A MEASUREMENT OF THE ANOMALOUS PRECESSION FREQUENCY OF THE POSITIVE MUON

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

David Allen Sweigart

May 2020

A MEASUREMENT OF THE ANOMALOUS PRECESSION FREQUENCY OF THE POSITIVE MUON

David Allen Sweigart, Ph.D.

Cornell University 2020

The measurement of the anomalous magnetic moment of the muon, $a_\mu \equiv (g_\mu - 2)/2$, has been a historic test of our theoretical understanding of elementary particles and their interactions. At present, the world average is in tension with the value predicted by the Standard Model of particle physics by more than three standard deviations, possibly caused by new physics interactions. To resolve this discrepancy, the Muon $g - 2$ experiment at Fermi National Accelerator Laboratory aims to measure a_μ to a record 140 parts per billon using data taken over four years from 2018 to 2021. The experimental method involves trapping a polarized beam of positive muons in a storage ring containing an extremely uniform magnetic field. The difference in the muons' cyclotron and spin-precession frequencies, ω_a, is directly proportional to a_μ. This dissertation motivates making an improved measurement of a_μ; outlines the experimental method, with a focus on the backend electronics; and details the algorithm used to reconstruct the decay positrons impacting the electromagnetic calorimeters around the ring. Using data taken in 2018, a blinded measurement of ω_a to 410 parts per billion is then presented, which will allow a_μ to be determined with a precision comparable to that of the world average.

BIOGRAPHICAL SKETCH

David Allen Sweigart received a Bachelor of Science *summa cum laude* in physics and in mathematics in 2013 from the University of Maryland, Baltimore County. As an undergraduate student, he researched nonlinear optical materials for photonic applications from 2010 to 2013. In the summers of 2011 and 2012, he became involved in particle physics, working on the ATLAS and IceCube experiments through the University of Michigan/CERN Research Experience for Undergraduates Program and German Academic Exchange Service's Research Internships in Science and Engineering Program at Technische Universität Dortmund, respectively. David then continued his education in particle physics at Cornell University starting in 2013, initially working on the CMS experiment before switching after a year to the Muon $g - 2$ experiment. From 2015 to 2019, his research was conducted at Fermi National Accelerator Laboratory, where he contributed to designing, building, commissioning, and operating the Muon $g - 2$ experiment. Upon receiving his Ph.D., David will continue at Cornell University as a Postdoctoral Associate to see through the publication of the Muon $g - 2$ experiment's first physics result.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank the many scientists, engineers, and technicians who have contributed to the Muon $g - 2$ experiment and made the work described in this dissertation possible. I particularly want to thank my research advisor, Lawrence Gibbons, for his continual support and guidance. I am also very grateful to Kevin Labe, David Hertzog, Brendan Casey, Aaron Fienberg, Jarek Kaspar, Chris Polly, Antoine Chapelain, and Nate Rider for their support, feedback, and encouragment over the years. In addition, I am deeply thankful to Fermi National Accelerator Laboratory for hosting me from 2015 to 2019.

TABLE OF CONTENTS

<center>CHAPTER 1</center>

<center>**INTRODUCTION**</center>

Increasingly precise measurements of subatomic particles' magnetic dipole moments have historically led to breakthroughs in our theoretical understanding of the elementary particles and their interactions. In particular, the electron's g-factor, relating its spin to its magnetic moment, has helped guide us from quantum mechanics to the current Standard Model of particle physics. It is in fact the most precise physical quantity presently measured and, through vacuum fluctuations, should be sensitive to every known (and unknown) particle. Since it can also be unambiguously predicted to high precision, g—or, equivalently, $a \equiv (g-2)/2$—provides a stringent test of the Standard Model's validity and a way to guide the search for new physics.

If unknown particles do exist, the resulting discrepancy in a between theory and experiment is expected to be about four orders of magnitude larger for the muon compared to the electron. This discovery potential has spurred over time a series of experiments measuring the muon's anomalous magnetic moment, a_μ, to ever more precision, culminating with the currently running Muon $g-2$ experiment—also called E989—at Fermi National Accelerator Laboratory (FNAL). These experiments deduce a_μ by trapping muons in a known magnetic field and measuring the rate at which their spin vectors rotate relative to their velocity, which is known as the *anomalous precession frequency*. By further refining the experimental method, this latest-generation experiment at FNAL—the focus of this dissertation—is poised to resolve the previous measurement's intriguing a_μ-discrepancy with the Standard Model of more than three standard deviations.

These introductory concepts leading to the Muon $g-2$ experiment at FNAL are expounded in this chapter, including the current state of the theory and speculative extensions that predict a discrepancy in a_μ. The details of the experimental method employed are reserved for the next chapter, while subsequent chapters focus on the author's direct and significant contributions to

the a_μ-measurement. In particular, the last two chapters present a blinded measurement of the anomalous precession frequency to 410 parts per billion using data taken during the experiment's first running year in 2018. Once the Muon $g-2$ collaboration finishes measuring the magnetic field seen by the trapped muons, this dissertation's result can then be directly used to calculate the value of a_μ—and in turn to evaluate the Standard Model.

1.1 Anomalous Magnetic Moment

The magnetic moment $\boldsymbol{\mu}$ corresponding to a subatomic particle at rest with charge e, mass m, and intrinsic angular momentum, i.e., spin, $\boldsymbol{S}$, is defined as

$$\boldsymbol{\mu} = g \cdot \frac{e}{2m}\boldsymbol{S} \tag{1.1}$$

where g—known as the g-factor—is a dimensionless proportionality constant characterizing how the magnetic moment is related to the spin. In classical physics, where a particle's spin could be imagined as the orbital angular momentum from a charge distribution rotating about an axis of symmetry, g is predicted to be equal to one [1]. The same prediction holds in (non-relativistic) quantum mechanics, where the subatomic particle's intrinsic and orbital angular momentum are treated analogously [4].

It has long been experimental fact, however, that $g \neq 1$. In classical and quantum mechanics, when a particle is placed at rest in a static, external magnetic field, $\boldsymbol{B}$, the magnetic moment is expected to behave according to the Hamiltonian

$$H = -\boldsymbol{\mu} \cdot \boldsymbol{B} = -g \cdot \frac{e}{2m}\boldsymbol{S} \cdot \boldsymbol{B} \tag{1.2}$$

which is simply the potential energy [1]. This physical manifestation of spin enabled early measurements, such as of the Zeeman effect, to deduce for the electron that $g \approx 2$ [2, and references therein]. This first, large discrepancy in g with the contemporaneous theory established that the

mathematical description of spin was insufficient. Thus, the g-factor is commonly interpreted as the deviation from the classical and non-relativistic quantum prediction.

This significant tension between theory and experiment was soon shown to be a relativistic effect—known as Thomas precession [3]. The theory of special relativity, however, would not be married with quantum mechanics in one framework until the Dirac equation

$$\left(i\slashed{\partial} - e\slashed{A} - m\right)\psi = 0 \tag{1.3}$$

for a wave function ψ, expressed in natural units where $\hbar = c = 1$ [6]. Here, an external electromagnetic 4-vector, $A^\mu = (A^0, \boldsymbol{A})$, has been incorporated using minimal coupling. In addition to implying that antiparticles exist, Eq. 1.3 predicts that $g = 2$ for any elementary, spin-1/2 particle such as the electron or muon [7]. This prediction can be more clearly recognized when taken to the non-relativistic limit, which results in the Hamiltonian

$$H = \frac{1}{2m}(\boldsymbol{p} - e\boldsymbol{A})^2 - 2 \cdot \frac{e}{2m}\boldsymbol{S} \cdot \boldsymbol{B} + eA^0 \tag{1.4}$$

where $\boldsymbol{p}$ is the particle's momentum. By comparing the second term in Eq. 1.4 with Eq. 1.2, it is evident that this theory predicts $g = 2$. Also, except for the non-trivial g-factor, Eq. 1.4 has the same form as the Pauli equation from the earlier theory of non-relativistic quantum mechanics [4]. With $g = 2$ consistent with contemporaneous measurements for the electron, the Dirac equation was, thus, considered a breakthrough in our mathematical understanding of spin.

Despite this success, however, the g-factor of nucleons was soon measured to differ significantly from two. As shown in Tab. 1.1, the proton's g-factor is found to be more than twice that of the electron, and the neutron's g-factor even carries the opposite sign. Such large deviations from the Dirac equation's prediction were the first indication that they were not elementary particles, but rather composed of previously unknown ones. Eventually, these discrepancies would be resolved by the quark model in the theory of quantum chromodynamics. This makes a particle's g-factor

Particle	Experimental value	Relative uncertainty	Reference
Electron	$2.002\,319\,304\,361\,82(52)$	2.6×10^{-13}	[8]
Muon	$2.002\,331\,841\,8(13)$	6.3×10^{-10}	[8]
Tau	$1.964(34)$	1.8×10^{-2}	[9]
Proton	$5.585\,694\,702(17)$	3.0×10^{-9}	[8]
Neutron	$-3.826\,085\,45(90)$	2.4×10^{-7}	[8]

Table 1.1: The g-factor recently measured for various particles. The proton's and neutron's large deviation from two is evidence that they are not elementary particles.

also a touchstone for its possible compositeness.

Moreover, as measurements of the electron's g-factor became more precise, it also turned into experimental fact that $g \neq 2$. Instead, it is larger than two by an amount on the order of 10^{-3}. This anomalous deviation—for any given particle—is conventionally parameterized by

$$a \equiv \frac{g-2}{2} \tag{1.5}$$

which is referred to as the particle's *anomalous magnetic moment*. Similar to g's relation to the classical prediction, a quantifies the deviation from the Dirac equation's prediction. This tension with theory helped lead us to another breakthrough with the advent of quantum electrodynamics (QED)—a renormalizable, relativistic quantum field theory [5].

In this new framework that supplants the Dirac equation, the electron's non-zero anomalous magnetic moment arises from its spin-1/2 field, ψ, interacting with the electromagnetic field, as mathematically described by the Lagrangian

$$\mathcal{L} = \bar{\psi}\left(i\slashed{\partial} - m\right)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} - e\bar{\psi}\gamma^{\mu}A_{\mu}\psi \tag{1.6}$$

where $F^{\mu\nu} = \partial^{\mu}A^{\nu} - \partial^{\nu}A^{\mu}$ is the electromagnetic field tensor [6]. However, in four spacetime dimensions, no exact solution to Eq. 1.6—or any interacting field theory—is known. Accordingly, the electron–photon interaction is treated as a perturbative series in the fine structure constant, $\alpha = e^2/4\pi$, which characterizes their coupling strength. Each higher power of α corresponds to

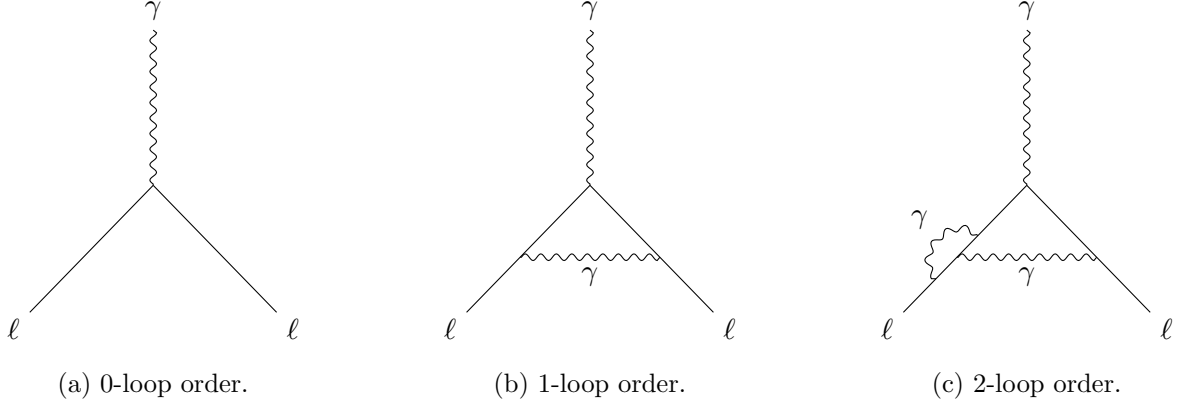(a) 0-loop order.   (b) 1-loop order.   (c) 2-loop order.

Figure 1.1: Feynman diagrams representing the lowest loop-order contributions to a_ℓ of a lepton $\ell \in \{e, \mu, \tau\}$. The (a) 0-loop and (b) 1-loop order diagrams contribute 0 and $\alpha/2\pi$, respectively. The (c) 2-loop diagram is one of nine diagrams whose contribution is mass dependent.

another closed loop containing virtual—i.e., intermediate—particles, whose contribution can be calculated by applying the Feynman rules [6]. A few examples of irreducible Feynman diagrams contributing to the lowest loop orders are shown in Fig. 1.1.

For a given diagram, the anomalous magnetic moment can be formally calculated by the matrix element describing the interaction. The most general matrix element—respecting Lorentz, charge, and parity symmetries—can be parameterized by

$$= (-ie)\bar{u}(p_2) \left[\gamma^\mu \cdot F_1(q^2) + \frac{i\sigma^{\mu\nu}q_\nu}{2m} \cdot F_2(q^2) \right] u(p_1) \qquad (1.7)$$

where $F_1(q^2)$ and $F_2(q^2)$ are independent form factors that encapsulate the properties of the particular interaction, $u(p)$ denotes the Dirac spinor with four-momentum p, and $\sigma^{\mu\nu} = (i/2)[\gamma^\mu, \gamma^\nu]$ [6]. Note that the form factors depend only on q^2, where $q = p_2 - p_1$ is the photon's momentum resulting in vacuum polarization. At leading, 0-loop order—often referred to as tree level—it can be shown that $F_1(q^2) = 1$ and $F_2(q^2) = 0$ [6].

For higher orders, the theory predicts that, in the limit of a static vector potential, i.e., $q^2 \to 0$,

the first form factor vanishes while the second can be shown to be

$$F_2(0) = a, \tag{1.8}$$

solely defining the anomalous magnetic moment [6]. The experimental fact that $a \neq 0$, therefore, must come from orders beyond tree level in the perturbation series. This can be intuited by recognizing that the second term in Eq. 1.7 would correspond at tree level to an effective interaction term in the Lagrangian of the form

$$\delta\mathcal{L}^{\mathrm{AMM}} \propto \bar{\psi}\sigma^{\mu\nu}F_{\mu\nu}\psi \tag{1.9}$$

which has a mass dimension of five [10]. In a renormalizable theory's Lagrangian, though, only operators with mass dimension less than five can appear. This means that at tree level the operator in Eq. 1.9 must be absent, i.e., $F_2(q^2) = 0$. The anomalous magnetic moment, therefore, can only enter the theory at a higher order through virtual loops. Accordingly, a can then be equivalently understood as the renormalization of the QED interaction vertex.

An extremely important consequence of $g = 2$ at tree level is that it allows the anomaly, a, to be *unambiguously* predicted. Absent a tree-level operator of the form in Eq. 1.9, a is not a free parameter of the theory and, thus, cannot be independently adjusted. Instead, it is defined by a relationship among the physical quantities, e.g., particle masses and coupling constants. Each renormalizable theory, therefore, is likely to have a unique prediction, making the measurement of a an excellent means to test the limitations of QED. This has lead to a decades-long endeavor to calculate the magnetic moment's radiative corrections to an increasingly higher order to match or exceed the contemporaneous measurements' precision of a.

The leading order, 1-loop contribution to a arises from only one Feynman diagram, illustrated in Fig. 1.1b, where one virtual photon is emitted and re-absorbed by the electron. First evaluated

by J. Schwinger, this dominant contribution is given by

$$a^{\text{Schwinger}} = \frac{\alpha}{2\pi} \approx 0.00116 \tag{1.10}$$

which is $\mathcal{O}(\alpha/\pi)$ [5]. Further, as Eq. 1.10 solely depends on α, the 1-loop contribution to a is the same for both the electron (a_e) and muon (a_μ). At higher orders, the contributions become mass-dependent, which lead to small differences between a_e and a_μ. For either lepton, though, the higher orders constitute less than 0.2 % of the entire QED contribution to a [10].

As the measurements of a_e and a_μ continued to improve, deviations from the theoretical prediction for both arose on the order of 10^{-12} and 10^{-8}, respectively. By this time, however, other theoretical developments of the strong and weak force had already predicted new particles, which were later experimentally discovered. In general, *any* particle coupling to leptons or photons can also be virtually exchanged at higher orders, thereby slightly contributing to a. A virtual quark–antiquark loop, for instance, can occur inside the virtual photon loop in Fig. 1.1b, accounting for most of the discrepancy between QED and experiment [10].

