. The muon anomalous magnetic moment: A harbinger for 'new physics'. *Phys. Rev.*, D64:013014, 2001. doi:10.1103/PhysRevD.64.013014.

[17] T. Aoyama, M. Hayakawa, T. Kinoshita, and M. Nio. Complete Tenth-Order QED Contribution to the Muon $g-2$. *Phys. Rev. Lett.*, 109:111808, 2012. doi:10.1103/PhysRevLett.109.111808.

[18] T. Aoyama, T. Kinoshita, and M. Nio. Theory of the Anomalous Magnetic Moment of the Electron. *Atoms*, 7(1):28, 2019. doi:10.3390/atoms7010028.

[19] R. Jackiw and S. Weinberg. Weak interaction corrections to the muon magnetic moment and to muonic atom energy levels. *Phys. Rev.*, D5:2396–2398, 1972. doi:10.1103/PhysRevD.5.2396.

[20] I. Bars and M. Yoshimura. Muon magnetic moment in a finite theory of weak and electromagnetic interaction. *Phys. Rev.*, D6:374–376, 1972. doi:10.1103/PhysRevD.6.374.

[21] K. Fujikawa, B. W. Lee, and A. I. Sanda. Generalized Renormalizable Gauge Formulation of Spontaneously Broken Gauge Theories. *Phys. Rev.*, D6:2923–2943, 1972. doi:10.1103/PhysRevD.6.2923.

[22] A. Czarnecki, W. J. Marciano, and A. Vainshtein. Refinements in electroweak contributions to the muon anomalous magnetic moment. *Phys. Rev.*, D67:073006, 2003. doi:10.1103/PhysRevD.67.073006. [Erratum: Phys. Rev. **D73**, 119901 (2006)].

[23] C. Gnendiger, D. Stoeckinger, and H. Stoeckinger-Kim. The electroweak contributions to $(g-2)_\mu$ after the Higgs boson mass measurement. *Phys. Rev.*, D88:053005, 2013. doi:10.1103/PhysRevD.88.053005.

[24] M. Davier, A. Hoecker, B. Malaescu, and Z. Zhang. Reevaluation of the hadronic vacuum polarisation contributions to the Standard Model predictions of the muon $g-2$ and $\alpha(m_Z^2)$ using newest hadronic cross-section data. *Eur. Phys. J.*, C77(12):827, 2017. doi:10.1140/epjc/s10052-017-5161-6.

[25] A. Keshavarzi, D. Nomura, and T. Teubner. Muon $g-2$ and $\alpha(M_Z^2)$: a new data-based analysis. *Phys. Rev.*, D97(11):114025, 2018. doi:10.1103/PhysRevD.97.114025.

[26] G. Colangelo, M. Hoferichter, and P. Stoffer. Two-pion contribution to hadronic vacuum polarization. *JHEP*, 02:006, 2019. doi:10.1007/JHEP02(2019)006.

[27] Martin Hoferichter, Bai-Long Hoid, and Bastian Kubis. Three-pion contribution to hadronic vacuum polarization. *JHEP*, 08:137, 2019. doi:10.1007/JHEP08(2019)137.

[28] M. Davier, A. Hoecker, B. Malaescu, and Z. Zhang. A new evaluation of the hadronic vacuum polarisation contributions to the muon anomalous magnetic moment and to $\alpha(m_Z^2)$. *Eur. Phys. J. C*, 80(3): 241, 2020. doi:10.1140/epjc/s10052-020-7792-2. [Erratum: Eur.Phys.J.C 80, 410 (2020)].

[29] A. Keshavarzi, D. Nomura, and T. Teubner. The $g-2$ of charged leptons, $\alpha(M_Z^2)$ and the hyperfine splitting of muonium. *Phys. Rev.*, D101:014029, 2020. doi:10.1103/PhysRevD.101.014029.

[30] A. Kurz, T. Liu, P. Marquard, and M. Steinhauser. Hadronic contribution to the muon anomalous magnetic moment to next-to-next-to-leading order. *Phys. Lett.*, B734:144–147, 2014. doi:10.1016/j.physletb.2014.05.043.

[31] Friedrich Jegerlehner. The Anomalous Magnetic Moment of the Muon. *Springer Tracts Mod. Phys.*, 274:1–693, 2017. doi:10.1007/978-3-319-63577-4.

[32] K. Melnikov and A. Vainshtein. Hadronic light-by-light scattering contribution to the muon anomalous magnetic moment revisited. *Phys. Rev.*, D70:113006, 2004. doi:10.1103/PhysRevD.70.113006.

[33] P. Masjuan and P. Sánchez-Puertas. Pseudoscalar-pole contribution to the $(g_\mu - 2)$: a rational approach. *Phys. Rev.*, D95(5):054026, 2017. doi:10.1103/PhysRevD.95.054026.

[34] G. Colangelo, M. Hoferichter, M. Procura, and P. Stoffer. Dispersion relation for hadronic light-by-light scattering: two-pion contributions. *JHEP*, 04:161, 2017. doi:10.1007/JHEP04(2017)161.

[35] Martin Hoferichter, Bai-Long Hoid, Bastian Kubis, Stefan Leupold, and Sebastian P. Schneider. Dispersion relation for hadronic light-by-light scattering: pion pole. *JHEP*, 10:141, 2018. doi:10.1007/JHEP10(2018)141.

[36] A. Gérardin, H. B. Meyer, and A. Nyffeler. Lattice calculation of the pion transition form factor with $N_f = 2 + 1$ Wilson quarks. *Phys. Rev.*, D100(3):034520, 2019. doi:10.1103/PhysRevD.100.034520.

[37] J. Bijnens, N. Hermansson-Truedsson, and A. Rodríguez-Sánchez. Short-distance constraints for the HLbL contribution to the muon anomalous magnetic moment. *Phys. Lett.*, B798:134994, 2019. doi:10.1016/j.physletb.2019.134994.

[38] G. Colangelo, F. Hagelstein, M. Hoferichter, L. Laub, and P. Stoffer. Longitudinal short-distance constraints for the hadronic light-by-light contribution to $(g-2)_\mu$ with large-N_c Regge models. *JHEP*, 03:101, 2020. doi:10.1007/JHEP03(2020)101.

[39] Vladyslav Pauk and Marc Vanderhaeghen. Single meson contributions to the muon's anomalous magnetic moment. *Eur. Phys. J.*, C74(8):3008, 2014. doi:10.1140/epjc/s10052-014-3008-y.

[40] I. Danilkin and M. Vanderhaeghen. Light-by-light scattering sum rules in light of new data. *Phys. Rev.*, D95(1):014019, 2017. doi:10.1103/PhysRevD.95.014019.

[41] M. Knecht, S. Narison, A. Rabemananjara, and D. Rabetiarivony. Scalar meson contributions to a_μ from hadronic light-by-light scattering. *Phys. Lett.*, B787:111–123, 2018. doi:10.1016/j.physletb.2018.10.048.

[42] G. Eichmann, C. S. Fischer, and R. Williams. Kaon-box contribution to the anomalous magnetic moment of the muon. *Phys. Rev.*, D101(5):054015, 2020. doi:10.1103/PhysRevD.101.054015.

[43] P. Roig and P. Sánchez-Puertas. Axial-vector exchange contribution to the hadronic light-by-light piece of the muon anomalous magnetic moment. *Phys. Rev.*, D101(7):074019, 2020. doi:10.1103/PhysRevD.101.074019.

[44] T. Blum, N. Christ, M. Hayakawa, T. Izubuchi, L. Jin, C. Jung, and C. Lehner. The hadronic light-by-light scattering contribution to the muon anomalous magnetic moment from lattice QCD. *Phys. Rev. Lett.*, 124(13):132002, 2020. doi:10.1103/PhysRevLett.124.132002.

[45] G. Colangelo, M. Hoferichter, A. Nyffeler, M. Passera, and P. Stoffer. Remarks on higher-order hadronic corrections to the muon $g - 2$. *Phys. Lett.*, B735:90–91, 2014. doi:10.1016/j.physletb.2014.06.012.

[46] M. Benayoun, L. DelBuono, and F. Jegerlehner. BHLS_2, a new breaking of the HLS model and its phenomenology. *The European Physical Journal C*, 80(2), feb 2020. doi:10.1140/epjc/s10052-020-7611-9.

[47] Sz. Borsanyi, Z. Fodor, J. N. Guenther, C. Hoelbling, S. D. Katz, L. Lellouch, T. Lippert, K. Miura, L. Parato, K. K. Szabo, F. Stokes, B. C. Toth, Cs. Torok, and L. Varnhorst. Leading hadronic contribution to the muon magnetic moment from lattice QCD. *Nature*, 593(7857):51–55, 2021. doi:10.1038/s41586-021-03418-1.

[48] T. Blum, P. A. Boyle, V. Gülpers, T. Izubuchi, L. Jin, C. Jung, A. Jüttner, C. Lehner, A. Portelli, and J. T. Tsang. Calculation of the hadronic vacuum polarization contribution to the muon anomalous magnetic moment. *Phys. Rev. Lett.*, 121(2):022003, 2018. doi:10.1103/PhysRevLett.121.022003.

[49] D. Giusti, V. Lubicz, G. Martinelli, F. Sanfilippo, and S. Simula. Electromagnetic and strong isospin-

breaking corrections to the muon $g-2$ from Lattice QCD+QED. *Phys. Rev. D*, 99(11):114502, 2019. doi:10.1103/PhysRevD.99.114502.

[50] E. Shintani and Y. Kuramashi. Hadronic vacuum polarization contribution to the muon $g-2$ with 2+1 flavor lattice QCD on a larger than $(10\text{ fm})^4$ lattice at the physical point. *Phys. Rev. D*, 100(3):034517, 2019. doi:10.1103/PhysRevD.100.034517.

[51] C. T. H. Davies et al. Hadronic-vacuum-polarization contribution to the muon's anomalous magnetic moment from four-flavor lattice QCD. *Phys. Rev. D*, 101(3):034512, 2020. doi:10.1103/PhysRevD.101.034512.

[52] D. Giusti and S. Simula. Lepton anomalous magnetic moments in Lattice QCD+QED. *PoS*, LATTICE2019:104, 2019. doi:10.22323/1.363.0104.

[53] C. Lehner and A. S. Meyer. Consistency of hadronic vacuum polarization between lattice qcd and the r ratio. *Phys. Rev. D*, 101:074515, Apr 2020. doi:10.1103/PhysRevD.101.074515.

[54] G. Colangelo, M. Davier, A. X. El-Khadra, et al. Prospects for precise predictions of a_μ in the standard model. 2022. doi:10.48550/ARXIV.2203.15810.

[55] Gilberto Colangelo, Martin Hoferichter, Massimiliano Procura, and Peter Stoffer. Dispersive approach to hadronic light-by-light scattering. *JHEP*, 09:091, 2014. doi:10.1007/JHEP09(2014)091.

[56] Gilberto Colangelo, Martin Hoferichter, Massimiliano Procura, and Peter Stoffer. Dispersion relation for hadronic light-by-light scattering: theoretical foundations. *JHEP*, 09:074, 2015. doi:10.1007/JHEP09(2015)074.

[57] T. Blum, N. Christ, M. Hayakawa, T. Izubuchi, L. Jin, C. Jung, and C. Lehner. Connected and Leading Disconnected Hadronic Light-by-Light Contribution to the Muon Anomalous Magnetic Moment with a Physical Pion Mass. *Phys. Rev. Lett.*, 118(2):022005, 2017. doi:10.1103/PhysRevLett.118.022005.

[58] En-Hung Chao, Renwick J. Hudspith, Antoine Gérardin, Jeremy R. Green, Harvey B. Meyer, and Konstantin Ottnad. Hadronic light-by-light contribution to $(g-2)_\mu$ from lattice QCD: a complete calculation. *Eur. Phys. J. C*, 81(7):651, 2021. doi:10.1140/epjc/s10052-021-09455-4.

[59] P. Athron, C. Balázs, D. H. J. Jacob, W. Kotlarski, D. Stöckinger, and H. Stöckinger-Kim. New physics explanations of a_μ in light of the FNAL muon g $-$ 2 measurement. *JHEP*, 09:080, 2021. doi:10.1007/JHEP09(2021)080.

[60] D. Hanneke, S. Fogwell Hoogerheide, and G. Gabrielse. Cavity Control of a Single-Electron Quantum Cyclotron: Measuring the Electron Magnetic Moment. *Phys. Rev. A*, 83:052122, 2011. doi:10.1103/PhysRevA.83.052122.

[61] P. J. Mohr, B. N. Taylor, and D. B. Newell. CODATA Recommended Values of the Fundamental Physical Constants: 2010. *Rev. Mod. Phys.*, 84:1527–1605, 2012. doi:10.1103/RevModPhys.84.1527.

[62] W. Liu et al. High precision measurements of the ground state hyperfine structure interval of muonium and of the muon magnetic moment. *Phys. Rev. Lett.*, 82:711–714, 1999. doi:10.1103/PhysRevLett.82.711.

[63] A. Fienberg. *Measuring the Precession Frequency in the E989 Muon $g-2$ Experiment*. PhD thesis, University of Washington, 2019.

[64] D. Stratakis, B. Drendel, J. P. Morgan, M. J. Syphers, and N. S. Froemming. Commissioning and First Results of the Fermilab Muon Campus. *Phys. Rev. Accel. Beams*, 22(1):011001, 2019. doi:10.1103/PhysRevAccelBeams.22.011001.

[65] K. S. Khaw et al. Performance of the Muon $g-2$ calorimeter and readout systems measured with test beam data. *Nucl. Instrum. Meth.*, A945:162558, 2019. doi:10.1016/j.nima.2019.162558.

[66] A. T. Fienberg et al. Studies of an array of PbF_2 Cherenkov crystals with large-area SiPM readout. *Nucl. Instrum. Meth. A*, 783:12–21, 2015. doi:10.1016/j.nima.2015.02.028.

[67] B.T. King, T. Albahri, S. Al-Kilani, et al. The straw tracking detector for the fermilab muon $g-2$ experiment. *Journal of Instrumentation*, 17(02):P02035, Feb 2022. doi:10.1088/1748-0221/17/02/p02035.

[68] D. Sweigart. *A Measurement of the Anomalous Precession Frequency of the Positive Muon*. PhD thesis, Cornell University, 2020.

[69] W. Gohn. Data Acquisition with GPUs: The DAQ for the Muon $g-2$ Experiment at Fermilab. *PoS*, ICHEP2016:174, 2016. doi:10.22323/1.282.0174.

[70] J. Hempstead. *Measuring the anomalous precession frequency ω_a for the Muon $g-2$ experiment*. PhD thesis, University of Washington, 2021.

[71] J. Kaspar et al. Design and performance of SiPM-based readout of PbF_2 crystals for high-rate, precision timing applications. *JINST*, 12(01):P01009, 2017. doi:10.1088/1748-0221/12/01/P01009.

[72] A. Anastasi et al. The laser-based gain monitoring system of the calorimeters in the Muon $g-2$ experiment at Fermilab. *JINST*, 14:P11025, 2019. doi:10.1088/1748-0221/14/11/P11025.

[73] S. Foster. Empirical pileup correction method. E989 DocDB 25130, 2021.

[74] G. W. Bennett et al. Statistical equations and methods applied to the precision muon $g-2$ experiment at BNL. *Nucl. Instrum. Meth. A*, 579:1096–1116, 2007. doi:10.1016/j.nima.2007.06.023.

[75] J. Mott and D Tarazona. Systematics 101: Cbo. E989 DocDB 27513, 2022.

[76] J. Mott. Fitting sum of data from all calorimeters. E989 DocDB 25788, 2021.

[77] K. S. Khaw and Y. Zeng. Fit function, acceptance correlation and 1.9 mhz peak. E989 DocDB 26207, 2021.

[78] S. Foster. Investigating the K_{loss} Slow Effect. E989 DocDB 25868, 2021.

[79] M. Abe, S. Bae, G. Beer, et al. A new approach for measuring the muon anomalous magnetic moment and electric dipole moment. *Progress of Theoretical and Experimental Physics*, 2019(5), 05 2019. ISSN 2050-3911. doi:10.1093/ptep/ptz030. 053C02.