At present, the theory of QED is subsumed by the Standard Model of particle physics, which encapsulates our best mathematical understanding of *all* known elementary particles and their interactions. Testing its validity, this more-comprehensive theory's prediction for a_e is consistent with the latest measurement, agreeing to 11 decimal places [11]. However, in comparison to the electron's, the muon's anomalous magnet moment is far more sensitive to physics interactions at high mass scales. It can be shown, for example, that the contribution from a heavy virtual particle of mass Λ to a_ℓ—for a lepton $\ell \in \{e, \mu, \tau\}$—is given by

$$a_\ell \propto \frac{m_\ell^2}{\Lambda^2} \tag{1.11}$$

to leading order in the lepton's mass, m_ℓ [10]. This arises from a helicity flip of the lepton, which yields $F_2 \propto m_\ell$. The contribution to a_μ relative to a_e is, thus, enhanced by $(m_\mu/m_e)^2 \approx 4 \times 10^4$.

Contribution	Theoretical prediction $\times 10^{-10}$	Fractional variance		Reference
a_μ^{QED}	$11\,658\,471.8971(75)$	0.00045	%	[11, 12]
a_μ^{HVP}	$684.7(2.4)$	46	%	[13]
a_μ^{EW}	$15.36(10)$	0.080	%	[14]
a_μ^{HLBL}	$9.8(2.6)$	54	%	[13]

Table 1.2: The a_μ-contribution for each of the four types of interactions, summing in Eq. 1.12 to the value predicted by the Standard Model of particle physics. While QED contributes the most to the theory's prediction, the hadronic terms dominate the theory's uncertainty.

This makes a_μ's value a deeper probe of the theory beyond QED and of any, as yet unknown, particles beyond the Standard Model. In that vein, a_τ should be an even deeper probe, but, as seen in Tab. 1.1, the tau is too short lived to perform a precise measurement of its anomalous magnetic moment. For these reasons, the measurement of a_μ provides an exceptional, overall test of the current theory and will be the focus of this dissertation.

1.2 Theoretical Prediction

In the Standard Model of particle physics, the muon's non-zero anomalous magnetic moment originates from the renormalization of the QED vertex. The inclusion of radiative corrections, where virtual particles are emitted and re-absorbed, sum to give a well-defined prediction for a_μ. These contributions to the a_μ-prediction, a_μ^{SM}, are typically parameterized as

$$a_\mu^{\mathrm{SM}} = a_\mu^{\mathrm{QED}} + a_\mu^{\mathrm{HVP}} + a_\mu^{\mathrm{EW}} + a_\mu^{\mathrm{HLBL}} \tag{1.12}$$

where $a_\mu^{\mathrm{HVP}} + a_\mu^{\mathrm{HLBL}}$ make up the hadronic contribution. Here, the a_μ-contributions are, in order, from quantum electrodynamics, hadronic vacuum polarization, electroweak, and hadronic light-by-light interactions. A recent prediction of each contribution is summarized in Tab. 1.2. Although QED contributes nearly 99.994 % of the value of a_μ^{SM}, it is also the best understood. In fact, in comparison to the other contributions, the uncertainty from the QED prediction adds a negligible

Measurement	Deduced value of α^{-1}	Relative uncertainty	Reference
a_e	137.035 999 084(51)	3.7×10^{-10}	[15]
h/m_{Rb}	137.035 999 037(91)	6.6×10^{-10}	[16]
h/m_{Cs}	137.035 999 046(27)	2.0×10^{-10}	[17]

Table 1.3: Recent measurements of physical quantities from which the fine structure constant, α, can be deduced. The value of a_e was measured using a one-electron quantum cyclotron, while the values of h/m_{Rb} and h/m_{Cs} were measured from the recoil velocity of a ^{87}Rb and ^{133}Cs atom, respectively, when it absorbs a photon.

fraction of the total uncertainty on a_μ^{SM}, $\delta a_\mu^{\mathrm{SM}}$. Instead, $\delta a_\mu^{\mathrm{SM}}$ is dominated by the hadronic contribution's prediction, spurring many ongoing theoretical initiatives to reduce its uncertainties. In this section, the current state of each of the four contributions will in turn be presented.

1.2.1 Quantum Electrodynamics

The a_μ-contribution from quantum electrodynamics, a_μ^{QED}, comes from Feynman diagrams including leptons and photons alone. It is expressed as a perturbation expansion in α/π:

$$a_\mu^{\mathrm{QED}} = \sum_{n=1}^{\infty} a_\mu^{(2n)} \left(\frac{\alpha}{\pi}\right)^n \tag{1.13}$$

where the order n in α/π corresponds to the number of closed loops in the diagrams contributing to it. To leading order—i.e., $n = 1$—the only contributing Feynman diagram yields $a_\mu^{(2)} = 1/2$, as previously found in Eq. 1.10. At next-to-leading order—i.e., $n = 2$—nine diagrams contribute, of which one is shown in Fig. 1.1c with $\ell = \mu$. For each successive order, as $\alpha/\pi \approx 0.0023 \ll 1$, the contribution to a_μ^{QED} will be increasingly smaller. This means that, by summing over enough orders, the QED prediction can ideally be calculated to arbitrary precision.

Beyond 1-loop order, however, the Feynman diagrams can have a virtual electron or tau loop, making their contributions dependent on m_μ/m_e or m_μ/m_τ, respectively [12]. As they are not predicted by the theory, these lepton-mass ratios must be obtained from experiments, introduc-

Order n in α/π	$a_\mu^{(2n)} \times 10^{-10}$	Relative uncertainty	Fraction of a_μ^{QED}	
1	11 614 097.331 8(77)	6.6×10^{-10}	99.6	%
2	41 321.762 91(90)	2.2×10^{-8}	0.354	%
3	3 014.190 248(41)	1.4×10^{-8}	0.0259	%
4	38.100 8(19)	5.0×10^{-5}	0.000327	%
5	0.509 38(70)	1.4×10^{-3}	0.00000437	%

Table 1.4: The coefficients up to 5-loop order of the perturbation series in Eq. 1.13 for the a_μ contributions from quantum electrodynamics. The values are taken from [12].

ing a small measurement uncertainty and limiting the QED prediction's precision. Furthermore, these ratios enhance the $a_\mu^{(2n)}$-coefficients for $n \geq 2$ and thereby slow the convergence rate of the perturbation series in Eq. 1.13.

Another required input from experiment is a precise value of the fine structure constant, $\alpha = e^2/4\pi$. This can be deduced from the extremely precise measurement of a_e since—as discussed in Sec. 1.1—it is less sensitive to physics beyond the Standard Model [15]. However, the theoretical calculation of a_e is highly correlated with that of a_μ, making this α-value undesirable. Instead, α can be independently deduced from an atom of mass M according to

$$\alpha^2 = \frac{2R_\infty}{c} \cdot \frac{M}{m_e} \cdot \frac{h}{M} \tag{1.14}$$

where R_∞ is the Rydberg constant. Here, the precision of α is limited by a measurement of h/M, which has recently been found for the ^{87}Rb and ^{133}Cs atom [16, 17]. As shown in Tab. 1.3, the uncertainty on their deduced values of α is now on the same order of magnitude as that from a_e, limiting the QED prediction's relative precision to $\mathcal{O}(10^{-10})$.

With such inputs, a_μ^{QED} has been most recently computed through 5-loop order by Aoyama et $al.$ [11, 12]. Using the value of α from the ^{87}Rb measurement, the contribution from each of the five orders is given in Tab. 1.4, with the total contribution reported to be

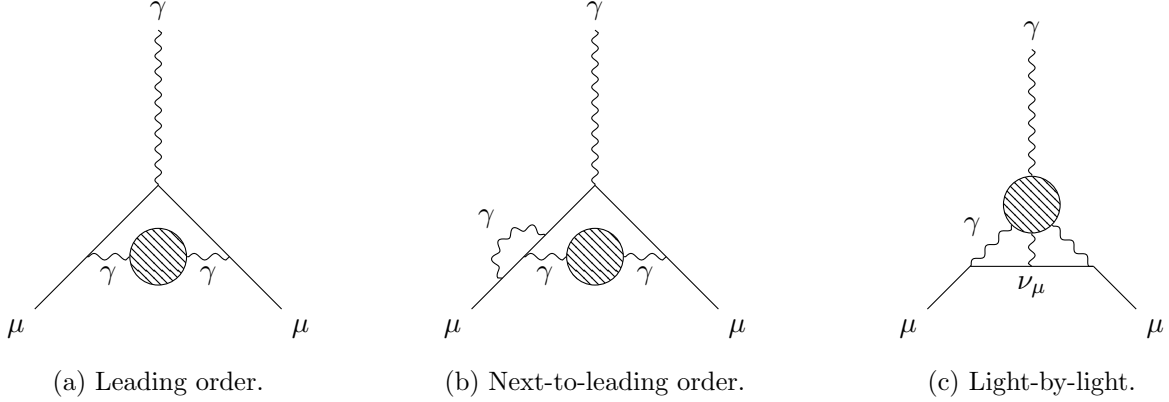$$a_\mu^{\text{QED}}(\alpha(\text{Rb})) = 11 658 471.897 1(7)(17)(6)(72) \times 10^{-10} \tag{1.15}$$

(a) Leading order.   (b) Next-to-leading order.   (c) Light-by-light.

Figure 1.2: The contributions at (a) leading and (b) next-to-leading order to a_μ^{HVP} from hadronic interactions, depicted by the shaded circles. Note that the next-to-leading order diagram shown is only one of many such diagrams with three virtual loops. The higher-order hadronic contribution from (c) light-by-light scattering is considered separately, forming a_μ^{HLBL}.

[11]. Here, from left to right, the assigned uncertainties are from the lepton-mass ratios, numerical error in computing 4-loop diagrams, numerical error in computing 5-loop diagrams, and the value of α—which dominates the total uncertainty. The numerical errors are primarily from statistical fluctuations in the numerical techniques employed to evaluate the mass-dependent, 4- and 5-loop Feynman diagrams. Importantly, the 4-loop order's contribution has been independently checked by analytical and semi-analytical calculation [18, 19]. The a_μ-contribution from quantum electrodynamics, therefore, is very well understood and presently limited only by the measurement of α from atomic experiments—not by the truncated series in Eq. 1.13.

1.2.2 Hadronic Vacuum Polarization

The a_μ-contribution from hadronic vacuum polarization (HVP), a_μ^{HVP}, comes from Feynman diagrams involving quarks but not $W^\pm$, Z^0, and H^0 bosons (which are included in the electroweak term). Analogously to a_μ^{QED}, a_μ^{HVP} may be expressed as a power-series expansion in the strong interaction's coupling constant, $\alpha_s(s)$. At low energies of $E = \sqrt{s}$, however, α_s becomes larger than one, causing the series to diverge. Such a perturbative quantum chromodynamics (pQCD)

approach, therefore, is only reliable above around 2 GeV and away from quark-flavor thresholds where α_s is also enhanced [10]. This makes the HVP contribution difficult to calculate, requiring a semi-phenomenological approach as well.

The leading order (LO) contribution to a_μ—depicted in Fig. 1.2a—consists of a virtual hadron loop inside a virtual photon loop. Fortunately, these loops can be identified with the photon self-energy function, giving rise to a dispersion relation [10]. The imaginary part of the self-energy in that relation can then be recast in terms of the total cross-section of e^+e^- annihilation to hadrons using the optical theorem:

$$2\mathrm{Im}\left[\;\rightsquigarrow\!\!\bigcirc\!\!\rightsquigarrow\;\right] = \int \mathrm{d}\Pi \left| \rightsquigarrow\!\!\bigcirc\!\!\vdots \right|^2 \tag{1.16}$$

where $\mathrm{d}\Pi$ is the differential phase space [6]. More precisely, to leading order, the HVP contribution to a_μ can be obtained from experimental data according to

$$a_\mu^{\mathrm{HVP,LO}} = \frac{m_\mu^2}{12\pi^3} \int_{m_{\pi^0}^2}^{\infty} \frac{\hat{K}(s)}{s} \cdot \sigma(e^+e^- \to \mathrm{hadrons})\,\mathrm{d}s \tag{1.17}$$

where $\hat{K}(s)$ is a slowly varying, positive function and σ is the (bare) cross-section [10]. For some ranges of s, though, pQCD can be accurately used in place of experimental data. Although, as the cross-section roughly follows $1/s$, the lower energies—where pQCD breaks down—contribute more to a_μ. In fact, the $\rho \to \pi^+\pi^-$ resonance around 770 MeV alone accounts for close to 73 % of $a_\mu^{\mathrm{HVP,LO}}$, and, for all channels, the energy region below 1.8 GeV makes up over 89 % [13].

At low energies, therefore, the available experimental data must be carefully combined. This is a non-trivial procedure, which must take into account the many differences among the wide range of colliders, detectors, and analysis methods. Typically, a χ^2-function is formed and minimized to find the best-fit cross-section in energy bins [13]. The primary difficulty here is to understand the energy-dependent, correlated uncertainties among each channel's data, while implementing them

into the χ^2-function without biasing the result. It should also be noted that some channels' data from different experiments—such as from BaBar and KLOE in the $\pi^+\pi^-$ channel—are in tension with each other by as much as a few standard deviations [13]. For such channels, the choice of how to treat their correlations can cause noticeable differences in their predicted a_μ-contribution—as seen in Tab. 1.5 between two independent analyses recently performed. The total contribution for $a_\mu^{\mathrm{HVP,LO}}$, however, tends to be in very good agreement among the analyses [13].

Nonetheless, the higher-order contribution to a_μ^{HVP} must also be evaluated for a precise prediction of the hadronic effects. At the next-to-leading order (NLO), these Feynman diagrams exhibit either another virtual photon—as depicted in Fig. 1.2b—or another virtual quark or lepton loop inserted in the photon line. Note that the so-called light-by-light type of diagrams are separately added in a_μ^{HLBL}. In the other types, the hadronic loops are also identified with the photon self-energy, allowing the NLO contributions to a_μ^{HVP} to be similarly evaluated from e^+e^--annihilation data. At the next-to-next-to-leading order (NNLO), the same dispersive techniques can once again be used. However, the total NNLO contribution is found to be $1.24(1) \times 10^{-10}$—approximately half of the LO contribution's uncertainty from the experimental data [21]. As such, further orders can be safely neglected until the cross-sections' precision improves. The total value of a_μ^{HVP} reported by Keshavarzi *et al.* through the NNLO contribution is given in Tab. 1.2.

Alternatively, the HVP contribution to a_μ can be predicted from first principles by employing lattice QCD and QED. In this technique, the theory is simulated on a finite lattice of points in Euclidean spacetime, recovering the continuum theory by taking the lattice's spacing to zero and volume to infinity [22]. In principle, this allows the theory to be solved to arbitrarily high precision by using a smaller spacing and a larger volume of the lattice. However, in practice, computing resources typically become the limiting factor. So far, the precision of lattice calculations of a_μ^{HVP} are above 2 % but are expected to eventually reach 0.2 %—almost half of that from recent dispersive

Channel(s)	Keshavarzi *et al.* [13]	Davier *et al.* [20]	Difference
$\pi^+\pi^-$	503.7 ± 2.0	507.1 ± 2.6	-3.4
$\pi^+\pi^-\pi^0$	47.70 ± 0.89	46.20 ± 1.45	1.50
$\pi^+\pi^-\pi^+\pi^-$	13.99 ± 0.19	13.68 ± 0.31	0.31
$\pi^+\pi^-\pi^0\pi^0$	18.15 ± 0.74	18.03 ± 0.54	0.12
K^+K^-	23.00 ± 0.22	22.81 ± 0.41	0.19
$K_S^0 K_L^0$	13.04 ± 0.19	12.82 ± 0.24	0.22
$1.8 \leq \sqrt{s} \leq 3.7$ GeV	34.54 ± 0.56	33.45 ± 0.65	1.09
Total	693.3 ± 2.5	693.1 ± 3.4	0.2

Table 1.5: The contribution from various channels to the leading order value of a_μ^{HVP}, determined by two independent analyses. While there are some significant differences in individual channels, the analyses' total estimates are in good agreement.

results [23, 13]. The dispersive results, though, will also improve over time as more experimental data are released. These improvements are important because, at present, the uncertainty on a_μ^{HVP} accounts for almost half of the total uncertainty on the theoretical prediction of a_μ.

1.2.3 Electroweak

The a_μ-contribution from electroweak (EW) theory, a_μ^{EW}, comes from Feynman diagrams with the virtual exchange of $W^\pm$, Z^0, and H^0 bosons. As with the QED contribution, a_μ^{EW} can be evaluated perturbatively in the coupling, i.e., in the number of virtual loops inside the diagrams. The leading-order, 1-loop Feynman diagrams, which are reminiscent of that from QED, are illustrated in Fig. 1.3. In particular, the analytical expression corresponding to the charged-current and neutral-current diagrams can be shown to be

$$a_\mu^{\mathrm{EW,LO}} = \frac{G_F m_\mu^2}{8\sqrt{2}\pi^2} \left[\frac{10}{3} + \frac{(1 - 4\sin^2\theta_W)^2 - 5}{3} + \mathcal{O}\left(\frac{m_\mu^2}{m_W^2} \right) \right] \qquad (1.18)$$

where $G_F \propto 1/m_W^2$ is the Fermi coupling constant and $\sin^2\theta_W = 1 - (m_W/m_Z)^2 \approx 0.22$ is the weak mixing parameter [24]. Here, the first and second term in square brackets are from the W- and Z-boson's contribution, respectively. The EW contribution is, thus, suppressed by a factor of

$(m_\mu/m_W)^2$, causing it to be around 0.0001 % of the QED contribution. The contribution from a Higgs-boson exchange, on the other hand, can be shown to be further suppressed by a factor of $(m_\mu/m_H)^2$, making it negligible at the order of 10^{-14} [10].