[80] W. Altmannshofer et al. Testing Lepton Flavor Universality and CKM Unitarity with Rare Pion Decays in the PIONEER experiment. In *2022 Snowmass Summer Study*, 3 2022. doi:10.48550/ARXIV.2203.05505.

[81] W. Altmannshofer et al. PIONEER: Studies of Rare Pion Decays. 3 2022. doi:10.48550/ARXIV.2203.01981.

[82] L. Bartoszek et al. Mu2e Technical Design Report. 10 2014. doi:10.2172/1172555.

[83] K. Arndt, H. Augustin, P. Baesso, et al. Technical design of the phase i mu3e experiment. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 1014:165679, 2021. ISSN 0168-9002. doi:https://doi.org/10.1016/j.nima.2021.165679.

[84] A. Blondel et al. Research Proposal for an Experiment to Search for the Decay $\mu \to eee$. 1 2013. doi:10.48550/ARXIV.1301.6113.

[85] A. M. Baldini, E. Baracchini, C. Bemporad, et al. The design of the MEG II experiment. *The European Physical Journal C*, 78(5):380, 5 2018. doi:10.1140/epjc/s10052-018-5845-6.

Appendix A

Muon loss correction C_{ml} in Run-1

This appendix discusses the calculation of the muon loss correction C_{ml} calculation for Run-1. This thesis focuses on Run-2/3, and the correction was irrelevantly small for these runs. Muon losses are described in subsection 5.3.2. A summary of the muon loss correction and its evaluation in Run-2/3 is given in subsection 3.2.5. These results below are published in [4] and [2].

A.1 Study 1: phase-momentum correlation

The muons arriving at the $g-2$ storage ring have a wide range of momenta, but the ring only accepts a small momentum slice. Different portions of the incoming distribution can be selected by changing the field setting, since the average stored momentum is proportional to the field strength. The simulated incoming momentum distribution with illustrative highlighted slices of the distribution is shown in Figure A.1. Positron data was taken at three field settings: nominal momentum, higher momentum, and lower momentum. After taking positron data at different field settings, the average phase at each momentum was extracted by performing a 5 parameter fit (Equation 5.34).

The beam arrives in the ring in sets of 16 pulses, which are comprised of a pattern of 8 pulse shapes repeating twice. The ω_a phases of the 8 beam pulse shapes vary by as much as 30 mrad due to differences in shape and timing. If the relative intensity of the pulses were to vary from dataset to dataset, for example, the extracted phase of the dataset could be artificially pulled by the pulse-to-pulse phase difference. Therefore, all 8 pulses were fit separately, and the final slopes were averaged together.

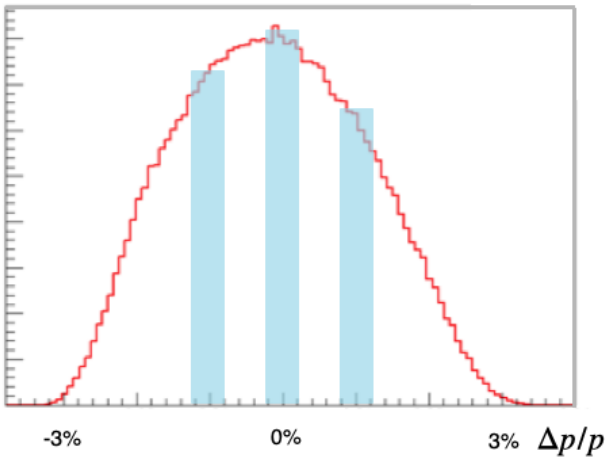


Figure A.1: Simulated momentum distribution upon injection, showing sample slices of the distribution which would be stored at different field settings.

The original intention of the study was to increase the magnetic field by 1%. To this end, the magnet current was increased by 1%. However, due to hysteresis effects in the magnet, a 1% increase in the magnet current did not correspond to a 1% increase in the actual field strength. Instead, a field increase of .7% was predicted.

The field change was calculated by using the τ parameter from the precession frequency fit. $\frac{\Delta B}{B} = \Delta\tau - 1 = 64.86 - 64.43 - 1 = 0.67\%$. This calculation is consistent with the prediction of the field team, so the field increase, and therefore the momentum increase, was determined to be .67%. A similar calculation can be done for the field decrease.

The phase of each fit is then plotted against the field setting for each pulse, and a linear fit is applied (Figure A.2). The final phase-momentum correlation is extracted by averaging the phase at each field setting and fitting a line to it. The final result is shown in Figure 3.8; the sign has been flipped relative to Figure A.2 as a result of sign conventions used in the publication.

The phase-momentum correlation uncertainty is measured to be 10 ± 1.6 mrad. This matches with the theoretically predicted value of $11.8 \pm 0.5 \frac{\text{mrad}}{\%\Delta p/p}$, shown in blue.

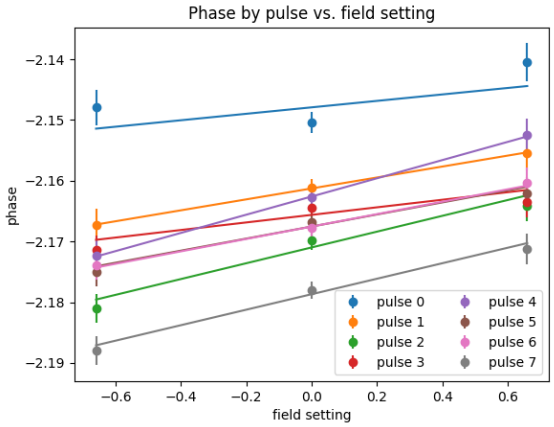


Figure A.2: Linear fits to phase as a function of field strength (proportional to momentum) setting for pulses 0-7 showing the correlation between phase and momentum.

A.1.1 Possible sources of uncertainty

The phase ϕ depends on the definition of $t = 0$. Ideally, $t = 0$ would to be defined at beam injection, but because the ω_a extraction does not depend on this definition, it has not been extracted precisely. However, because the magnetic field change affects the ω_a frequency, a shift in $t = 0$ can cause a phase advance between the two different measurements that are being compared. The exaggerated schematic in Figure A.3 illustrates the problem. If the two oscillations have a different frequency, a shift in the $t = 0$ definition between the measurements will cause a false phase advance. For example, if $t = 0$ in this plot is the true start of precession, one would conclude that the phase of both fits are 0; however, if the precession started at $t = -0.1$, one would conclude that the phases are non-zero and different. For this reason it is important to determine how close $t = 0$ is to the injection time after applying all time corrections.

The time correction scheme can be seen in Figure A.4. The T0 start-time detector and each calorimeter crystal receive a laser synchronization pulse. Each channel is time-aligned to the sync pulse. Then global $t = 0$ is defined to be the time centroid of the beam pulse in the T0 detector. A shift in global $t = 0$ could result from the long laser fiber that brings the sync pulse from the calorimeter 22 crate to the T0 detector, causing a constant offset in the arrival of the calorimeter sync pulse and the T0 sync pulse, which is not corrected for. This $t_{\mathrm{T0}}^{\mathrm{fiber}}$ time would cause the beam to arrive artificially at $t > 0$ in the calorimeters, even after accounting for the time the beam takes to get from the T0 detector to the calorimeters. There are several

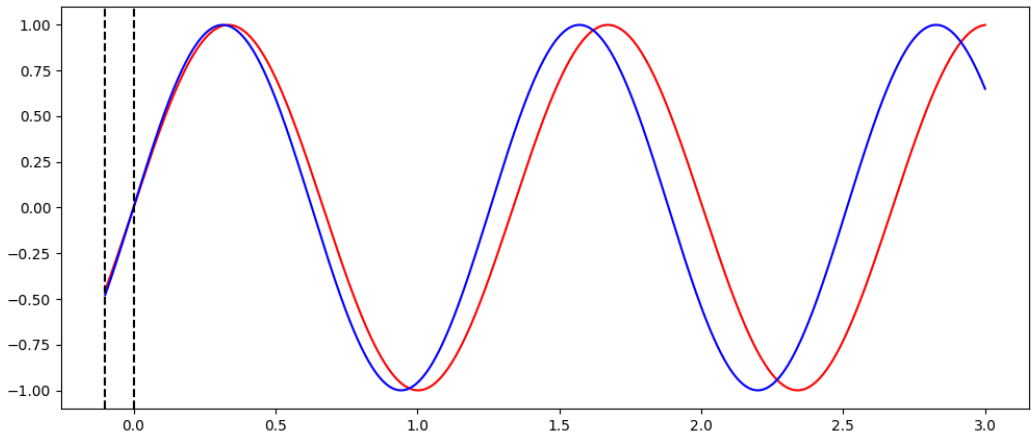


Figure A.3: Exaggerated cartoon showing the phase advance effect of a shifting global $t = 0$ when two functions have different frequencies.

ways to get a handle on this effect.

One strategy is to look at the time of the beam splash in the calorimeters after all timing corrections have been applied. The splash is impossible to fit exactly because it saturates the calorimeters and bombards them with many particles, causing a chain-fit reconstruction be be very slow. However, the splash timing can be imprecisely studied by using the faster GPU fitting and weighting by the energy. A plot of this GPU-fit splash as compared to the T0 detector pulse timing can be seen in Figure A.5, where all time corrections have been applied (so that $t = 0$ is defined to be the centroid of the T0 detector pulse). Based on this plot, the splash is probably arriving in the early calorimeters within ≈ 50 ns of the T0 pulse, although it is hard to extract the timing more precisely than that.

A second strategy is to extract the $t = 0$ time from the fast rotation analysis. $t = 0$ can be extracted by minimizing the fast rotation amplitude of the nonphysical region ($R > 45$ mm, where the beam cannot exist). An incorrect $t = 0$ distorts the distribution so that that some of the beam intensity appears in the nonphysical region. A scan of $t = 0$ for the nominal momentum and high momentum cases in order to minimize the nonphysical amplitude is seen in Figure A.6. This method gives a $t = 0$ value of $\approx 66 - 67$ ns in both cases.

This is consistent with the value extracted from the GPU fit method, so the $t = 0$ time is correct to

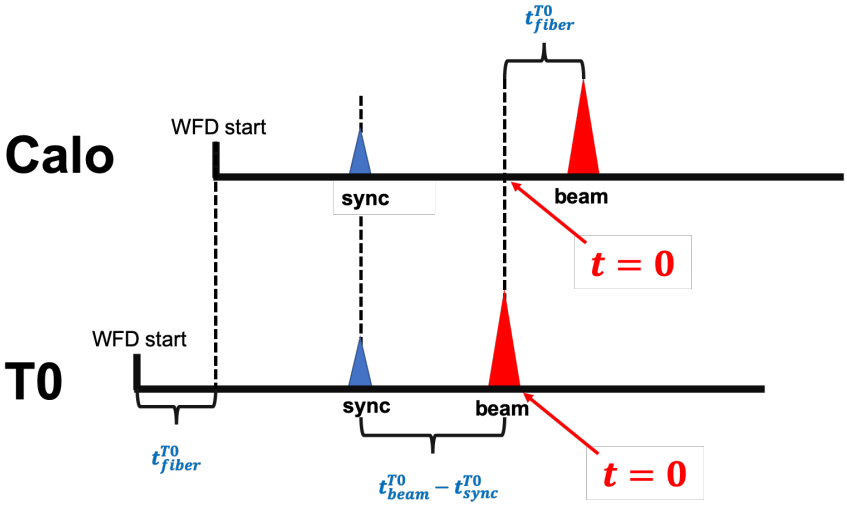


Figure A.4: Cartoon showing the time alignment scheme in the reconstruction and the effect on global $t = 0$ of a long fiber between calo 22 and the T0 detector.

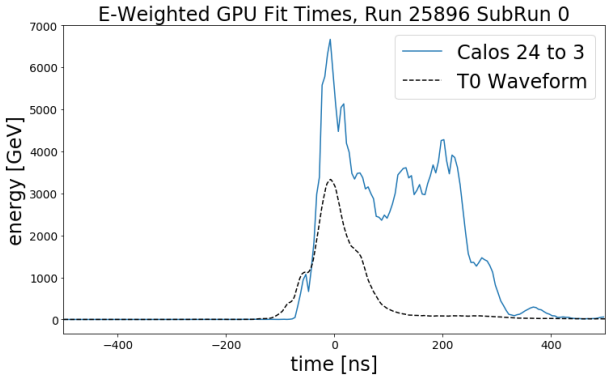


Figure A.5: Energy-weighted GPU fits of the splash in the early calorimeters as compared to the timing of the beam passing through the T0 detector. Reproduced from A. Fienberg.

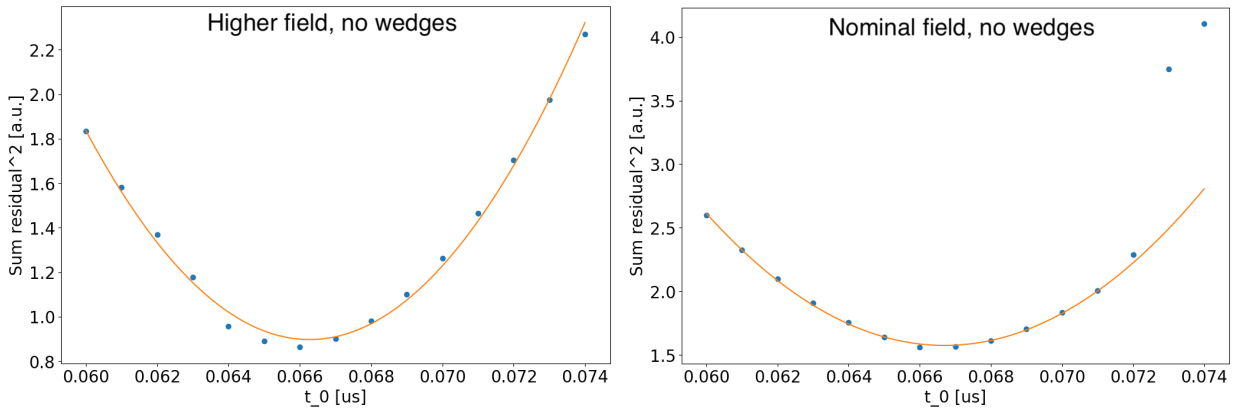


Figure A.6: Fast rotation $t = 0$ scans for the high field and nominal field cases showing the residuals of the nonphysical region as a function of the $t = 0$ time. Minimizing the residuals allows us to estimate $t = 0$. Reproduced from K.S. Khaw.

within 100 ns as a conservative estimate. If there is a shift of $\Delta t_0 = 100$ ns, then the phase will be affected by $\Delta\phi = \Delta t_0 \cdot \omega_a \cdot \frac{\Delta B}{B} = 100 \text{ ns} \cdot \frac{1.4 \text{ rad}}{\mu s} \cdot 0.67\% \approx 1$ mrad. So there is an uncertainty of < 1 mrad due to the global $t = 0$ phase advance effect.

The original intention of this first study was to run at a +1% increased field, but the actual field increase was 0.67%. Therefore, the inflector current was increased by 1% to cancel out the increased field value, meaning that for the high or low field cases, there was a nonzero field in the inflector. This would cause a phase advance in the high or low field cases compared to the nominal field case. The phase change due to this field is

$$\begin{aligned} \Delta\phi &= \frac{d\phi}{dt} \cdot t \\ &= \omega_a^{\text{infl}} \cdot t \end{aligned} \tag{A.1}$$

Where ω_a^{infl} is the the ω_a frequency in the inflector field B_{infl}. Then using the nominal $\omega_a^{\text{nom}} = \frac{1.4 \text{ rad}}{1\ \mu\text{s}}$ in

the nominal field $B_{\text{nom}} = 1.45$ T,

$$\begin{aligned}\Delta\phi &= \omega_a^{\text{nom}} \cdot \frac{B_{\text{infl}}}{B_{\text{nom}}} \cdot \frac{L}{c} \\ &= \frac{1.4 \text{ rad}}{1 \times 10^{-6} \text{ s}} \cdot 0.0044 \cdot \frac{2 \text{ m}}{3 \times 10^{8} \text{ m/s}} \\ &= 41 \ \mu\text{rad}\end{aligned} \tag{A.2}$$

So the effect is $\approx .04$ mrad and can be ignored.