At the 2-loop order, the Feynman diagrams can have either an electromagnetic correction or a hadronic loop included. These diagrams' contributions are all suppressed by a factor of α/π with respect to Eq. 1.18 [10]. Although, in some diagrams, the Higgs-boson exchange's extra suppression by $(m_\mu/m_H)^2$ can now be avoided by having H^0 couple to a particle other than the muon. At the same time, those diagrams containing a closed fermion loop are also greatly enhanced by $\log(m_Z/m_f)$, where m_f is the fermion's mass. These types of enhancement help make the 2-loop order's contribution to a_μ^{EW} non-negligible.

The perturbative approach to evaluating the diagrams with light quarks breaks down due to low-energy, strong interactions—as in the HVP contribution. To account for them, calculations using increasingly refined techniques—including constituent quark models, effective field theory, and so-called large-N_c QCD—have been carried out [10, and references therein]. Currently, the variation among the estimates is small enough to give confidence in the 2-loop contribution at $\mathcal{O}(10^{-11})$. In comparison, the leading diagrams at 3-loop order have been estimated to be on the order of 10^{-12}, making them negligible until the hadronic effects are better understood [14]. The total value of a_μ^{EW} in Tab. 1.2 reported by Gnendiger *et al.* through 2-loop order accounts for the neglected 3-loop order diagrams in the uncertainty.

1.2.4 *Hadronic Light-by-Light Scattering*

The a_μ-contribution from hadronic light-by-light scattering, a_μ^{HLBL}, is given by (high order) Feynman diagrams where four virtual photons couple to a hadronic, intermediate state—as pictured in Fig. 1.2c. Historically, these diagrams have been the most difficult a_μ-contribution to compute
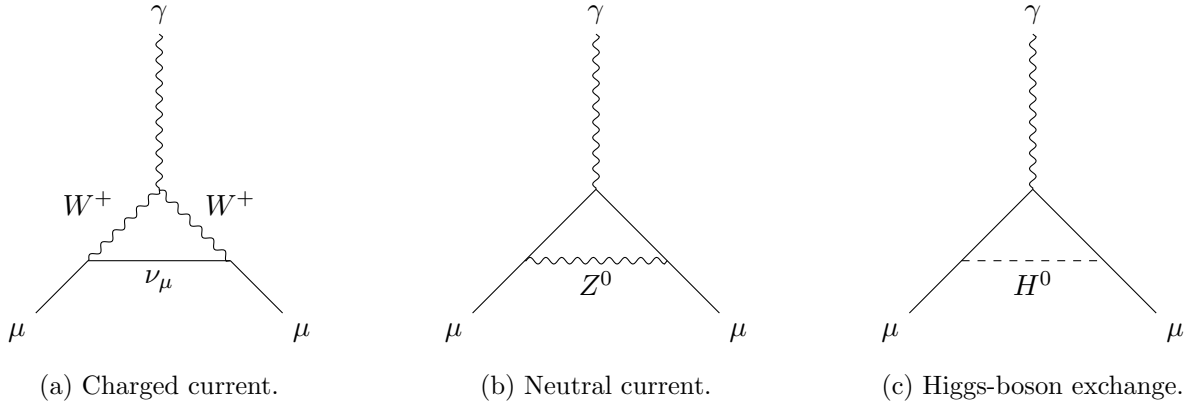
(a) Charged current.   (b) Neutral current.   (c) Higgs-boson exchange.

Figure 1.3: The 1-loop order contributions to a_μ^{EW} from electroweak interactions that involve the W^+, Z^0, or H^0 boson. Due to the small coupling of the Higgs boson to the muon, only the (a) charged and (b) neutral current processes contribute significantly to leading order.

and have been calculated using many low-energy, hadronic models, which have large uncertainties [10]. This approach was needed because, for this type of diagram, no simple analog exists to the dispersive method in Eq. 1.17.

In recent years, however, a more complex relation for a_μ^{HLBL} has been derived where every possible intermediate state is treated separately [25, and references therein]. For states with up to two particles, the connection to the relevant experimental observable is known, and arguments have been made as to which ones should contribute the most. The leading contribution, for example, is well known to be from a π^0 coupling to two photons at two separate vertices and has been recently evaluated from available data [26]. In the not too distant future, a complete estimate—including many more intermediate-state contributions—is expected to be realized and, hopefully, with an uncertainty competitive to those from model-dependent estimates [25].

In addition, there are two major, ongoing efforts to calculate a_μ^{HLBL} from first principles on the lattice [27]. Broadly speaking, they differ by whether QED is or treated analytically or simulated on the lattice as is done for QCD. So far, encouraging progress has been made with no major disagreements with the model estimates. In the near future, the lattice estimates' uncertainties are

likely to reach close to 10 %—about half of the widely accepted model estimate [13].

A typical value for the leading HLBL contribution, using a model estimate, is given in Tab. 1.2. Its uncertainty is currently over half of the total uncertainty on a_μ^{SM}, making the aforementioned alternative approaches essential. At the next-to-leading order, the dominant HLBL contribution is expected to come from an additional lepton loop inserted into the photon or muon line [28]. However, using a hadronic model, these diagrams' contributions have been found to be suppressed to $\mathcal{O}(10^{-11})$—an order of magnitude smaller than the model estimate's uncertainty. Nonetheless, adding their contribution to all of the other contributions in this section, the total theoretical prediction of a_μ^{SM}—as recently reported by Keshavarzi *et al.*—becomes

$$a_\mu^{\text{SM}} = 11\,659\,182.0(3.6) \times 10^{-10} \tag{1.19}$$

with a relative uncertainty of 310 ppb [13]. Therefore, to fully test the Standard Model of particle physics, an equally precise value must be measured and compared to the theoretical prediction. If a statistically significant discrepancy is found between the two values, a_μ-contributions from new physics will have been discovered—hopefully leading to the next theoretical breakthrough.

1.3 Previous Measurements

As seen in Sec. 1.1, increasingly precise measurements of the muon's g-factor—or, equivalently, anomalous magnetic moment—have played an important role in developing the current theory of the Standard Model. In essence, to make g experimentally accessible, they were all performed by placing muons in a static, external magnetic field, $\boldsymbol{B}$. According to the Hamiltonian in Eq. 1.2, when placed at rest, their magnetic moments obey the time-evolution operator

$$\mathcal{U}(t, 0) = \exp\left[-iHt\right] = \exp\left[-iS_z\left(-g\frac{eB}{2m} \cdot t\right)\right] \tag{1.20}$$

from an initial time $t = 0$ [4]. Here, the magnetic field is chosen to be along the z-axis where S_z and B are the corresponding components. This expression is identical to that corresponding to a rotation about the z-axis by an angle of $\omega_s \cdot t$, where

$$\omega_s = -g\frac{eB}{2m} \tag{1.21}$$

is known as the "spin precession" frequency [4]. Therefore, each muon's spin vector will precess around the magnetic field's vector at a rate that is directly proportional to its g-factor, returning to its initial direction after $t = 2\pi/\omega_s$. Thanks to parity violation of the weak interaction, the spin direction in turn is correlated with the momentum of the electron or positron emitted when the muon inevitably decays. All together, such phenomena have allowed g_μ to be experimentally measured for over half a century by simply measuring the decay products' direction/energy over time. The exact experimental technique applied, however, has had to be continually refined in order to match or exceed the latest theoretical prediction's precision.

Early g_μ-measurements, starting in 1957, were performed at the Nevis cyclotron by using the above technique [29]. In particular, they counted the number of decay positrons over 35 MeV from stopped muons in an external magnetic field. They then deduced ω_s—and, by Eq. 1.21, g_μ—by changing the spin-precession rate by adjusting the magnetic-field strength. Their final result, though, on the muon anomaly, a_μ, was limited to a precision of around 6.6 %—only being able to test QED's prediction at 1-loop order [29]. Fortunately, it was soon realized that a_μ could in fact be directly measured by letting the muon have kinetic energy.

In this case, the Lorentz force will cause muons to undergo cyclotron motion with a frequency of $\omega_c = -eB/m$. Therefore, their spin vector *relative* to their momentum vector will precess at $\omega_c - \omega_s \propto a_\mu$. This frequency difference, ω_a—typically referred to as the anomalous precession frequency—enables a_μ to be measured to much higher precision and gave rise to a decades-long succession of so-called Muon $g - 2$ experiments trying to improve the technique, starting with a

Experiment	Year(s)	$a_\mu \times 10^{10}$	Relative uncertainty	Reference
CERN I	1961	11 450 000(220 000)	4 300 000 ppb	[30]
CERN II	1962–1968	11 661 600(3 100)	270 000 ppb	[31]
CERN III	1975–1976	11 659 240(850)	7 300 ppb	[32]
BNL	1997–2001	11 659 208.0(6.3)	540 ppb	[33]

Table 1.6: A summary of the previous measurements of the muon's anomalous magnetic moment, a_μ. The years refer to that in which data was collected. The latest measurement, performed at BNL, is the first one sensitive to EW and HLBL contributions to a_μ.

series of three experiments at CERN.

The first experiment, known as CERN I, used the CERN synchro-cyclotron to inject polarized, 83-MeV muons into a narrow magnetic trap [30]. Due to a field gradient, the muons would drift in a helical motion to the other end of the apparatus where they were extracted from the field and stopped in a target. The forward/backward asymmetry of the decay positrons was then used to extract a_μ to 0.4 %, as shown in Tab. 1.6 [30]. This allowed for the first test of QED's prediction at 2-loop order. Higher orders could not be tested because of the measurement's large statistical uncertainty—primarily limited by the muon's short lifetime at rest.

To overcome this limitation, the next iteration of the experiment—CERN II—used the newly built CERN proton synchrotron to create higher-energy muons, trapping them inside a magnetic storage ring [31]. More specifically, high-energy protons hit a pion-production target in the ring, resulting in a small fraction of muons being stored from the pions' decay. As the muons circled around the ring, a radial gradient in the magnetic field provided vertical focusing to keep them trapped. The muons then eventually decayed into lower-energy positrons, which were detected by counters located around half of the ring. This improved technique allowed a_μ to be measured to 270 ppm—now testing QED's prediction at 3-loop order [31].

To test the hadronic contributions to a_μ, the third experiment—CERN III—further refined the method in two key ways. First, the target was moved outside of the ring, allowing pions to be in-
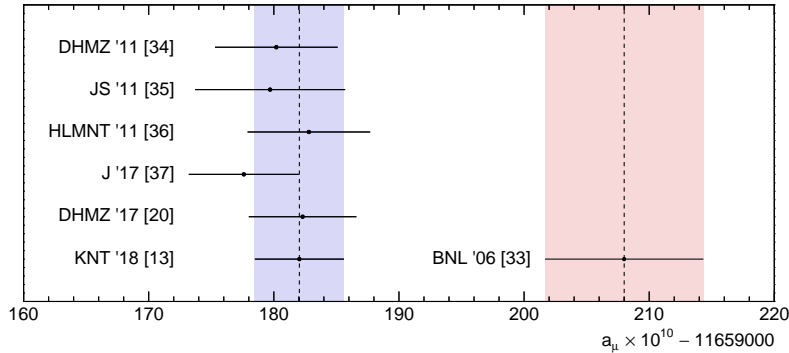
Figure 1.4: A comparison between the final a_μ-measurement from BNL and six recent theoretical predictions, labeled by their authors' initials. These predictions were found using a variety of approaches and input data for some of the contributions to a_μ. They all differ by between three and four standard deviations from the measured a_μ-value.

jected directly into the storage ring. This allowed for counting detectors to be installed around the remainder of the ring. Second, an electric field was introduced to provide the muons' vertical focusing. This allowed the magnetic-field strength—required to deduce a_μ—to be more precisely measured. Together, these improvements resulted in an a_μ-value measured to 7 ppm, confirming the predicted HVP contribution to a_μ at the five standard deviation level [32].

After about a decade had passed, the theoretical prediction's uncertainty had reduced enough such that the EW contribution could also be tested. This motivated the most recently completed experiment, conducted at Brookhaven National Laboratory (BNL) using the alternating gradient synchrotron [33]. To achieve the sub-ppm precision desired, the BNL experiment further refined the storage-ring method in many ways to reduce both the statistical and systematic uncertainties of the CERN III measurement. As an example, to increase the storage efficiency and lower the background, the ring was redesigned for muons to be directly injected. Instead of discrete resistive magnets, this design also used a continuous superconducting magnet in order to improve the magnetic field's uniformity around the ring. Together with improved positron detectors, the BNL experiment was able to measure a_μ to 540 ppb—nearly a 14-fold improvement as seen in Tab. 1.6 [33]. The a_μ-value found, as a result, dominates the current world average and confirms the EW

contribution to almost three standard deviations [22].

However, as seen in Sec. 1.2, the theoretical prediction for a_μ can reasonably differ based on one's choice of analytical/numerical approaches and experimental data, such as the value of α in Tab. 1.3. Accordingly, the BNL measurement must be judged against the range of possible theoretical predictions. As shown in Fig. 1.4, recent a_μ-predictions differ from the measured value, a_μ^{measured}, by between three and four standard deviations. Specifically, the most precise prediction to date—reported by Keshavarzi *et al.*—indicates a difference of

$$a_\mu^{\text{measured}} - a_\mu^{\text{SM}} = 26.0(7.3) \times 10^{-10} \tag{1.22}$$

which is non-zero with a significance of about 3.6 standard deviations [13]. This corresponds to a 1 in 3000 chance of simply being due to a statistical fluctuation and is almost twice the EW contribution predicted. Ever since its publication, the BNL result and theoretical prediction have both been intensely scrutinized, looking for any mistakes or unaccounted for biases. The muon's electric dipole moment, for example, was also measured, which would have reduced the measured value of ω_a [10]. Nonetheless, the tension between the measured and predicted a_μ-value has not only persisted to today but has become more statistically significant.

This lasting tension has spurred a community-wide initiative to both measure and predict the value of a_μ even more precisely—with the aim of settling whether the discrepancy is real. As part of that effort, the latest-generation Muon $g-2$ experiment has been developed at Fermi National Accelerator Laboratory (FNAL) with a goal of measuring a_μ to 140 ppb—almost a four-fold improvement [38]. This current experiment, as detailed in Ch. 2, continues to refine the storage-ring technique, employing the higher-intensity muon source at FNAL and upgraded instrumentation. This dissertation will present the ω_a-measurement from this experiment's first year of running in 2018, which is expected to yield an a_μ-value slightly more precise than the BNL result. If the central value of a_μ remains unchanged, the discrepancy with theory would exceed five standard

deviations—strongly suggesting new physics beyond the Standard Model.

1.4 New Physics Interpretations

The Standard Model of particle physics is already known to be incomplete. For instance, it does not explain the experimental fact of neutrino oscillation or the baryon asymmetry in the universe [39]. To explain such effects, many extensions have been proposed to the Standard Model, often incorporating new particles. Although their direct production has yet to be observed at colliders, these hypothesized particles should perturb the value of a_μ by different amounts. For example, with a 1-loop order proportionality factor of α/π, Eq. 1.11 suggests that a shift of $\mathcal{O}(10^{-9})$ may be explained by some new particle with a mass around 100 GeV [10]. The current a_μ-discrepancy in Eq. 1.22, therefore, can be interpreted as evidence of radiative corrections from such an as yet undiscovered particle. In this way, the measured a_μ-value has been (and will continue to be) able to tightly constrain the parameter space of a wide range of "new physics" models, some of which will be briefly surveyed in this section.

1.4.1 Heavy New Physics

In general, the possible extensions to the Standard Model can be divided by the mass scale of the new physics, Λ. For so-called heavy new physics above the electroweak scale—$\Lambda \gtrsim 100$ GeV—the contribution to a_μ is expected to take the form $\mathcal{C} \cdot (m_\mu/\Lambda)^2$, where $\mathcal{C}$ is highly model dependent. In some such scenarios, for example, the $SU(3) \times SU(2) \times U(1)$ local symmetry of the Standard Model is extended to a larger gauge group, postulating new charged (W') and neutral (Z') gauge bosons. Their contributions to a_μ will, thus, enter at the 1-loop order—analogous to W and Z in the Feynman diagrams in Fig. 1.3—with $\mathcal{C} = \mathcal{O}(\alpha/\pi)$ [40]. This loop-suppression factor leads to a_μ-contributions typically on the order of 10^{-11}—much too small to explain the (current) a_μ-

discrepancy in Eq. 1.22. The same conclusion also tends to hold in many other Standard Model extensions, including little Higgs and some extra-dimension models [41, 42].