A.2 Study 2: momentum dependence of losses

The first half of the study demonstrated that there is a correlation between muon phase and momentum. This means that if the losses are momentum dependent, there will be a systematic shift to ω_a.

The goal of the second half of the study is to build a data-driven loss probability function (the probability that a muon will be lost as a function of momentum). Then this loss function can be used to evolve the stored momentum distribution through time in the fill. As the average momentum of the stored distribution changes, so will the muon phase, because muons with a certain phase will be preferentially lost. This procedure will give the average phase of the stored distribution as a function of the time in the fill. Then the pull on ω_a can be evaluated from the time changing phase.

In order to measure the loss probability function, a set of special runs were taken where the incoming muon momentum distribution was changed using delivery ring (DR) collimators. In the DR, the radial position of the muons is correlated with momentum, with higher momentum muons tending to be towards the outside of the DR. Therefore the DR collimators on one side or the other are used to preferentially remove muons with a certain momentum. This shifts the average momentum of the stored muon distribution downstream. In the storage ring, the number of lost muons for each DR collimator setting are counted to determine the dependence of lost muons on momentum.

A.2.1 Data taking procedure

For this study, which was during Run-2, the Run-1 data conditions were replicated as closely as possible. Run-1 losses were an order of magnitude higher than Run-2 losses, due to the damaged quad resistors, so

the broken resistors were replaced for this study.

The DR collimator set points were determined using the reduction in the number of CTAGs, estimates of decay positrons, relative to nominal data taking. For example, "1/5 low" means that the *high* momentum side of the distribution was scraped off in the DR until only 1/5 of the nominal CTAGs remained downstream in the storage region. Therefore, data was taken for longer periods when the CTAGs were reduced more. A “mid momentum" dataset was also taken, where both DR collimators were inserted in order to select the most central momentum muons. In total, 8 datasets with different momentum distributions were taken.

A.2.2 Results

The equilibrium radius distribution for each systematic dataset was determined by the Cornell fast rotation (FR) analysis. Maximum background subtraction was applied, because the analysis is so sensitive to noise at the edges of the distribution. These perturbed FRs are shown in Figure A.7. The gray line corresponds to the nominal distribution (no DR collimators) for normal data taking. The colored lines correspond to the collimator-biased distributions, with blue being the lowest momentum dataset and red being the highest. The integral of each $F(x)_i$ is normalized so that the integral of the “1/5 CTAG, low momentum" curve is 1/5 the nominal integral.

The muon loss rate vs. time was measured for each systematic dataset. Sample triples spectra, normalized to the total CTAGs for the dataset, for the 1/5 CTAG low and high momentum are shown in Figure 3.9. From this plot the asymmetry between low and high momentum losses is evident. At the beginning of the fill, there were more muon losses in the low momentum case than in the high momentum case. Further, this figure demonstrates that the muon loss behavior is changing throughout the fill; later in the fill, the losses are more symmetric between the two datasets. These data are used to construct an empirical mapping between the perturbed FR and its corresponding loss spectrum in order to quantify the loss probability as a function of muon momentum. This loss probability will evolve in time to account for the changing behavior of the loss spectrum.

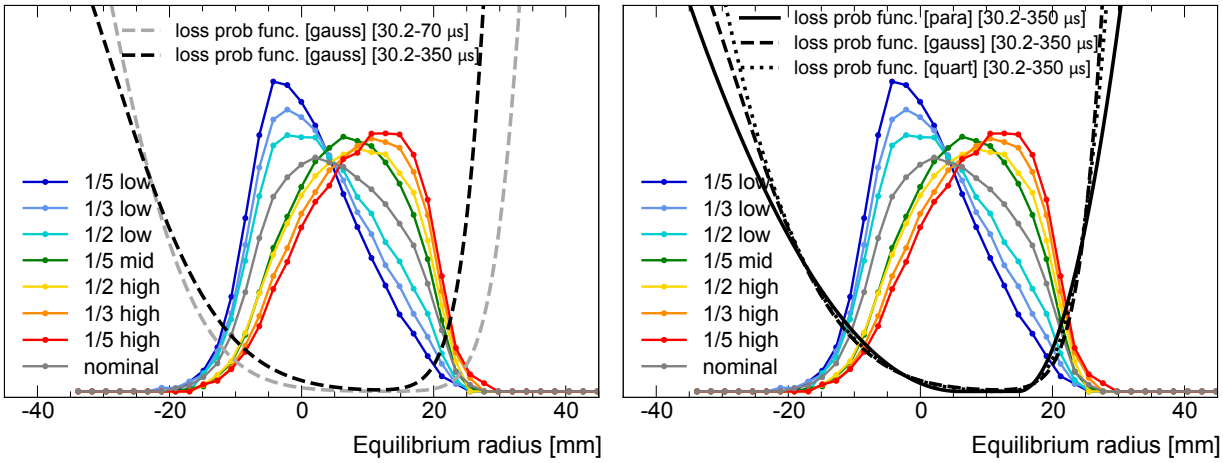


Figure A.7: Perturbed fast rotation distributions for the eight systematic datasets, created by different DR collimator settings. The integral of each curve is normalized to demonstrate the intensity relative to the nominal (no DR collimator) case, in gray. The plot on the left shows a comparison of the Gaussian loss probability function $l(x)$ early in the fill and over the whole integration period, showing that $l(x)$ starts off more asymmetric and then becomes more symmetric. The plot on the right shows a comparison of the three functional forms of $l(x)$ showing that they differ slightly but demonstrate the same qualitative behavior.

A.2.3 Constructing the loss probability function

Empirical loss probability functions as a function of time and equilibrium radius ($l(x,t)$) are constructed using the FR distributions and muon loss spectra measured during the systematic runs. Because there are only 8 data points, an analytical form of the loss probability function must be assumed. Based on input from simulation, the loss probability function is expected to peak at the edges of the storage distribution (high or low momentum) and be lower near the center of the distribution. That said, the relative shapes and amplitudes of the function can be different at the high and low edges of the storage region. In order to quantify the effect of this choice of analytical form on the uncertainty calculation, 3 functions were chosen:

- Double Gaussian, with the center of each Gaussian at the minimum and maximum radii of the storage distribution, $R_{\text{min}} = -45$ mm, $R_{\text{max}} = 45$ mm.
  $l(x) = A_{\text{min}} \exp(-\frac{x-R^2_{\text{min}}}{2\sigma^2_{\text{min}}}) + A_{\text{max}} \exp(-\frac{x-R^2_{\text{max}}}{2\sigma^2_{\text{max}}})$

- Double piecewise parabola, with 0 amplitude between the two parabolas.

$$l(x) = \begin{cases} A_1(x - B_1)^2, \text{ for } x < B_1 \\ 0, \text{ for } B_1 \leq x \leq B_2 \\ A_2(x - B_2)^2, \text{ for } x > B_2 \end{cases}$$

- Double piecewise quartic function.

$$l(x) = \begin{cases} A_1(x - B_1)^4, \text{ for } x < B_1 \\ 0, \text{ for } B_1 \leq x \leq B_2 \\ A_2(x - B_2)^4, \text{ for } x > B_2 \end{cases}$$

The 4 parameters of the Gaussian (A_{min}, σ_{min}, A_{max}, σ_{max}) and the 4 parameters of the parabolic and quartic functions (A_1, B_1, A_2, B_2) are determined by a two dimensional curve fit to a system of equations of the form:

$$\int_{R_{min}}^{R_{max}} F(x)_i l(x,t) dx = \frac{1}{\text{C}_\text{i}} \int_{t_s}^{t} e^{t'/\gamma\tau_\mu} L(t')_i dt', \tag{A.3}$$

where i runs from 1 to 8 for each of the special runs. The left side is an integral over the storage region, convolving the fast rotation distribution ($F(x)_i$) with the loss probability function. Each $F(x)_i$ is normalized to one. Convolved, these functions represent the momentum distribution of the muons that have been lost at that point in the fill. The right side integrates over the triples spectrum $L(t)_i$ for each systematic dataset. The additional term of $e^{t/\gamma\tau}$ is included so that the left side follows the convention of loss metric $f_{loss}(t)$, which is the fractional losses as a function of time. $f_{loss}(t)$ will provide the absolute scaling of the losses. $L(x,t)$ is integrated from the start time t_s to the current time t for the same reason. $t_s = 30\mu s$ for Run-1a, b, and c, and $t_s = 50\mu s$ for Run-1d. C_i is the total number of CTAGs for that dataset, so that all 8 datasets can be accurately compared.