There are some models, however, where the loop-suppressed $\mathcal{C}$ can be enhanced. A well-studied example is supersymmetry (SUSY), which is a compelling extension since it can also explain the hierarchy problem. In the simplest SUSY model, a new symmetry is postulated, requiring twice as many particles to exist. That is, for every particle in the Standard Model, there is an associated, so-called superpartner particle whose spin differs by a half-integer. For example, the neutrino's superpartner is the sneutrino, which contributes to a_μ via a sneutrino–chargino loop as depicted in Fig. 1.5. Numerically, the superpartners' total contribution from leading 1- and 2-loop order diagrams can be shown to be approximately

$$a_\mu^{\text{SUSY}} \approx \pm 13 \times 10^{-10} \cdot \tan\beta \left[\frac{100 \text{ GeV}}{m_{\text{SUSY}}} \right]^2 \tag{1.23}$$

where, for simplicity, the superpartners' masses are all taken to be m_{SUSY} [40]. Here, the sign of the a_μ-contribution is determined by that of the μ-parameter in the supersymmetric Lagrangian. Therefore, if the discrepancy in Eq. 1.22 is confirmed, its sign alone will effectively rule out half of all SUSY models [40]. This is especially significant because, otherwise, the μ-parameter's sign would be difficult to deduce solely from collider experiments [43].

In addition, the SUSY contribution's magnitude in Eq. 1.23 can be substantially enhanced by $\tan\beta = v_2/v_1$—the ratio of the two Higgs fields' vacuum expectation values. For example, some SUSY models predict this enhancement to $\mathcal{C}$ to be on the order of $m_t/m_b \approx 40$—the ratio of the top and bottom quarks' masses [40]. As seen in Fig. 1.5, such models can account for the current a_μ-discrepancy if the superpartners' masses are around 500 GeV. Conversely, the a_μ-discrepancy can be viewed as constraining $\tan\beta$ if m_{SUSY} is one day measured. This is noteworthy because $\tan\beta$ can only be poorly determined by collider experiments alone. More quantitatively, for some models, the uncertainty of $\tan\beta$ is expected to be halved if future m_{SUSY}-measurements at the
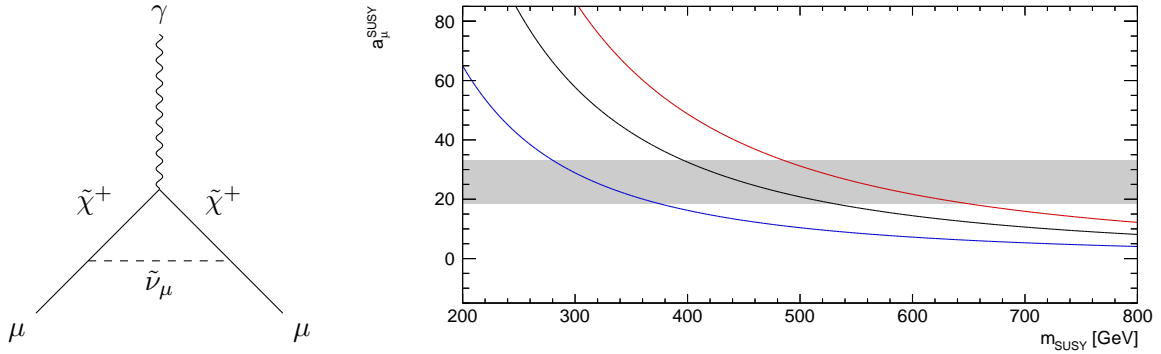
Figure 1.5: Left: The leading Feynman diagram contributing to a_μ^{SUSY}, involving the hypothesized sneutrino, $\tilde{\nu}_\mu$, and chargino, $\tilde{\chi}^+$. Right: The positive value of a_μ^{SUSY} predicted in Eq. 1.23 as a function of the degenerate superpartners' mass, m_{SUSY}, with a $\tan\beta$ of 20 (blue), 40 (black), and 60 (red). The shaded region denotes the $\pm 1\sigma$ bounds of the current a_μ-discrepancy in Eq. 1.22.

Large Hadron Collider (LHC) are combined with a_μ even at current precision [43].

Of course, a more detailed analysis incorporating non-degenerate masses is needed for a_μ^{SUSY}-predictions in specific SUSY models. This has in fact been performed for numerous benchmark models, some of which cannot be distinguished at the LHC alone [44]. Fortunately, these models predict very different a_μ-contributions, including different signs, and, thus, a precisely known value of a_μ can help resolve possible LHC degeneracies in the future. This is but one example of how the a_μ-measurement complements other ongoing searches for new physics. It also illustrates how diverse experiments will likely be needed to pin down the right new physics model.

1.4.2 Light New Physics

As discussed above, barring enhancement factors like $\tan\beta$ in SUSY, it is hard to account for the currently observed a_μ-discrepancy with $\mathcal{C} = \mathcal{O}(\alpha/\pi)$ unless the mass scale is very small—i.e., $\Lambda \lesssim 1$ GeV. Such light new physics scenarios as with the W' and Z', however, have already been mostly excluded by searches at colliders [40]. Nonetheless, possible exceptions exist where the new particles interact with the muon but, due to very small couplings, are virtually hidden from the searches. A minimal example is to extend the Standard Model's gauge group by a hidden
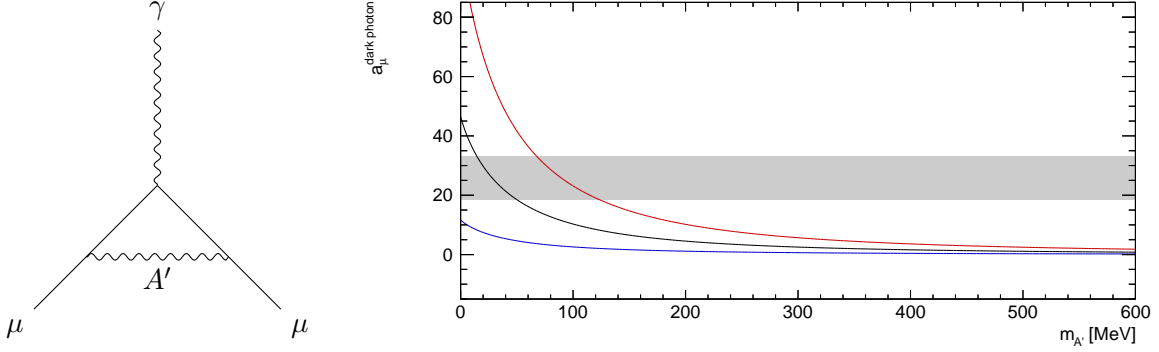
Figure 1.6: Left: The leading Feynman diagram contributing to $a_\mu^{\text{dark photon}}$, involving the hypothesized dark photon, A'. Right: The value of $a_\mu^{\text{dark photon}}$ predicted in Eq. 1.24 as a function of the dark photon's mass, $m_{A'}$, with an ε of 1×10^{-3} (blue), 2×10^{-3} (black), and 3×10^{-3} (red). The shaded region denotes the $\pm 1\sigma$ bounds of the current a_μ-discrepancy in Eq. 1.22.

$U(1)$ gauge symmetry. The corresponding gauge boson, A', can then interact (very weakly) with the muon through a new kinetic-mixing term, $-e\varepsilon\bar{\psi}\gamma^\mu A'_\mu\psi$, in the Lagrangian, where ε is a free parameter that suppresses the coupling strength [45].

This scenario with A'—typically called the dark photon—has become popular in recent years, originally being motivated by cosmology [45]. It can contribute to a_μ at 1-loop order, as depicted in Fig. 1.6, which is analytically given by

$$a_\mu^{\text{dark photon}} \approx \frac{\alpha}{2\pi}\varepsilon^2 \int_0^1 \frac{2z(1-z)^2}{(1-z)^2 + z(m_{A'}/m_\mu)^2}\, \mathrm{d}z \qquad (1.24)$$

where $m_{A'}$ is the dark photon's mass [22]. As shown in Fig. 1.6, the current a_μ-discrepancy in Eq. 1.22 can be resolved with certain values of ε and $m_{A'}$. For example, for $\varepsilon \approx 3 \times 10^{-3}$, the measured a_μ can be explained by a dark-photon mass around 100 MeV. By now, though, current experimental constraints have excluded the entire ε–$m_{A'}$ parameter space that is consistent with the BNL measurement [46]. However, more exotic variants of dark-photon models have yet to be ruled out as the explanation for the a_μ-discrepancy [47].

In any case, the dark-photon scenario serves as yet another example of how measuring a_μ has helped guide our search for physics beyond the Standard Model. Some other light new physics

models proposed include a dark Higgs and axion-like particles [48, 49]. More generally, as seen in this chapter, a_μ plays a key role in our theoretical understanding of the fundamental interactions. It has already excluded many new physics models, but many exciting options still exist to explain a_μ. However, even if the final a_μ measured by the latest experiment at FNAL is consistent with the Standard Model's prediction, it will nonetheless set a stringent constraint on the new physics models needed to explain already known gaps in the current theory.

CHAPTER 2

EXPERIMENTAL METHOD

The Muon $g - 2$ experiment at Fermi National Accelerator Laboratory (FNAL) aims to measure the muon's anomalous magnetic moment, a_μ, to 140 parts per billion using data taken over four years from 2018 to 2021. This four-fold improvement over the latest measurement will hopefully confirm or deny the current tension with the theoretically predicted a_μ-value (Sec. 1.3). To realize such precision, this latest-generation experiment builds on the foundation and lessons learned from the last one, exploiting the more-intense muon production available at FNAL to reduce the statistical uncertainty. A complementary reduction of the systematic uncertainty—as exemplified across this dissertation—is achieved through upgraded instrumentation and analysis procedures. This chapter will broadly review the experimental method behind the a_μ-measurement performed on data collected during the first running year—designated as *Run 1*—in 2018.

2.1   Experiment Principles

The latest-generation Muon $g - 2$ experiment at FNAL employs the same underlying technique used by the preceding one at Brookhaven National Laboratory (BNL). In general, a pulsed beam of polarized muons is injected into a uniform, external magnetic field, $\boldsymbol{B}$, orthogonal to its velocity. Each muon—with charge e and velocity $\boldsymbol{v}$—will experience a Lorentz force, $e(\boldsymbol{v} \times \boldsymbol{B})$, and travel in a circular orbit as in a cyclotron. Simultaneously, its spin direction will rotate in the magnetic field in the same plane. In the absence of other external fields, the *cyclotron*—i.e., orbital—and *spin* rotational frequencies are, respectively, given by

$$\boldsymbol{\omega}_c = -\frac{e\boldsymbol{B}}{m_\mu \gamma}, \tag{2.1}$$

$$\boldsymbol{\omega}_s = -g_\mu \frac{e\boldsymbol{B}}{2m_\mu} - (1 - \gamma)\frac{e\boldsymbol{B}}{m_\mu \gamma} \tag{2.2}$$

where m_μ is the muon mass; $\gamma = 1/\sqrt{1 - \beta^2}$, with $\beta = v/c$, is the Lorentz factor; and g_μ is the muon's g-factor [1]. The second term in Eq. 2.2 is a relativistic correction, known as Thomas precession, which accounts for the muon's rotating reference frame.

From these expressions, it is clear that, if $g_\mu = 2$, the cyclotron and spin frequencies would be equal. Each muon's velocity and spin, in other words, would always remain at a fixed angle with respect to each other. However, radiative corrections, as explained in Ch. 1, sightly increase g_μ, which in turn causes the spin to rotate slightly faster than the velocity. This frequency difference between $\boldsymbol{\omega}_s$ and $\boldsymbol{\omega}_c$ is formally given by

$$\boldsymbol{\omega}_a \equiv \boldsymbol{\omega}_s - \boldsymbol{\omega}_c = -a_\mu \frac{e\boldsymbol{B}}{m_\mu} \tag{2.3}$$

which is known as the *anomalous precession frequency*. As it is directly proportional to the muon anomalous magnetic moment, $a_\mu \equiv (g_\mu - 2)/2$, ω_a is the observable upon which the Muon $g - 2$ experiment is based. Experimentally speaking, it is vital that ω_a depends on a_μ as opposed to g_μ since, in comparison, this enables an a_μ-measurement several orders of magnitude more precise. It is also crucial that ω_a is independent of γ, which would be challenging to measure concurrently to high precision. As a result, ω_a can be found very precisely by effectively measuring the muon ensemble's mean spin direction as a function of time spent in the magnetic field. An overview of the specific techniques used will be presented in Sec. 2.5.

Nonetheless, the overall precision of a_μ will also be limited by that of the other quantities in Eq. 2.3. The currently accepted value of e/m, in particular, has an uncertainty of around 26 ppb [8]. To reduce this precision floor, a_μ in Eq. 2.3 can be rewritten as

$$a_\mu = \frac{g_e}{2} \frac{\mu_p}{\mu_e} \frac{m_\mu}{m_e} \frac{\omega_a}{\omega_p} \tag{2.4}$$

where g_e is the electron's g-factor, μ_p/μ_e is the proton-to-electron magnetic moment ratio, m_μ/m_e is the muon-to-electron mass ratio, and ω_p is the Larmor frequency of the free proton [38]. Using

Quantity	Experimental method	Uncertainty	Reference
$g_e/2$	Quantum cyclotron spectroscopy	0.00028 ppb	[15]
μ_p/μ_e	Hydrogen spectroscopy	3.0 ppb	[8]
m_μ/m_e	Muonium spectroscopy	22 ppb	[8]

Table 2.1: The precision of the three quantities, which, together with ω_a and ω_p measured in the Muon $g-2$ experiment, determine a_μ in Eq. 2.4. Their quadrature sum forms a precision floor, which is much less than the experiment's target precision of 140 ppb.

the uncertainties from previous experiments in Tab. 2.1, the rearrangement in Eq. 2.4 leads to a slightly lower precision floor of almost 22 ppb—much less than the experiment's target precision of 140 ppb. This limit is reached if the magnetic field—encapsulated in ω_p—is perfectly known. In practice, the Muon $g-2$ experiment aims to measure ω_p to 70 ppb. To achieve that precision, the magnetic-field region where the muons travel is instrumented with an extensive network of probes based on pulsed nuclear magnetic resonance [38]. The field's strength and spatial non-uniformity were thereby measured throughout Run 1, enabling ω_p to be weighted by the muons' spatial distribution. Details of the ω_p-measurement can be found in [38].

2.2 Muon Production and Injection

As aforementioned, the experimental method consists of storing longitudinally polarized muons in a uniform magnetic field, after which the mean spin direction is determined. The first stage of the experiment, therefore, is the production of such muons—most readily obtained from pion decay. As will be expounded shortly, this is accomplished with FNAL's vast accelerator complex, which delivers muons to the storage region in short temporal bunches—known as *muon fills*. Illustrated in Fig. 2.1, these muons enter the magnetic field through a specially designed *inflector* magnet, radially offset from the design storage orbit. To "kick" them onto the nominal path, an electrostatic magnet—called the *kicker*—is quickly pulsed as the muons pass by on their first turn. With the
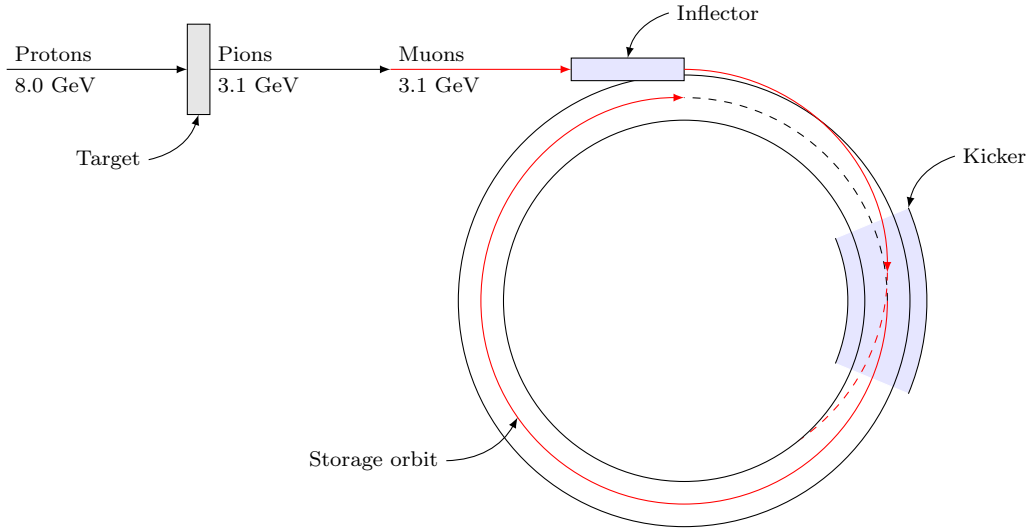
Figure 2.1: A schematic of the production and injection of muons into the experiment's storage region. In temporal bunches, accelerated protons impinge on a high-Z target, producing copious charged pions. By selecting on the highest-energy muons from their decay, bunches of polarized muons are injected into the storage region's magnetic field through an inflector magnet. A transverse impulse is then applied by a kicker magnet on their first turn to trap them on the design storage orbit. Adapted from [10].

muons trapped, the frequency at which their spin vector advances relative to their velocity—i.e., ω_a—can then be deduced from their decay products, as explained in Sec. 2.5.

Each muon fill originates from a bunched beam of protons that was accelerated to 8 GeV via FNAL's linac and booster synchrotron [50, and references therein]. Through FNAL's recycler synchrotron, the proton bunch is further manipulated into tighter—about 120-ns long—bunches of approximately 10^{12} protons. This re-bunching is required to end up with a tolerable flux rate at the positron detectors used to measure ω_a (Sec. 2.5). Further, the tighter bunches are separated by at least 10 ms, following the cyclical time structure in Fig. 2.2 with an average rate of about 11.4 Hz. This bunch separation also provides enough time to record the detectors' data. One at a time, each bunch is then directed to a pion-production target, composed of a nickel–iron alloy core. This high-Z material causes charged pions—the lightest hadron in the Standard Model—to be copiously produced in the proton–nucleus interactions that result.
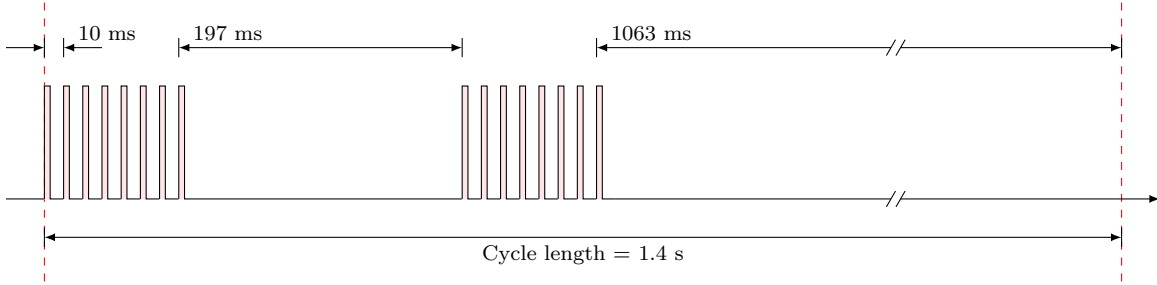
Figure 2.2: The time structure of the proton bunches in FNAL's recycler synchrotron, which is the same as that of the muons entering the experiment's storage region. Each cycle consists of 16 bunches—separated by at least 10 ms—averaging to about 11.4 Hz.

All positively charged particles produced within ± 10 % of 3.11 GeV are subsequently collected using a lithium lens and a bending magnet. They are next steered along a long beamline designed to capture 3.094-GeV muons from the weak decay of charged pions in flight:

$$\pi^+ \to \mu^+ + \nu_\mu \tag{2.5}$$

which is the main decay mode with a ≥ 0.9998 branching fraction [22]. The parity-violating nature of the weak interaction, together with the momentum selection, also results in the captured muons being highly polarized. To conserve momentum, the muon and neutrino must be emitted with equal and opposite momenta in the pion's rest frame. As the pion has a spin of zero and neutrinos—in the minimal Standard Model theory—are always left-handed, to conserve angular momentum, the muon must further have the same handedness as the neutrino. In the laboratory frame, this correlates the muon spin and momentum direction. Accordingly, by capturing only the highest-momentum, positive muons in the beamline, about a 95 % polarized muon beam is formed—with spin and momentum anti-parallel [38].

By the end of this beamline, roughly 80 % of the pions have decayed into muons. Aside from the remaining pions, however, the beam will also contain some proton and deuteron secondaries that passed the momentum selection. To separate the muons from these heavier particles, the beam is then circulated around FNAL's delivery ring [50]. After four turns—when essentially all of the
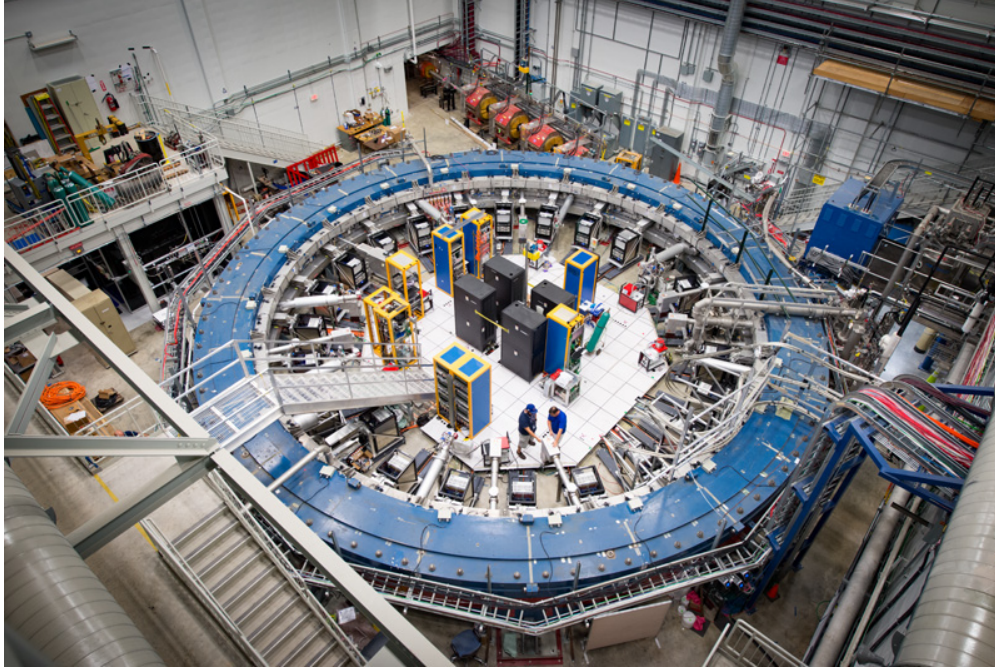
Figure 2.3: A photograph of the Muon $g-2$ experiment's storage ring at FNAL in 2017. The top of the storage ring's magnet—painted blue—is visible along with the 24 calorimeter stations positioned uniformly around the ring's inner circumference. The four quadrupole magnets leading up to the inflector magnet is also visible towards the top of the photograph. Reprinted with permission. Copyright 2017 by Fermi National Accelerator Laboratory.

remaining pions have decayed—the protons, for example, will have fallen behind the muon bunch by about 200 ns on account of their smaller Lorentz factor. This allows a pulsed electromagnetic magnet to remove almost all of these heavier particles. The high intensity and high purity of the resulting muon bunch is one of the major improvements over the previous experiment.

The muon bunch is then extracted and, using four quadrupole magnets, focused through a small aperture to finally enter the experiment's storage region. This requires, however, traveling through the region's fringe magnetic field. Therefore, to enter the uniform field without being deflected, the muons are injected through the so-called inflector magnet [38]. Specially designed for the previous experiment, this magnet primarily uses superconducting coils to cancel the fringe field and to prevent its own magnetic field from perturbing the storage region.

The muons then exit the infector on a tangent with the storage region, circulating in the uni-

form magnetic field at the cyclotron frequency. If they exit on the design storage orbit, however, they would collide with the inflector magnet after one turn. Therefore, the inflector is instead displaced 77 mm radially outward. As the muons first cross the nominal orbit—after circulating by $\pi/2$—they then receive a transverse impulse from the so-called kicker [38]. This magnet is comprised of three sections of curved plates, azimuthally spanning $\pi/4$ in total. Each section's plates carry current in opposite directions, powered by Blumlein triaxial transmission lines that can be independently timed to when the muons exit the inflector. The kicker field generated is ideally maintained at the required 200–280 G over the muon bunch's temporal width and returns to zero before the bunch re-enters the aperture on its second turn. However, the inevitable spread of angles exiting the inflector, for example, means that an ideal kick cannot be given to every muon. This leads to dynamics of the stored muons as they circulate, which are important for the ω_a-measurement and will be further discussed in Sec. 2.4.

2.3   Muon Storage

The storage region's uniform magnetic field is produced by the same *storage-ring* magnet used by the previous experiment, pictured in Fig. 2.3 [51]. It employs a super-ferric design, where a steel magnet is excited by three superconducting coils connected in series. Illustrated in Fig. 2.4, it is intended to create a highly uniform, 1.451-T dipole field in a toroidal region where the muons circulate. With a major radius of 7.112 m and a minor radius of 45 cm, this region is able to trap muons that are within roughly 0.15 % of 3.094 GeV. Of the muons injected in each bunch, only about 1–2 % will end up being stored given this small range.

The superconducting coils, in particular, are cooled to 5.0 K and operated at 5200 A. Two of them are positioned radially inward, while the third—with twice the number of windings in the opposite direction—is located radially outward. For indirect cooling and mechanical support, each
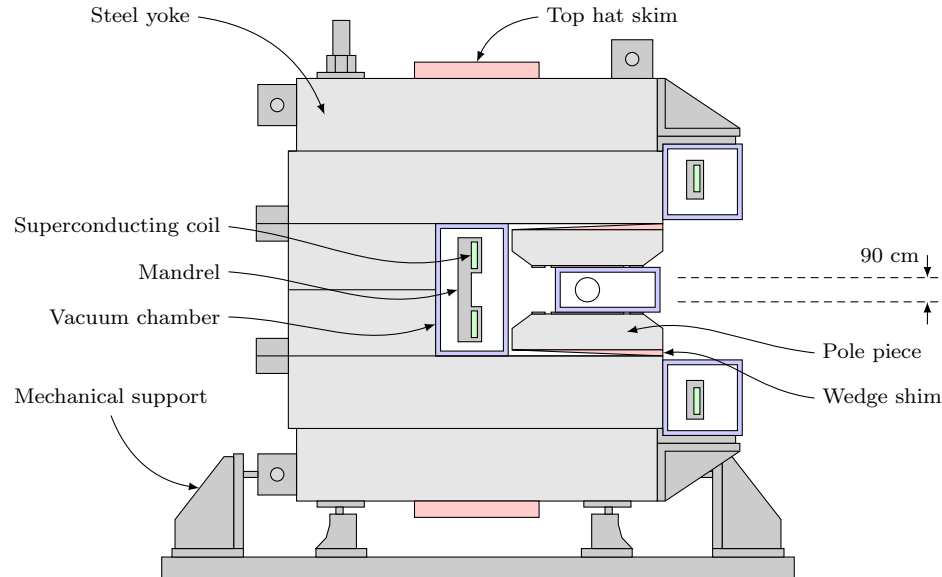
Figure 2.4: A cross-sectional view of the storage ring magnet. As described in the text, the steel yoke is excited by superconducting coils (green) to produce a highly uniform, 1.451-T dipole field between the pole pieces. The muons travel into the page inside the circular aperture shown. The field can be adjusted locally by the position of various steel skims (red). Adapted from [38].

coil resides in an aluminum mandrel, which in turn hangs inside a cryostat using low thermal-conductivity straps for thermal insulation. When the coils are cooled, they can contract radially inward by as much as 22 mm while, when they are powered, the Lorentz force can push them outward by as much as 4 mm [52]. Therefore, to maintain the field's uniformity between thermal and/or power cycles, the radial positions had to be repeatable to ± 0.5 mm [38]. To that end, when the storage ring was first cooled down and powered up at FNAL in 2015, the outer coil's position was directly measured at different azimuthal locations through small viewports in the cryostat. The results shown in Fig. 2.5 are in excellent agreement with those made during the previous experiment at BNL. From that time onward, the coil positions were continuously monitored using indirect measurements such as from strain gauges on their straps.

The magnetic field produced is shaped by the surrounding steel to return the flux across the magnet's gap uniformly. The storage region's field, therefore, largely depends on the quality and
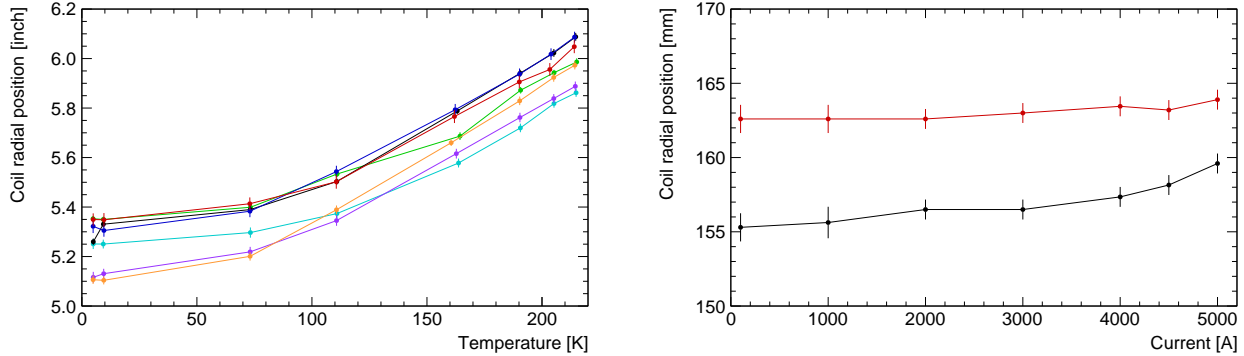
Figure 2.5: The radial position of the outer-radius coil relative to a reference radius when it was first cooled down (left) and powered up (right) at FNAL in 2015 [52]. Each color denotes one of seven azimuthal positions around the storage ring where a viewpoint through the cryostat is located. These positions agree well with those taken by the previous experiment at BNL.

geometry of the ferric material used. More specifically, the field aberrations primarily depend on that of the pole pieces above and below the storage region. These pieces are, therefore, partially decoupled by a 1.5-cm air gap from the bulk of the magnet, which was made from a lower-quality steel. This gap also provided a region where steel wedges could be inserted by varying amounts, allowing the field to be passively shimmed around the ring. Further, a collection of ferric material on the pole faces and at the top and bottom of the magnet provided more than 1000 knobs for azimuthally localized changes. Even further, 100 concentric coils on a printed circuit board were attached to the pole pieces' surface, enabling a final refinement. All together, the magnetic field's azimuthal variations were reduced down to $\mathcal{O}(100\,\mathrm{ppm})$—as shown in Fig. 2.6—with the azimuthally averaged variation spatially across the storage region held to $\mathcal{O}(1\,\mathrm{ppm})$.

Moreover, the muons travel within an aluminum vacuum chamber, which is situated between the pole pieces. The vacuum maintained inside it helps prevent the muons from scattering off gas molecules and potentially being displaced outside of the storage region. It also reduces the likelihood of a spontaneous discharge when the kicker plates—placed inside the vacuum chamber—hold a high voltage. They are further scalloped, as illustrated in Fig. 2.7, so that positrons from the muons' decay can travel unobstructed to the detectors described in Sec. 2.5.2.
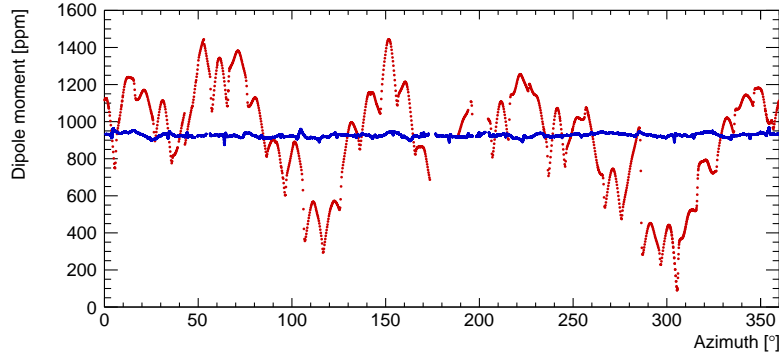
Figure 2.6: The dipole magnetic field in azimuth before (red) and after (blue) skim adjustments during 2015 and 2016. The peak-to-peak field variations around the storage ring were reduced roughly from 1400 to 100 ppm. Adapted from [53].

Also inside the vacuum chamber are electrostatic quadrupole plates that are used to confine the muons vertically. When the muons exit the inflector, they will inevitably have a spread of both horizontal and vertical angles. As the aforementioned magnetic field is vertically aligned, it will supply a restoring force horizontally—but not vertically. Any muon with a vertical component to its velocity will simply spiral upwards or downwards, eventually leaving the storage region. A simple solution is to tilt the pole faces into a wedge shape to provide classic weak focusing, but that would spoil the experiment's uniform field.