The python code used to create the $l(r,t)$ functions is as follows. For each analytical form of $l(r,t)$ a function called `integrateLossProb`, which performs a piecewise integration of $F(x)$ and $l(x)$.

```
def integrateLossProb(X, p0, p1, p2, p3):
    xs,ys = X
    return np.trapz([y*lossProb(x, p0, p1, p2, p3) for y, x in zip(ys, xs)], xs)
```

The input `X` is a tuple `(xs_list, FRs_list)`, where `xs_list` is the list of the equilibrium radius coordinates of the 8 FR distributions and `FRs_list` is a list of their intensities. `p0, p1, p2, p3` are the four free parameters in the loss probability functions. `lossProb` is one of the three $l(r)$ functional forms described above.

The fit parameters are extracted using:

```
popt, _ = curve_fit(integrateLossProb, (xs_list, FRs_list), integrated_triples)
```

Where `curve_fit` is scipy curve fitting. `integrated_triples` is a list of integrated normalized triples for the 8 datasets, integrated from the start time to the current time bin. So `xs_list` and `FRs_list` are lists of 8 lists and `integrated_triples` is a list of 8 constants. `curve_fit` is then able to return the best fit parameters for the 4 $l(x)$ constants. This procedure can be repeated for each time bin, integrating from the fit start, to create all the $l(x,t)$ functions. The changing values of `integrated_triples` relative to each other create the time dependent behavior of $l(x,t)$.

Figure A.7 (left) shows an example of how $l(x,t)$ varies with time to account for the changing behavior of the triples spectrum. At the beginning of the fill, integrating from [30 μs, 70 μs], there are more losses coming from the low momentum side of the storage region. Over the whole fill, the losses are more symmetric. Figure A.7 (right) shows a comparison of the 3 analytical forms of $l(x)$ for the whole integration period. Each functional form behaves slightly differently, but they all peak at the edge of the storage region and dip in the center, as desired.

Leave-one-out cross-validation

One method of testing the predictive power of the $l(x)$ models is by leave-one-out cross-validation (LOOCV). Each of the 8 systematics datasets described in the results section is left out one by one, so that the $l(x)$ model is created using the other 7 datasets. Then the number of losses in the last dataset is predicted. The results of this procedure for the three $l(x)$ models are shown in Figure A.8. It is important to note that the more extreme distributions (i.e. 1/5 low and 1/5 high) contain more information about $l(x)$ than the middle distributions, and therefore it makes sense that excluding these datasets would lead to worse results. This plot suggests that the parabolic model has the most predictive power. However, in the "real life" case, which is predicting the nominal number of triples from the systematic datasets, all three models are predictive.

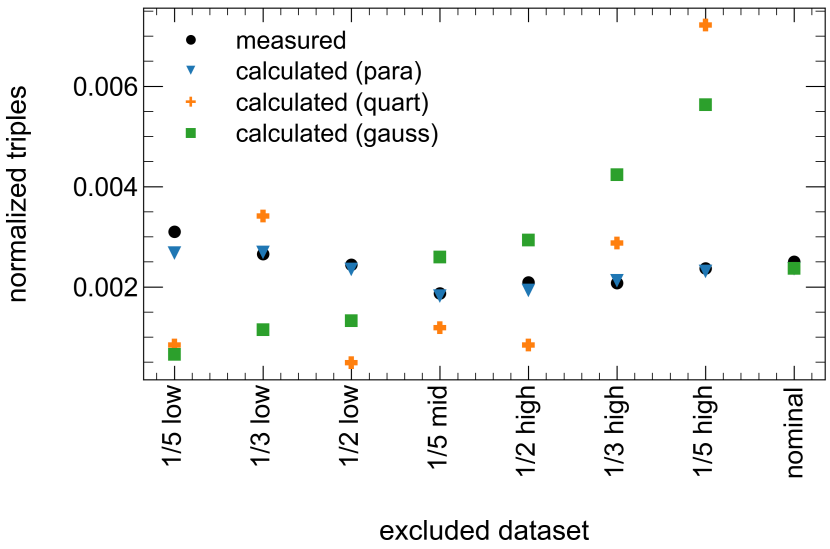


Figure A.8: Leave-one-out cross-validation results for the three $l(x)$ models. The number of normalized triples (to predict) is on the y axis, and the dataset that is excluded from the model building is on the y axis.

Calculating the systematic shift with all three models will give a conservative evaluate of the uncertainty uncertainty.

A.3 Extracting C_{ml}

We have demonstrated experimentally that phase and momentum are correlated and that muon losses are momentum dependent. This means that the stored muon average phase is changing throughout the fill, which will result in a pull on ω_a. A method for quantifying this systematic uncertainty using the data we have taken follows.

The starting point is the Run-1 fast rotation distribution for a certain dataset, $F_0(x)$. These distributions are shown in Figure 3.5. The goal is to evolve the FR distribution through the fill, extracting the average equilibrium radius at each time. The equilibrium radius (x) can then be converted to momentum using the equation

$$\frac{\Delta p}{p} \approx \frac{1-n}{R_0} x, \tag{A.4}$$

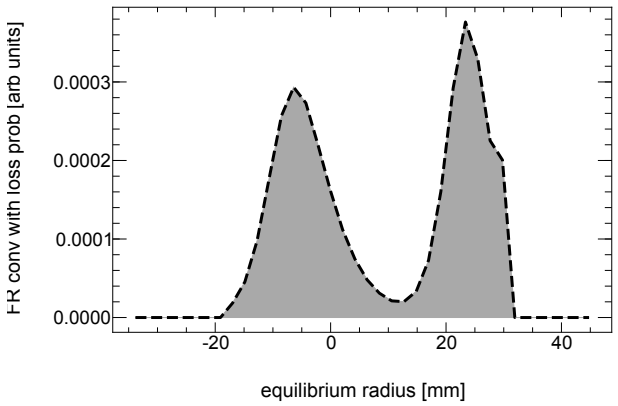


Figure A.9: Sample convolution of Run-1a fast rotation distribution with the Gaussian loss probability function over the whole fill, showing generally symmetric behavior.

where $R_0 = 7112$ mm is the magic radius and $n = 0.1075$ the field index at the 13.1/18.3 kV quad setting used for this study. Finally, momentum units will be converted to phase units using the measured phase-momentum correlation of $10.0 \pm 1.6 \frac{mrad}{\%\Delta p/p}$. This procedure creates $\langle \phi \rangle$ vs. time in the fill, which can be used to quantify the shift on ω_a.

The time loop begins 5 μs after the fit start time t_s so that there are enough statistics to integrate over and ends at 350 μs, after which there are very few losses. At each time bin t in [t_s+5 μs, 350 μs], $l(x,t)$ is calculated using Equation A.3. The loss function is then convoluted with the original fast rotation distribution, resulting in Figure A.9. These plots represent the momentum distribution shape of the total losses up to the time t. (As a side note, the bumpy feature on the right side of the convolution is due to the background subtraction for the Run-1 FR distributions.) However, because $F(x)_i$ are on an arbitrary scale, and because the triples in the $L(t)_i$ spectra have an unknown acceptance, the area under the convolution curve must be scaled using information from the ω_a analysis. Specifically, the area under the convolution curve is scaled using the $f_{loss}(t)$ value for the dataset so that the total losses are the correct percentage of the original distribution. The convolved distribution is then subtracted from the original $F(x)_i$ to give the equilibrium radius distribution at time t, $F_{curr}(x,t)$. In equation form, $F_{curr}(x,t)$ is given by

$$F_{curr}(x,t) = F_0(x) - f_{loss}(t) \frac{F_0(x) l(x,t) dx}{\int_{R_{min}}^{R_{max}} F_0(x) l(x,t) dx} \tag{A.5}$$

F_{curr} will then have a slightly different average equilibrium radius than the original F_0, showing that the lost muons have preferentially depleted one side of the momentum distribution. As the lost muons deplete the momentum distribution throughout the fill, the average momentum of the stored muons changes with time.