Instead, a similar effect is created using four flat plates to form an electrostatic quadrupole field. Shortly before the muons are injected, the plates above and below the storage region are held at a positive voltage whereas the side plates are at a negative voltage. The resulting field creates a restoring force vertically. However, the quadrupole plates only cover 43 % of the storage region azimuthally so the vertical spread of muons will be slightly larger at some locations around the ring. At the end of every fill's measurement period, the voltages are then reset to zero to release any trapped, low-energy particles. While such an electric field is needed, it also alters the muon's anomalous precession frequency. Fortunately—as will be seen in Sec. 2.7—the ω_a-correction that arises is almost eliminated for a muon momentum of 3.094 GeV.
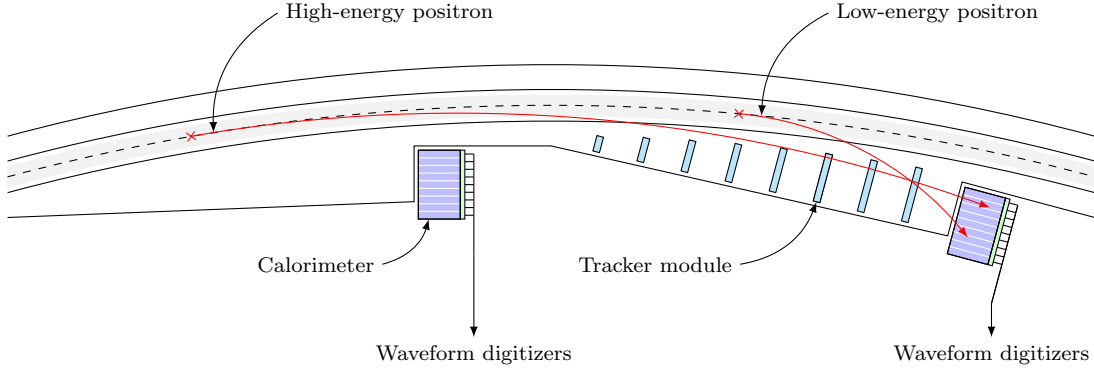
Figure 2.7: A top-down view of a section of the vacuum chamber, depicting two possible decay positron trajectories (red). The dashed line delineates the muons' design storage orbit, which is at the center of the shaded muon-storage region. The solid lines on either side of the nominal orbit show the position of the electrostatic quadrupole plates.

2.4 Muon Dynamics

In an ideal experiment, the muons would be injected and kicked onto the design storage orbit, so that their momentum only has a longitudinal component. Every muon would then exhibit perfect cyclotron motion in the horizontal mid-plane—with no need for radial or vertical focusing. In practice, however, muons will enter the storage ring with a range of angles with respect to the design orbit and will be subsequently kicked to a range of radii. As will be seen in this section, such displacements will cause the muons to oscillate about different equilibrium orbits, yielding dynamics of the ensemble that will be important for the ω_a-measurement in Ch. 5.

To predict the muons' motion in the storage fields, first consider one at the design momentum of $p_0 = m_\mu \gamma v_0 = 3.094$ GeV, where v_0 is its velocity. Further suppose that the magnetic dipole field and the electric quadrupole field are ideal and independent of azimuth as given by

$$\boldsymbol{B} = B_0 \cdot \hat{y}, \tag{2.6}$$

$$\boldsymbol{E} = kx \cdot \hat{x} + ky \cdot \hat{y} \tag{2.7}$$

respectively, where the quadrupole field's strength is characterized by its gradient, $k = \partial E_x / \partial x =$

$\partial E_y / \partial y$. In this coordinate system, $x = r - R_0$ and y are the radial and vertical displacements, respectively, from the center of the quadrupole field at radius R_0.

In such fields, the positive muon will experience both the centripetal and Lorentz force, which, by Newton's second law, can be expressed as

$$m_\mu \gamma \cdot \ddot{x} = \frac{m_\mu \gamma v_0^2}{R_0 + x} - e v_0 B_0 + ekx \approx \frac{m_\mu \gamma v_0^2}{R_0} \left(1 - \frac{x}{R_0} \right) - e v_0 B_0 + ekx, \tag{2.8}$$

$$m_\mu \gamma \cdot \ddot{y} = -eky \tag{2.9}$$

where the binomial approximation is applied to the centripetal force's term. This is reasonable since the storage region's aperture is only 45 mm in radius as compared to $R_0 = 7.112$ m, making $x/R_0 < 0.007$. The trivial solution to these ordinary differential equations—$x = 0$ and $y = 0$—simply describes a muon stored perfectly on the design orbit with $p_0 = eR_0 B_0$ in the longitudinal direction. Its cyclotron frequency, ω_c, is, therefore, $eB_0/m_\mu \gamma$. To find a non-trivial solution, it is helpful to rewrite these equations of motion as

$$\ddot{x} = -\omega_c^2 (1 - n) \cdot x, \tag{2.10}$$

$$\ddot{y} = -\omega_c^2 n \cdot y \tag{2.11}$$

where n is defined as the *field index* given by

$$n \equiv \frac{kR_0}{v_0 B_0}. \tag{2.12}$$

In this form, it is evident that a non-zero radial or vertical displacement will result in simple harmonic motion—typically referred to, in this context, as *betatron oscillation*. That is, the ordinary differential equations in Eqs. 2.10 and 2.11 have non-trivial solutions of the form

$$x(t) = A_x \cos \left(\omega_x \cdot t + \phi_x \right), \tag{2.13}$$

$$y(t) = A_y \cos \left(\omega_y \cdot t + \phi_y \right) \tag{2.14}$$

with angular frequencies given by

$$\omega_x = \omega_c \cdot \sqrt{1 - n}, \tag{2.15}$$

$$\omega_y = \omega_c \cdot \sqrt{n}. \tag{2.16}$$

For each muon, the betatron oscillations' amplitude and phase will depend on the exact injection angle and kick it received. From Eq. 2.15, it is also clear that n must be less than one for a restoring force to constrain the muon radially.

Moreover, note that the equilibrium orbit is centered at $x = 0$ in Eq. 2.13. However, for an off-momentum muon—with $p \neq p_0$—this balance between the centripetal and Lorentz forces in Eq. 2.8 will change. By setting $\ddot{x} = 0$ in its equation of motion analogous to Eq. 2.8, such a muon's equilibrium radius with respect to R_0, x_e, can be shown to be

$$x_e \approx \frac{R_0}{1 - n} \cdot \frac{p - p_0}{p_0} \tag{2.17}$$

to leading order, where terms of order $(x_e/R_0)^2$ are neglected. Further, as $\omega_c \approx (R_0 + x_e)v$, the 0.1-% spread in p of off-momentum muons that can be stored leads to a spread in cyclotron periods and, in turn, betatron frequencies. The multiplicative factors in front of ω_c in Eqs. 2.15 and 2.16, though, are already approximations good to only about 1 % because the quadrupole plates do not cover the entire ring [54].

As well, the spread in cyclotron periods causes the injected muon bunch to spread out in azimuth over time, eventually filling the entire storage region. This *debunching* behavior is aided by inevitable, small nonlinearities in the fields and is found to occur with a characteristic time constant of around 30 μs. Additionally, this debunching—together with the betatron frequency's momentum dependence—will cause any coherent betatron motion of the muon ensemble to smear out over time. At early enough times after injection, however, such muons' dynamics will leave an imprint in the number of muon decays detected by the calorimeters that will be introduced in the

next section. The implications for the ω_a-measurement are evaluated in Sec. 5.6.

Muons too far off-momentum will not be stored for long. In total, only about 1–2 % of the muons injected are trapped longer than 30 μs. This loss also complicates the ω_a-measurement. To alleviate such difficulties, muons near the edges of the storable phase space are forcefully removed. At injection, the electrostatic quadrupole plates are charged asymmetrically to manipulate the muons' equilibrium orbit [38]. The muons on the edge will then strike one or more copper collimators, causing it to lose energy and exit the storage ring. These collimators, therefore, define the toroidal storage region. After about 7 μs, the quadrupole plates are finally charged to their nominal voltages—with a time constant of about 5 μs—so that, by 30 μs after injection, the muon ensemble surviving is centered about the design orbit. All together, this *scraping* process mitigates the rate of muons being lost at times past 30 μs. Those that are lost afterward will be characterized and accounted for in Sec. 5.5.

Additionally, perturbations in the magnetic or electric field can cause a muon to be lost from the storage ring through changes to its betatron motion. This effect is enhanced, in particular, if a muon's oscillating orbit encounters the *same* local field perturbation every turn. A resonance occurs, therefore, when a linear combination of the betatron frequencies equals a multiple of the cyclotron period for its momentum, i.e.,

$$i \cdot \omega_x + j \cdot \omega_y = k \cdot \omega_c \tag{2.18}$$

where $i, j, k \in \mathbb{Z}$. When such a condition is met, the muons' spatial distribution can widen until most of them are lost. More importantly, a muon's spin will advance by a different amount relative to its velocity in the local field perturbations. This is, in general, tolerable since the perturbation averaged over azimuth is only of $\mathcal{O}(1 \text{ ppm})$. However, if near such a resonance, the muons will experience a slightly different, average magnetic and/or electric field, which is difficult to predict. To avoid introducing a direct systematic bias to a_μ, n was, in part, specifically chosen to be far from

the lowest-order—i.e., lowest-k—resonance condition in Eq. 2.18.

2.5 Anomalous Precession Frequency Measurement

As the muons trapped in the ring are unstable—with a time-dilated lifetime of around 64.44 μs—
they will all eventually decay, primarily through the three-body decay channel,

$$\mu^+ \rightarrow e^+ + \nu_e + \bar{\nu}_\mu \tag{2.19}$$

which has a branching fraction of very nearly 100 % [22]. As will be discussed in this section, the
emitted positron's observables—direction and energy—are correlated with its parent muon's spin
direction. Therefore, by detecting and characterizing these positrons, the average direction of the
muon ensemble's spin can be measured over time. From that, the anomalous precession frequency,
ω_a, can be extracted, and its precision can be optimized by analyzing only certain subsets of the
decay positrons observed. The specific techniques introduced in this section build the framework
for this dissertation's ω_a-measurement in Ch. 5.

2.5.1 Measurement Principles

Akin to the charged pion's decay in Sec. 2.2, the muon's decay is mediated by the weak interac-
tion and, as a result, exhibits parity violation that manifests as a correlation between the muon's
spin direction and the decay positron's emission direction. Furthermore, the muon's spin at the
time of decay is also correlated with the decay positron's energy. As a concrete example, con-
sider when the positron is emitted in the muon's rest frame with the *maximum* allowed energy,
$E_{\max} \approx m_\mu/2 \approx 53$ MeV. In this extreme case, ν_e and $\bar{\nu}_\mu$ must be emitted collaterally in the oppo-
site direction and, as neutrinos (antineutrinos) are always left-handed (right-handed), their spins
will cancel. To conserve angular momentum, the decay positron then must carry the *same* spin as
the muon when it decayed. Thanks to the correlation from parity violation previously mentioned,

these high-energy decay positrons will also be preferentially emitted parallel to the muon's spin direction. Therefore, aside from their momentum direction, their energy encodes experimentally accessible information on the muon's spin orientation at decay.

More specifically, in the muon's rest frame, the differential probability for a decay positron to be emitted with energy E at an angle θ with respect to the muon's spin direction can be derived from V-A theory [55]. Assuming that $E \gg m_e$, it can be written of the form

$$\frac{\partial^2 P}{\partial E \, \partial \cos \theta} \propto n(E) \left[1 + a(E) \cos \theta \right] \tag{2.20}$$

where, with $y \equiv E/E_{\text{max}}$, the functions of energy are

$$n(E) = 2y^2(3 - 2y), \tag{2.21}$$

$$a(E) = \frac{2y - 1}{3 - 2y}. \tag{2.22}$$

The "asymmetry" function, $a(E)$, encodes the degree to which the positron's momentum direction and muon's spin direction are correlated when the decay positron's energy is between E and $E + \mathrm{d}E$. As shown in Fig. 2.8, $a(E)$ monotonically increases with E, crossing zero at $E = E_{\text{max}}/2$ and reaching one at $E = E_{\text{max}}$. Thus, positrons with $E > E_{\text{max}}/2$ are more likely to be emitted in the direction of the muon's spin. In fact, since the number density, $n(E)$, also monotonically increases with E, the decay positrons are *overall* more likely to be found along the muon's spin direction—i.e., when integrated over E.

The stored muons, however, are highly relativistic in the laboratory frame with $\gamma \approx 29.3$. The decay positrons' range of emission directions will, therefore, be very nearly parallel to the muon's momentum direction. This makes measuring the positrons' average θ over time—to extract ω_a—experimentally inaccessible. The positrons' energy in the laboratory frame, on the other hand, is related to that in the muon's rest frame by

$$E = \gamma E^* \left[1 - \beta (\boldsymbol{p}_e^* \cdot \boldsymbol{p}_\mu) \right] \approx \gamma E^* \left(1 + \cos \alpha \right) \tag{2.23}$$
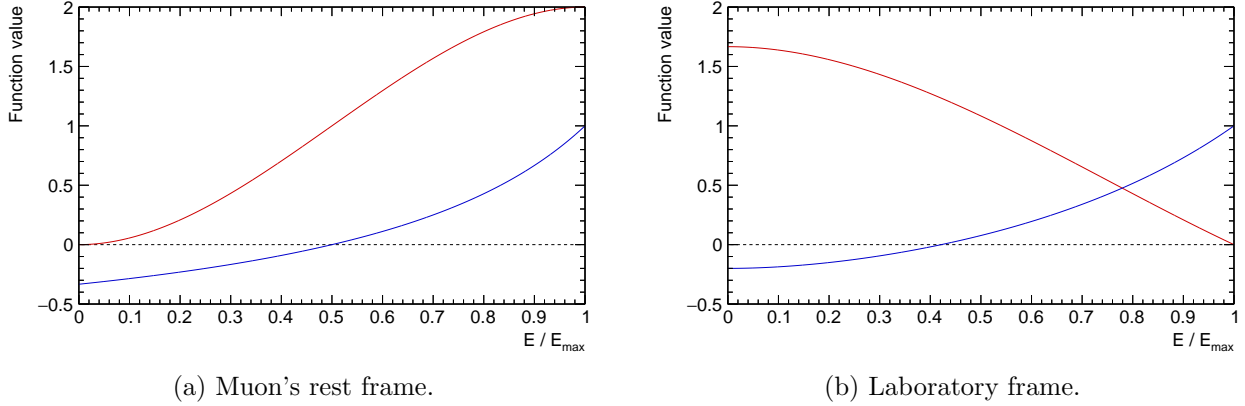
(a) Muon's rest frame.

(b) Laboratory frame.

Figure 2.8: The number density (red) and asymmetry (blue) functions in the (a) muon's rest frame, where $E_{\text{max}} \approx 53$ MeV, and (b) laboratory frame of reference, where $E_{\text{max}} \approx 3.1$ GeV.

neglecting the positron's mass, where E^* and $\boldsymbol{p}_e^*$ are the positron's energy and momentum in the muon's rest frame, respectively, and $\boldsymbol{p}_\mu$ is the muon's momentum in the laboratory frame. From Eq. 2.23, it is evident that the muon's energy is strongly dependent on the emission angle with respect to the muon's momentum direction, α.

This means that high-energy positrons in the laboratory frame must be high-energy positrons in the muon rest frame *and* be emitted in the direction of the muon's momentum—i.e., of the Lorentz boost. Thanks to the muon decay's aforesaid correlations, the latter requirement is satisfied only when the muon's spin is parallel to its momentum. As the muons' spins precess about their momenta in the ring's magnetic field with $\alpha = \omega_a t$, the highest-energy positrons will, therefore, be observed only at certain times, i.e., when they are aligned. Further, in a given time range, the number of these highest-energy positrons observed will then be maximal when all of the stored muons are perfectly polarized at injection.

More generally, at a time t after injection, the energy distribution of the decay positrons observed in the laboratory frame can be shown from Eqs. 2.20–2.23 to be of the form

$$\frac{\partial P}{\partial E} \propto N(E) \left[1 + \mathcal{P} \cdot A(E) \cos(\omega_a t + \phi) \right] \tag{2.24}$$

where $\mathcal{P} \approx 0.95$ and ϕ are the muon ensemble's polarization and α at injection, respectively.
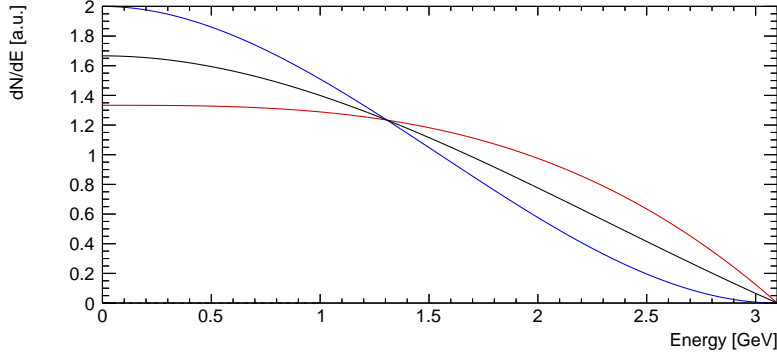
Figure 2.9: The decay positron's energy distribution in the laboratory frame when the muon's spin and momentum are aligned (red), orthogonal (black), and anti-aligned (blue). The probability of observing a higher-energy positron from a muon's decay increases with the alignment. When the stored muons' spins precess relative to their momenta in the magnetic field, the observed energy distribution will then oscillate at ω_a, thereby enabling a measurement of it in Ch. 5.

With $y \equiv E/E_{\mathrm{max}}$ and $E_{\mathrm{max}} \approx 3.1$ GeV, the number density and asymmetry functions are

$$N(E) = \frac{1}{3}(y-1)(4y^2 - 5y - 5), \tag{2.25}$$

$$A(E) = \frac{1 + y - 8y^2}{4y^2 - 5y - 5} \tag{2.26}$$

respectively. Hence, as conveyed in Fig. 2.9, very nearly the positron's *entire* energy distribution also oscillates at ω_a between the extreme cases of the muons' spin and momentum being aligned and anti-aligned. This means that, as exploited in Sec. 2.5.3, positrons with energy less than the highest, E_{max}, also contain information about the muon spin's orientation. Measuring and then fitting a subset of the decay positron's oscillating energy spectrum as each injected muon bunch decays in the ring, therefore, forms the basis for an ω_a-measurement.

2.5.2   Positron Detectors

To conserve momentum, each positron emitted from a muon decay must carry away a momentum less than that of the parent muon. In the magnetic field, the decay positrons will, thus, begin to curl towards the ring's center, with the high-energy ones curling the least. To intercept primarily high-energy positrons—which have the strongest correlation with the muon's spin direction—24

electromagnetic calorimeters are stationed uniformly around the ring's inner circumference. As depicted in Fig. 2.7, every calorimeter is specifically positioned flush with the vacuum chamber's corresponding scalloped section. Illustrated in Fig. 2.10, this configuration can intercept over 80 % of decay positrons above about 2 GeV. The low-energy positrons, in contrast, tend to exit the vacuum chamber between calorimeters and go undetected.

These calorimeters are designed to measure each incident positron's energy and time of arrival. They are constructed from a 6-high × 9-wide array of PbF_2 crystals, each of which is a $25 \times 25 \times 140$ mm^3 rectangular cuboid [56]. When a positron enters a crystal, an electromagnetic shower develops, yielding Čerenkov radiation that propagates towards the back of the crystals. To optically isolate the crystals, they are separated by a layer of matte black, non-reflective Tedlar. Further, the crystal's short radiation length of roughly 9.3 mm allows the shower to be well contained longitudinally along its 14-cm depth. This is important to keep the calorimeter short so as to minimize the number of low-energy positrons entering the side of the calorimeter. In general, the higher the flux at the calorimeter, the harder it is to measure the incident positrons' properties individually. This motivated, in part, the segmentation of the calorimeter to allow for spatial separation of incident positrons close in time.

Each crystal's Čerenkov light is read out by a 16-channel, 12×12 mm^2 silicon photomultiplier (SiPM), mounted to its back face [57]. This solid-state device is composed of an array of discrete avalanche photodiodes—known as pixels—that work as Geiger-mode counters. Their output is an electrical current, which is amplified and shaped by a custom-designed printed circuit board. The resulting voltage signal scales fairly linearly with the light's intensity. This design leads to each positron's shower producing a short pulse—with a width of only about 10 ns—in the SiPMs of crystals near where it entered. Importantly, it can also operate well in the fringe magnetic field where they are located while not perturbing the storage region's field.

As expounded in Ch. 3, these pulses are digitized by custom-designed waveform digitizers, which define each muon fill's measurement period to be around 700 μs—close to 11 time-dilated muon lifetimes of about 64.44 μs. The collected data are then stored to disk via the experiment's data-acquisition system, based on the MIDAS software framework [58]. Afterward, the positrons' pulses can be identified and fit to extract their energies and arrival times in Ch. 4. As will be introduced later in this section, those with high enough energy can then be histogrammed in time and fit to extract ω_a—which will be performed in detail in Ch. 5.

Because the spin-precession signal is time-based, the calorimeter's time resolution must be excellent. In contrast, its energy resolution needs only to be modest. A resolution around 5 % for positrons beyond 2 GeV, for example, provides adequate energy discrimination so as not to significantly wash out their measured asymmetry. To help ensure these specifications are met, the calorimeter design was repeatedly characterized throughout its development using incident electrons at the End Station Test Beam at SLAC [59, 60].

A time-dependent bias in the energy measured, on the other hand, can pose a systematic bias to the ω_a extracted, as explained in Sec. 2.6. To keep every SiPM's response stable over time, a laser-calibration system was developed, which can fire a laser pulse into each crystal on demand to provide a reference signal. As shown in Ch. 4, this system also allowed for the SiPMs' output signals to be time-aligned—a critical step needed to analyze the pulses.

The muons' dynamics predicted in the previous section, moreover, will introduce modulations (not at ω_a) of a high-energy positron's chance of hitting a calorimeter. To avoid a systematic bias to the ω_a-value extracted, such betatron motion must then be well understood. Therefore, as depicted in Fig. 2.7, leading up to two of the calorimeters in the vacuum chamber's scalloped section is a set of eight tracking modules—collectively known as a *tracker* [61]. Each of these modules is composed of layers of straw tubes, providing a non-destructive measurement of a positron's loca-
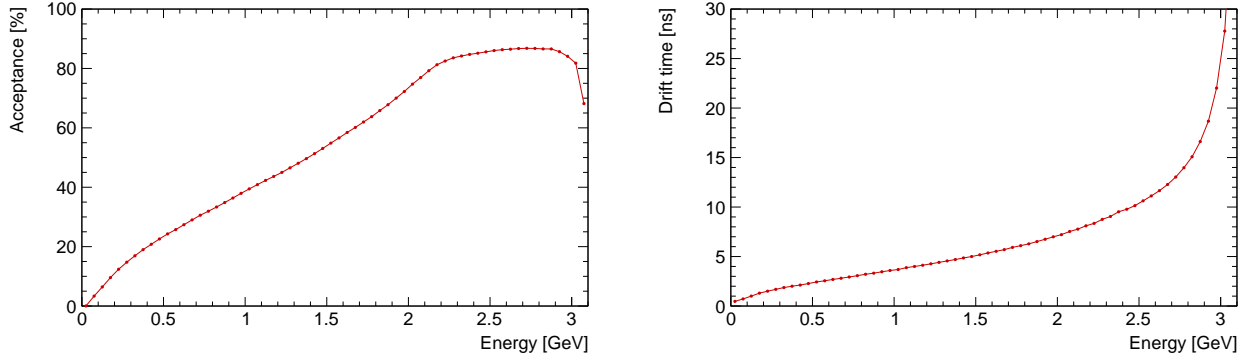
Figure 2.10: The probability for a decay positron to enter a calorimeter (acceptance) and, of those that enter, the average time of flight between the muon decay's vertex and hitting a calorimeter (drift time). These data were found using the `Geant4`-simulation described in Sec. 4.8.

tion. Together, the trajectory of positrons as they travel to the calorimeters can be reconstructed and monitored to infer the muon dynamics. The trackers' results are used in Sec. 5.6 to inform how the muon dynamics should be incorporated into the ω_a-measurement.

2.5.3   Measurement Technique

When the experiment's data-taking period is finished, the objective is to extract ω_a as precisely as possible from the recorded calorimeter signals. From Sec. 2.5.1, the flux of the highest-energy decay positrons detected—near $E_{\max} \approx 3.1$ GeV—will modulate the most at ω_a. A straightforward approach to determine ω_a, therefore, is to reconstruct every decay positron's energy and arrival time from the voltage pulses (Ch. 4). The number of reconstructed positrons above a (non-zero) energy threshold, E_{th}, will then oscillate at ω_a. If the correct time-dependent model is known, these higher-energy positrons can be histogrammed in their arrival time and, in the limit of normal statistical fluctuations, fit for ω_a by minimizing a χ^2-function.

More formally, from Eq. 2.24, the number of positrons observed at a given calorimeter as a function of the muon-decay time is given by

$$N(t) \propto e^{-t/\tau} \int_{E_{\mathrm{th}}}^{E_{\max}} w(E) \frac{\partial P}{\partial E} \, \mathrm{d}E \qquad (2.27)$$

$$\propto e^{-t/\tau} \left[\langle w \rangle + \mathcal{P} \cdot \langle wA \rangle \cos(\omega_a t + \phi) \right] \tag{2.28}$$

where the first factor describes the muons' exponential decay with the time-dilated muon lifetime of $\tau \approx 64.44$ μs. Without loss of generality, $w(E)$ is a weight factor assigned to each positron of energy E, and $\langle f \rangle$ represents the value of f averaged over the number of positrons at each energy above E_{th}. Neglecting muon-dynamics effects (Sec. 2.4), the time histogram of positrons with energy above E_{th} can then be described by

$$N(t) = N_0 \cdot e^{-t/\tau} \cdot \left[1 + A_0 \cdot \cos(\omega_a t + \phi_0) \right] \tag{2.29}$$

which is referred to as the *five-parameter model*. In addition to ω_a, the overall normalization N_0, time-dilated muon lifetime τ, overall asymmetry A_0, and the initial phase at injection ϕ_0 can be extracted from a fit. Note that N_0, A_0, and ϕ_0 will depend on the energy threshold applied. The last dependence arises because, as shown in Fig. 2.10, the drift time of a positron from its decay vertex to the calorimeter depends on its energy—thanks to the energy dependence of its radius of curvature. Each calorimeter, therefore, observes positrons from a larger spread of muon decays in azimuth as the energy threshold is lowered.

As the decay of a given muon is a Poisson process, the statistical fluctuations of the histogram's bin counts are governed by a Poisson distribution—but can be well approximated by a normal distribution as long as the number of counts is large enough. In this case, they can reasonably be fit to Eq. 2.29 using a χ^2-minimization method. The best-fit ω_a-value's statistical uncertainty can then be estimated from the inverse of the Hessian matrix according to

$$\sigma_{\omega_a} = \left(\frac{1}{2} \frac{\partial^2 \chi^2}{\partial^2 \omega_a} \right)^{-1/2} \approx \frac{\sqrt{2}}{\tau} \cdot \sqrt{\frac{\langle w^2 \rangle}{N_e \langle wA \rangle^2}} \tag{2.30}$$

which is approximated to leading order, where N_e is the number of positrons used in the histogram [62]. This relationship highlights the balance struck between the statistics and the asymmetry in
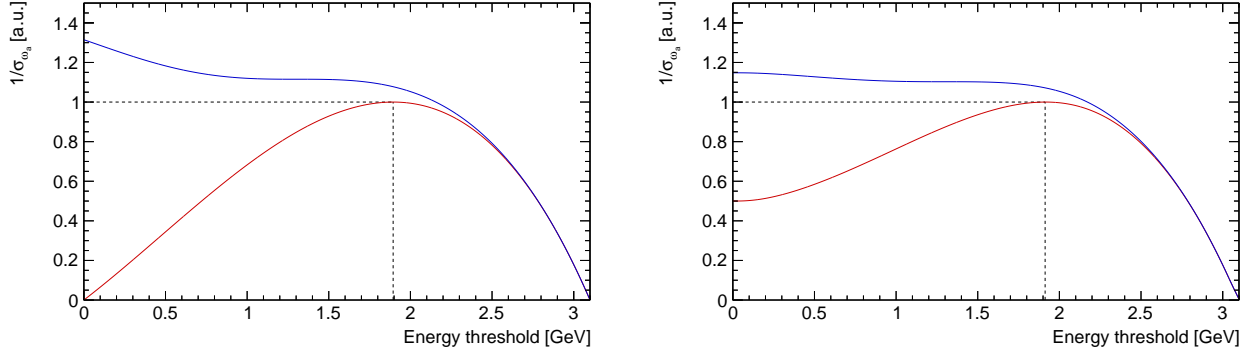
Figure 2.11: The relative statistical precisions obtained from Eq. 2.30 when using different energy thresholds with a weighting method of $w(E) = 1$ (red) and $w(E) = A(E)$ (blue). An estimate of the calorimeter's acceptance and energy resolution has been incorporated in the right plot.

Eqs. 2.25 and 2.26, which can be influenced by the weighting function $w(E)$. The extraction of ω_a can, thus, be optimized by choosing both a suitable $w(E)$ and energy threshold.

Neglecting detector effects, the baseline method—with $w(E) = 1$—reaches a maximum precision at $E_{\mathrm{th}} \approx 1.9$ GeV, as shown in Fig. 2.11. This can be improved by about 31 %, however, when weighting instead by the asymmetry—with $w(E) = A(E)$—which emphasizes the ω_a-oscillation's amplitude. Instead of deconstructively interfering, when the low-energy positrons' asymmetry changes sign, so do their assigned weights. In fact, weighting by $A(E)$ can be shown to be the optimal method [62]. The improvement over the baseline method, though, becomes more modest once the calorimeter's acceptance and resolution are included. As also illustrated in Fig. 2.11, adding such effects estimated from simulation reduces the statistical gain to about 15 %.

2.6 Systematic Effects

As explained above, after reconstructing the positrons from their calorimeter pulses, ω_a can be found by making a time histogram of weighted counts and fitting it to the model in Eq. 2.29. In addition to being straightforward, this technique—being a frequency measurement—is also robust against energy-dependent effects. Knowing the exact energy dependence of the calorimeters' ac-

ceptance or the positrons' drift time in Fig. 2.10, for example, is not required. Many such effects simply shift the normalization, asymmetry, and phase of the ω_a-oscillation that is fit. However, if these effects are also time-dependent, they can shift the ω_a extracted too. It is, therefore, imperative to understand potential biases of ω_a. Accordingly, a large part of the ω_a-measurement involves assessing the size of each effect's possible influence on ω_a (Ch. 6). For each one, a possible correction—to the underlying data, the fit model, or the extracted ω_a—can then be made, and a *systematic* uncertainty on ω_a can be assigned. After taking data for four years, the experiment's goal is to limit the combined systematic uncertainty to only 70 ppb [38].

As a concrete example, consider the energy-dependence of the positrons' drift time, $t_d(E)$. A key consequence is that a high-energy positron encodes the spin-precession phase of a muon that decayed earlier in time than that for a low-energy positron. Replacing t by $t + t_d(E)$ in Eq. 2.24, the drift time can then be seen to correspond to a *constant* phase shift, $\omega_a t_d(E)$, at each energy. Therefore, when integrated in Eq. 2.28, the number of positrons detected will be governed by the same functional form. While the best-fit ϕ_0's interpretation changes, there is no inherent bias to the best-fit ω_a. This also holds for the analysis' choice of weighting scheme, i.e., $w(E)$. In that vein, energy nonlinearities in the positron reconstruction cannot pose an inherent ω_a-bias either, since they are analogous to applying a different weighting function.

For an ω_a-bias to occur, the given effect must also have a time dependence. The muons' betatron motion, for instance, modulates the calorimeters' energy-dependent acceptance and, thereby, the average energy detected. This in turn modulates the average drift time, which introduces a time-dependence of the phase. In general, such an "early-to-late" effect poses a direct bias to the frequency extracted according to

$$\cos(\omega_a \cdot t + \phi_0(t)) \approx \cos((\omega_a + \dot{\phi}_0(0)) \cdot t + \phi_0) \tag{2.31}$$

where $\phi_0(t)$ is expanded as a Maclaurin series to first order. As a first approximation, therefore,

the best-fit ω_a will be offset from the true ω_a by a constant, $\dot{\phi}_0(0)$. For reference, a linear drift in the positrons' mean arrival time of only 50 ps over the 700-μs muon fill will shift the best-fit ω_a by more than 70 ppb. If higher-order terms are included as well, the frequency itself will become dependent on the time into the fill.

To attain the systematic uncertainty below 70 ppb, the experimental apparatus detailed in this chapter was designed to limit such phase-pulling effects via the lessons learned from the previous experiment's measurement. The positron detectors and electronics, for example, were specifically built to reduce the likelihood that two or more positrons would arrive at a calorimeter too close in time and space to be individually resolved. When such a *pileup* occurs, only one positron is reconstructed with an energy close to the sum of the unresolved positrons' energies, thereby changing the number of positrons entering the histogram. If we consider two-particle pileup, for instance, two below-threshold positrons can sum to surpass E_{th}, falsely *adding* a positron count. Alternatively, two above-threshold positrons are seen only as one above-threshold positron, falsely *subtracting* a count. These two effects do not cancel, however, because the acceptance is energy-dependent and causes them to occur at different rates.