This procedure allows for the construction of $\langle p \rangle(t)$ throughout the fill for the stored muons. $\langle p \rangle(t)$ is converted to $\langle \phi \rangle(t)$ using the phase-momentum correlation of $10 \pm 1.6 \frac{mrad}{\%\Delta p/p}$ measured in the first half of the study. $\langle \phi \rangle(t)$ for the four Run-1 datasets is shown in Figure 3.10. The shaded uncertainty bands come from calculating with the 3 different analytical forms of $l(x)$. There is some divergence at the beginning of the fill due to the rapidly changing behavior and limited statistics, but the functional forms all give relatively convergent results.

In order to quantify the effect of the time changing phase on ω_a, $\langle \phi \rangle(t)$ is fit with a high degree polynomial function. This function is then plugged into a toy wiggle model. Points are generated using the function

$$N = N_0 e^{-\frac{t}{\gamma\tau}}(1 + A\cos(\omega_a t - \phi(t))), \tag{A.6}$$

where $\phi(t)$ is the polynomial fit function to the empirical $\phi(t)$ points, with the other constants taken to be representative values from an ω_a fit. These points are then fit using a standard five parameter fit function, which assumes a constant ϕ. The error is taken to be the square root of the counts in each time bin. This procedure replicates the ω_a analysis, where the data (which might contain a changing ϕ) is fit with a constant ϕ. Finally, the shift on ω_a is calculated in parts per billion by $\Delta\omega_a = 10^9 \times \frac{\omega_a' - \omega_a^0}{\omega_a^0}$, where ω_a' is the new ω_a extracted from the toy model with ϕ_t and ω_a^0 is the input $\omega_a = 1.43729$ MeV.

In summary, the procedure for calculating $\Delta\omega_a$ for each dataset is as follows:

1. Start with the nominal FR distribution.

2. At each time bin t:

   (a) Get the fraction loss correction $f_{loss}(t)$ from the ω_a analysis.

   (b) Find integrated normalized triples from the fit start time to t for the 8 systematic datasets.

(c) Using the 8 FR distributions, the 8 numbers from the previous step, and an assumed form of the loss probability function, fit for the loss function.

(d) Propagate the nominal FR distribution with this loss probability function to get the integrated losses at that time, and then scale to the fractional loss correction $f_{loss}(t)$.

(e) Subtract this convoluted distribution from the nominal FR distribution to get the distribution of the stored muons at time t.

(f) Calculate the average radius of the new FR distribution at t and convert from equilibrium radius units to $\Delta p/p$ units.

(g) Convert $\Delta p/p$ units to $\langle\phi\rangle$ units using the measured phase-momentum distribution, giving (t, $\langle\phi\rangle$) coordinate.

3. Perform polynomial fit to get a functional form of $\langle\phi\rangle$ vs. t.

4. Generate a set of wiggle points based on a 5 parameter function with $\langle\phi\rangle(t)$ as input.

5. Perform a 5 parameter fit using a time-independent ϕ to find $\Delta\omega_a$ between the input ω_a value and the fit ω_a value.

This procedure is repeated for every combination of $f_{loss}(t)$ function (Recon East without residual gain correction, Recon West with residual gain correction, Recon West without residual gain correction) and analytical form of $l(x)$ (Gaussian, parabolic, and quartic) for the four Run-1 datasets.

The shift $\Delta\omega_a$ is positive, meaning that the measured value of ω_a in the real data (ω_a') will be *higher* than the "true" physical value of ω_a. Therefore, a *negative* muon loss correction term, C_{ml}, is needed to correct the value of ω_a. The full C_{ml} results are listed in Table A.1.

Sources of uncertainty

There are three sources of uncertainty that are included in the final result. The first is the phase-momentum correlation uncertainty of 16%, which comes from the extracted phase-momentum correlation of $10 \pm 1.6\frac{mrad}{\%\Delta p/p}$. The second comes from the choice of analytical form of $l(x)$ (Gaussian, parabolic, or quartic) and is quantified using Table A.1. To quantify the uncertainty due to the choice of $l(x)$, the average and

	Recon west + resid			Recon West no resid			Recon east no resid		
	Gauss	Para	Quart	Gauss	Para	Quart	Gauss	Para	Quart
Run-1a	-17	-14	-14	-13	-11	-11	-16	-13	-14
Run-1b	-3	-3	-3	-2	-2	-2	-4	-3	-3
Run-1c	-9	-8	-8	-6	-5	-5	-9	-8	-8
Run-1d	-19	-17	-19	-16	-14	-16	-18	-16	-18

Table A.1: C_{ml} results in ppb from all combinations of $f_{loss}(t)$ function (recon east without residual gain correction, recon west with residual gain correction, recon west without residual gain correction) and analytical form of $l(x)$ (Gaussian, parabolic, and quartic) for the four Run-1 datasets.

standard deviation was taken of each set of three numbers for the same dataset with the same $f_{loss}(t)$. For example, for Run-1a, there are three numbers (-17, -14, -14) using Recon West with residual gain correction. The average of these numbers is -15, and the standard deviation is 2, for -15 ± 2. This is repeated for the other two sets of numbers to give -12 ± 1 and -14 ± 2 respectively. So for Run-1a, the average uncertainty due to the choice of $l(x)$ is the average of (1, 2, 2) = 2 ppb.

The third source of uncertainty is the different $f_{loss}(t)$ functions from the different analyzers. From the previous step, there are three central values for Run-1a from the three reconstructions. Then the average and standard deviation of these three values are taken, which for this case gives a central value of -14 ppb with an uncertainty of 2 ppb.

This gives the first column in Table A.2. For Run-1a, $C_{ml} = -14$ ppb. The phase-momentum correlation uncertainty is 16%, so the contribution to $\sigma_{C_{ml}}$ is 2 ppb. As calculated above, there is a 2 ppb uncertainty due to the form of $l(x)$ and the choice of f_{loss} function. This procedure is repeated for the other three datasets.

For each dataset, the uncertainties are treated as fully correlated and added linearly, giving the final results for $\sigma_{C_{ml}}$ in the bottom row of Table A.2. Notably, datasets Run-1b and Run-1c have lower C_{ml} because of their lower loss rates; this is due to the different quad settings. For all datasets, $|C_{ml}| < 17$ ppb and $\sigma_{C_{ml}} < 6$ ppb. The fit start times for Run-1a, b, and c was $30\mu s$, whereas the fit start time for Run-1d was $50\mu s$. This later start time reduced the magnitude of C_{ml} slightly.

	Run-1a	Run-1b	Run-1c	Run-1d
C_{ml}	-14	-3	-7	-17
Phase-momentum correlation	2	0	1	3
Form of $l(x)$	2	0	1	1
Choice of f_{loss} function	2	1	2	2
Linear sum ($\sigma_{C_{ml}}$)	6	1	4	6

Table A.2: Final C_{ml} and σ_{ml} results for the Run-1 data.

A.4 Conclusions

This appendix presented a successful measurement of the phase-momentum correlation lost muon systematic shift, C_{ml}, for Run-1. Data taking occurred in two parts, with the first part measuring the phase-momentum correlation itself and the second measuring the momentum dependence of the losses. The overall scaling of the shift came from a third empirical input, the fractional loss correction from ω_a fitting. The results from the first measurement agreed well with the prediction of the end-to-end simulation, lending credence to the simulation. For the second measurement, the chosen functional forms of the loss probability function were informed by a toy simulation of muon losses in the storage ring, although an assumption still had to be made.

The worst case scenario for the magnitude of the shift to ω_a, C_{ml}, was > 100 ppb. In the pathological case, muons are lost only from one side of the distribution. For example, all of the lost muons would have originated in the low side of the stored momentum distribution. With our measurement, we were able to show that the losses are only slightly asymmetric (for example, see Figure A.7). As a result, the phase-momentum correlation effect was suppressed relative to the pathological case. The magnitude of C_{ml} was < 17 ppb for all datasets, and the calculated shift will be applied to the Run-1 ω_a for each dataset. Further, the uncertainty $\sigma_{C_{ml}}$ was measured to be < 6 ppb, which is tiny on the scale of the large statistical uncertainties for Run-1.

C_{ml} was measured to be negative, as the phase-momentum correlation effect causes the measured value ω_a to be erroneously high. Therefore, it needs to be corrected downward using the C_{ml} term.

Appendix B

Covariance matrices

Full covariance matrices are given for Run-2 and Run-3NO. The covariance matrix for Run-3BM is given in the main text in Figure 5.23.

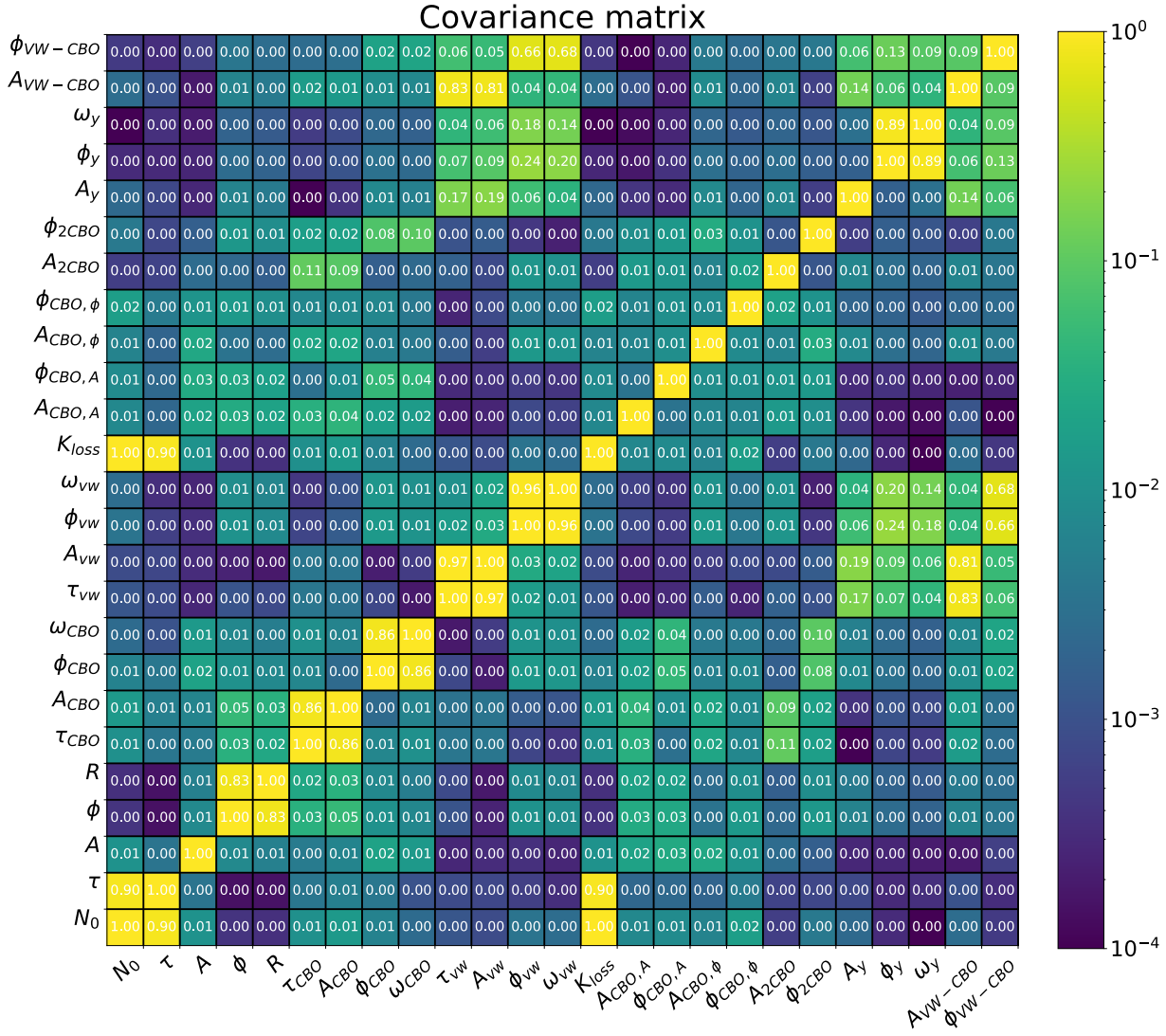


Figure B.1: Run-2 full fit covariance matrix.

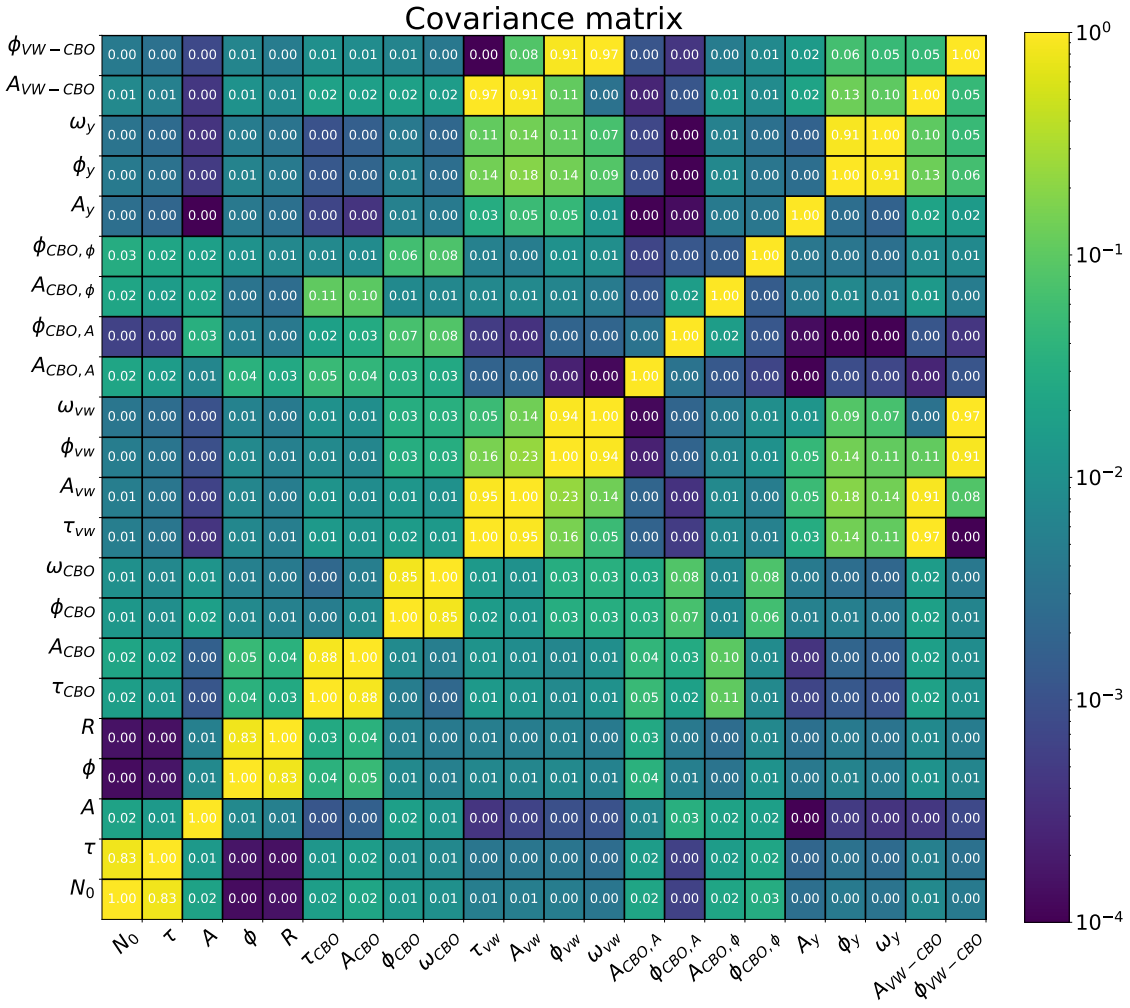


Figure B.2: Run-3NO full fit covariance matrix.