Further, this incorrect energy will encode the wrong spin-precession phase since the drift time is energy-dependent. The time distribution of pileup should, thus, oscillate with its own average initial phase, $\phi_{\text{pileup}} \neq \phi_0$. By itself, this phase difference does not pose an inherent ω_a-bias so long as the pileup and pileup-free distributions have the *same* time dependence. In that case, the sum of the distributions has a functional form according to

$$A \cos(\omega_a t + \phi) + A_{\text{pileup}} \cos(\omega_a t + \phi_{\text{pileup}}) = A' \cos(\omega_a t + \phi') \tag{2.32}$$

with the same frequency, ω_a, but a new asymmetry, A', and phase, ϕ'. Therefore, not *all* early-to-late effects will, in fact, bias the ω_a-extraction. The rate of M-positron pileup, however, is nearly proportional to the positron rate at the calorimeter to the power of M. Pileup is accordingly a

significant source of bias, motivating the calorimeter's segmentation to enable spatial separation of close-in-time positron hits. Additionally, the rate of pileup is further reduced by maintaining a high signal fidelity in the waveform digitizers (Ch. 3) and by applying a well-designed reconstruction algorithm (Ch. 4). The pileup observed after that must then be corrected for as part of the ω_a-measurement, with the correction assessed to estimate its systematic uncertainty. There are also many other sources of systematic ω_a-biases to be considered in Chs. 5 and 6.

2.7 Electric Field and Pitch Correction

In practice, muons will be injected into the storage ring with a range of vertical angles, which led to the need for an electric quadrupole field to provide vertical focusing. An external electric field, however, will modify the ideal $\boldsymbol{\omega}_c$ and $\boldsymbol{\omega}_s$ frequencies in Sec. 2.1 [1, 63]. The spin rotational frequency will further be altered by the muon's vertical momentum component, which is parallel to the magnetic field [63]. As a result, the frequency difference between $\boldsymbol{\omega}_s$ and $\boldsymbol{\omega}_c$ observed in the experiment becomes

$$\boldsymbol{\omega}_s - \boldsymbol{\omega}_c = -\frac{e}{m_\mu}\left[a_\mu \boldsymbol{B} - a_\mu\left(\frac{\gamma}{\gamma+1}\right)(\boldsymbol{\beta}\cdot\boldsymbol{B})\boldsymbol{\beta} - \left(a_\mu - \frac{1}{\gamma^2-1}\right)\frac{\boldsymbol{\beta}\times\boldsymbol{E}}{c}\right] \qquad (2.33)$$

where constants of c have been explicitly included. This expression's magnitude is, in principle, equal to ω_a only if $\boldsymbol{\omega}_s$ and $\boldsymbol{\omega}_c$ are parallel. They can be reasonably approximated as the same, though, since the betatron motion described in Sec. 2.4 must be small due to the narrow storage aperture as well as fast compared to ω_a due to the experiment's chosen field index between 0.108 and 0.120 for Run 1. That is, the angle between $\boldsymbol{\omega}_s$ and $\boldsymbol{\omega}_c$ for stored muons is always small and averages over an ω_a-period to zero.

As will be seen in this section, the second and third terms in Eq. 2.33 lower the true ω_a relative to the measured ω_a. Therefore, two corrections—known as the pitch and electric-field correction, respectively—must be applied to the ω_a extracted from the histogram fit. These corrections are

also of $\mathcal{O}(100 \text{ ppb})$, making them non-negligible compared to experiment's target precision of 70

ppb [38]. Before they can be applied, however, they must be expressed in terms of observables that

can be directly measured, such as the muons' average vertical or radial position. This can be done

by separately considering the two new terms.

To start, suppose that the muons are all injected with *no* vertical angle and, thus, no vertical

betatron motion—i.e., $\boldsymbol{\beta} \cdot \boldsymbol{B} = 0$. In this case, the second term in Eq. 2.33 vanishes but not the

third term. For that to also occur, the muons' Lorentz factor must precisely be

$$\gamma = \sqrt{\frac{1}{a_\mu} + 1} \approx 29.3 \tag{2.34}$$

which corresponds to a "magic" muon momentum of about 3.094 GeV—the storage ring's design

momentum. The electric-field correction, therefore, is only needed for off-momentum muons with

$p \neq p_0$, making it small but non-negligible given the storage ring's momentum acceptance around

0.1 %. With β almost constant over this velocity range, the fractional shift of $|\boldsymbol{\omega}_a|$ in Eq. 2.3

relative to $|\boldsymbol{\omega}_s - \boldsymbol{\omega}_c|$ is given by

$$\frac{\Delta\omega_a}{\omega_a} = -\frac{\beta E_x}{cB_0}\left(1 - \frac{1}{a_\mu \beta^2 \gamma^2}\right) \approx -2\frac{\beta E_x}{cB_0}\frac{p - p_0}{p_0} \tag{2.35}$$

to leading order in $(p - p_0)/p_0$, where $E_x = kx$ is the radial component of $\boldsymbol{E}$ in Eq. 2.7. Further,

using Eq. 2.12 and 2.17, this fractional shift can be equivalently written in terms of the muons'

radial position, x, and equilibrium radius, x_e, with respect to R_0:

$$\frac{\Delta\omega_a}{\omega_a} = -2\frac{n(1-n)\beta^2}{R_0}xx_e \approx -2\frac{n(1-n)\beta^2}{R_0}\langle x_e^2 \rangle \tag{2.36}$$

when averaged over a spin precession's period. This latter approximation holds since $\omega_x/\omega_a > 27$

for the field indices used during Run 1. For each data set analyzed, $\langle x_e^2 \rangle$ can be inferred from the

debunching behavior (Sec. 2.4) imprinted in the calorimeter data at early times into the muon fill

[38]. This correction's systematic uncertainty then comes from that on $\langle x_e^2 \rangle$ and on how accurately

R_0 is centered in the quadrupole potential—i.e., on the placement of the quadrupole plates, and on the muons' mean vertical position.

Alternatively, suppose that the muons are all injected with the magic momentum and with a small vertical angle—i.e., $\boldsymbol{\beta} \cdot \boldsymbol{B} \neq 0$. In this case, the third term in Eq. 2.33 vanishes but not the second term. The muons will also undergo betatron oscillations, alternating the sign of the radial and vertical components of $\boldsymbol{\beta}$. As $\omega_x/\omega_a > 27$ and $\omega_y/\omega_a > 9$ in Run 1, the components parallel to $\boldsymbol{\beta}$ in the second term will, thus, nearly average to zero over a spin precession's period. The perpendicular components, on the other hand, will be reduced based on the "pitch" angle, $\psi(t) = \psi_{\text{max}} \cos(\omega_y t)$, between the muon's momentum and $\hat{x} \times \hat{y}$. Using small-angle approximations, the fractional ω_a-shift can be shown to be

$$\frac{\Delta \omega_a}{\omega_a} = -\frac{\langle \psi^2 \rangle}{2} = -\frac{n}{2R_0^2} \langle y^2 \rangle \tag{2.37}$$

when averaged over a spin-precession period [64]. In the last equality, the pitch angle was related to the vertical position using $\psi \approx \tan \psi = y\sqrt{n}/R_0$. This pitch correction can, thus, be directly measured using the trackers. An associated systematic uncertainty then comes from that on $\langle y^2 \rangle$ and from any non-ideal radial component of the magnetic field. All together, the best-fit ω_a must be slightly increased by these two corrections to obtain an unbiased measurement, from which a_μ can be ultimately determined.

CHAPTER 3

BACKEND ELECTRONICS

The ω_a-measurement relies on knowing the energy and arrival time of each decay positron that hits a calorimeter. This is enabled in part by the calorimeter system's *backend electronics*, which capture and transfer the silicon photomultipliers' output signals to the data-acquisition system to store to disk. More specifically, these electronics encompass a suite of custom-designed waveform digitizers that convert the signals from analog to digital. This conversion must be done while retaining adequate signal fidelity to allow the waveforms to be analyzed to extract each positron's required time and energy information (Ch. 4). The clock driving this digitization is supplied by a central distribution system, which also partially controls the backend electronics' behavior via a set of commands. This chapter will comprehensively review these systems' design and operation during Run 1, along with their characterization important to the ω_a-measurement.

3.1   Hardware Architecture

The backend electronics are built around the Micro Telecommunications Computing Architecture (MicroTCA) standard initially developed for small telecommunication systems [65]. Each waveform digitizer, in particular, is custom-designed as a printed circuit board conforming to the Advanced Mezzanine Card (AMC) specification [66]. This allows the digitizers (and those boards employed to distribute the clock and commands) to plug into the backplane of a commercially available MicroTCA chassis, which provides a common platform for interconnections, cooling, and power and clock distribution. We specifically use the VadaTech VT892 chassis shown in Fig. 3.1, which can support up to 12 digitizers—enough for one calorimeter. To minimize the signal-cable length from the silicon photomultiplier (SiPM) to the waveform digitizer, a MicroTCA chassis is placed close
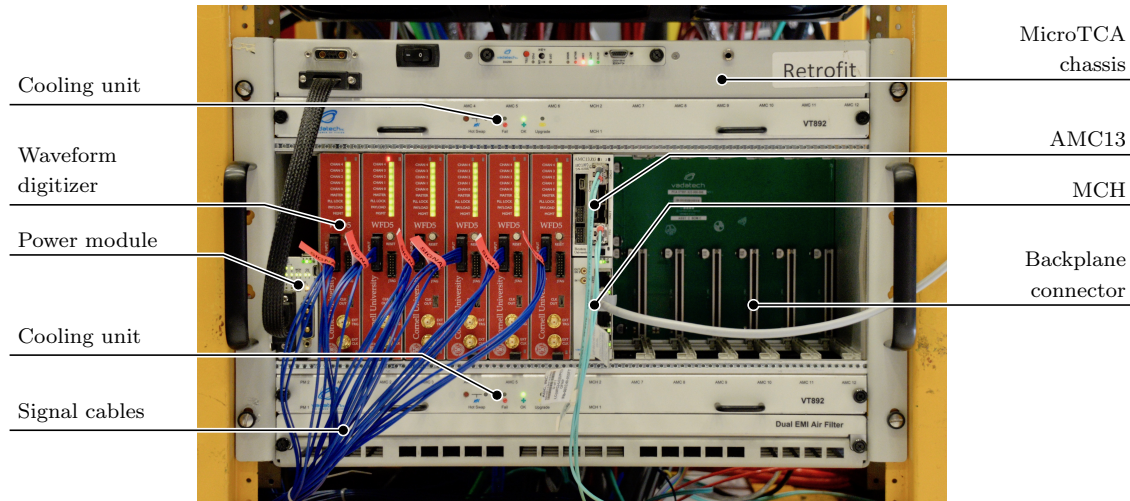
Figure 3.1: The VadaTech VT892 chassis' layout, as deployed to capture the auxiliary detectors' signals. It is populated with a power module, a MCH, an AMC13, and six waveform digitizers (Sec. 3.3) in slot numbers 1–6. Each VT892 chassis has two sets of cooling fans that run at their maximum rate to provide vertical airflow through the modules to help dissipate heat.

to each calorimeter in the magnetic fringe field of 40–60 G. To mitigate its perturbation to the muon storage region's uniform field, they have been retrofitted by the vendor for the Muon $g-2$ experiment to replace, where possible, ferromagnetic material with aluminum. Furthermore, the chassis' redundant power module, which delivers and manages the power to the digitizers, is not installed because it is unnecessary for this application.

Each chassis, per the MicroTCA standard, also has a MicroTCA Carrier Hub (MCH), which supplies Ethernet and Intelligent Platform Management Interface (IPMI) communication to each digitizer via the chassis' backplane for configuration and monitoring. In particular, the MCH (Va- daTech UTC002) negotiates the power-up of a digitizer. When the chassis is turned on, the power module distributes a 3.3-V supply to each digitizer—enough to power, per the AMC standard, its Module Management Controller (MMC). The MCH regularly polls each MMC—implemented as a micro-controller—for the board's low-level information, such as sensor readings and its power requirements. If its status is acceptable, the MCH permits a 12-V supply to be delivered to the board, which is capable of driving enough current for the digitizer's main operation. This pay-

load supply is revoked, however, if any MMC sensor reading goes out of a configurable safe range, thereby ensuring the MicroTCA system's health when collecting calorimeter data. Moreover, each MCH's network is configured such that only the data-acquisition system (DAQ) is able to communicate with the backend electronics [67]. This protects against any stray configuration change during data collection, such as those meant for a test stand.

In place of a redundant MCH, an AMC13 module is installed, which was originally custom-designed by Boston University for the CERN CMS experiment [68]. The AMC13 is primarily responsible for aggregating and transmitting the digitized data from each waveform digitizer enabled in the MicroTCA chassis to the DAQ. With a dedicated 5-Gb/s serial link to each digitizer, the AMC13 reads out their data—which are internally buffered—in parallel over the chassis' backplane using a custom protocol. It then transmits the buffered data asynchronously to the DAQ over a 10-Gb/s fiber-optic link using a stripped-down version of the Internet protocol suite. Given the waveform digitizer's design to be described in Sec. 3.3, the AMC13's buffer can hold about eight fills of data without compression—significantly less than that of each digitizer.

Further, the AMC13 is responsible for distributing the experiment's clock and commands to each waveform digitizer over the chassis' backplane. This information is received over fiber optic from the distribution system (Sec. 3.2) using the timing, trigger, and control (TTC) protocol—a two-channel, biphase mark code at 160 MBaud [69]. The "A" channel exclusively encodes a trigger bit, while the "B" channel transmits 8-bit commands. Each command requires an 8-bit overhead of start/stop bits and a Hamming code to verify the data's integrity. Accordingly, the TTC protocol sends one bit of each channel at 40 MHz. As a level transition is almost guaranteed between the channels, the AMC13 is able to robustly recover the underlying 40-MHz clock and transmit it to each digitizer so that it can decode the TTC signal—also transmitted. As detailed in Sec. 3.3, the waveform digitizers will additionally use this 40-MHz clock to derive the higher-frequency one

Priority	4-bit signal	TTS status	Description
1	0000 or 1111	Disconnected	Hardware failure or broken cable
2	1100	Error	Any other state that prevents functionality
3	0010	Synchronization Lost	Not synchronized with the DAQ
4	0100	Busy	Cannot accept triggers
5	0001	Overflow Warning	Imminent buffer overflow
6	1000	Ready	Read to accept triggers

Table 3.1: The trigger throttling system (TTS) specification [70]. Each AMC continuously reports one of five statuses to the AMC13 over the MicroTCA chassis' backplane. If an AMC is disconnected, a sixth status will be detected. The highest-priority status is relayed upstream over fiber optic to the TTC-distribution system (Sec. 3.2).

driving its analog-to-digital conversion.

The AMC13's final responsibility is to relay the MicroTCA system's overall status upstream to the TTC-distribution system in Sec. 3.2. This is transmitted over fiber optic using the trigger throttling system (TTS) protocol, which is an 8b/10b code at 400 MBaud [70]. In this feedback system, each digitizer repeatedly sends the AMC13 over the chassis' backplane one of six possible 4-bit signals—given in Tab. 3.1—defining its status. After considering its own status, the AMC13 relays the highest-priority status signal upstream. As detailed in the next section, the reported status may cause the TTC-distribution system to pause—or stop altogether—the calorimeters' data collection by withholding the triggers used by the digitizers to initiate the buffering of their digitized signal values. Thanks to the fast encoding, the MicroTCA system's status can be updated as quickly as a TTC trigger can be encoded. This allows the TTC-distribution system to react swiftly, maximizing the experiment's uptime for collecting data used in the ω_a-measurement.

3.2 Clock and Commands Distribution

A central distribution system encodes the experiment's clock and commands into a TTC signal and transmits it over fiber to the AMC13 at each calorimeter station. Located in a temperature-

controlled room outside the experimental hall, this system conducts the experiment's backend electronics by issuing triggers and commands to control their behavior in concert with the accelerator's beam cycle. In addition to the calorimeters, the systems under control include part of the electronics used in the laser-calibration, straw-tracker, quadrupole, and kicker systems. The clock used to encode the TTC signal employs a time reference independent from that used by the accelerator complex, protecting the ω_a-measurement from any bias introduced by effects synchronous to the accelerator's timing. As in the waveform-digitizer setup, this system to distribute the clock and commands—often referred to as the Clock and Commands Center (CCC)—is built around custom-designed AMCs in a MicroTCA chassis. This section will survey the CCC's hardware and firmware designs, together with how they fulfill its responsibilities.

3.2.1   Hardware Design

A block diagram of the backend electronics is presented in Fig. 3.2. The primary, 40-MHz clock source is synthesized by a Stanford Research Systems (SRS) SG382 signal generator, which has a 1-μHz resolution. Its timebase is externally supplied by a 10-MHz, Rb frequency standard (SRS FS725), exhibiting excellent short-term stability and negligible aging. This frequency standard is in turn phased-locked to a 1-pulse per second (PPS) reference from an EndRun Technologies (ET) Meridian connected to a GPS antenna (ET 0610-0009-000), mounted to the building's roof. This ensures long-term stability, and, all together, the 40-MHz clock produced is found to be sufficient for an almost negligible ω_a-systematic uncertainty in Ch. 6.

The 40-MHz clock is distributed as a TTC signal to the AMC13s in the experimental hall through an encoding/fanout network. This network is implemented in a MicroTCA chassis by three AMCs—known as FC7s—which are custom-designed by CERN and Imperial College London for the CMS experiment [71]. As pictured in Fig. 3.3, the FC7 is centered on a Xilinx Kintex-7
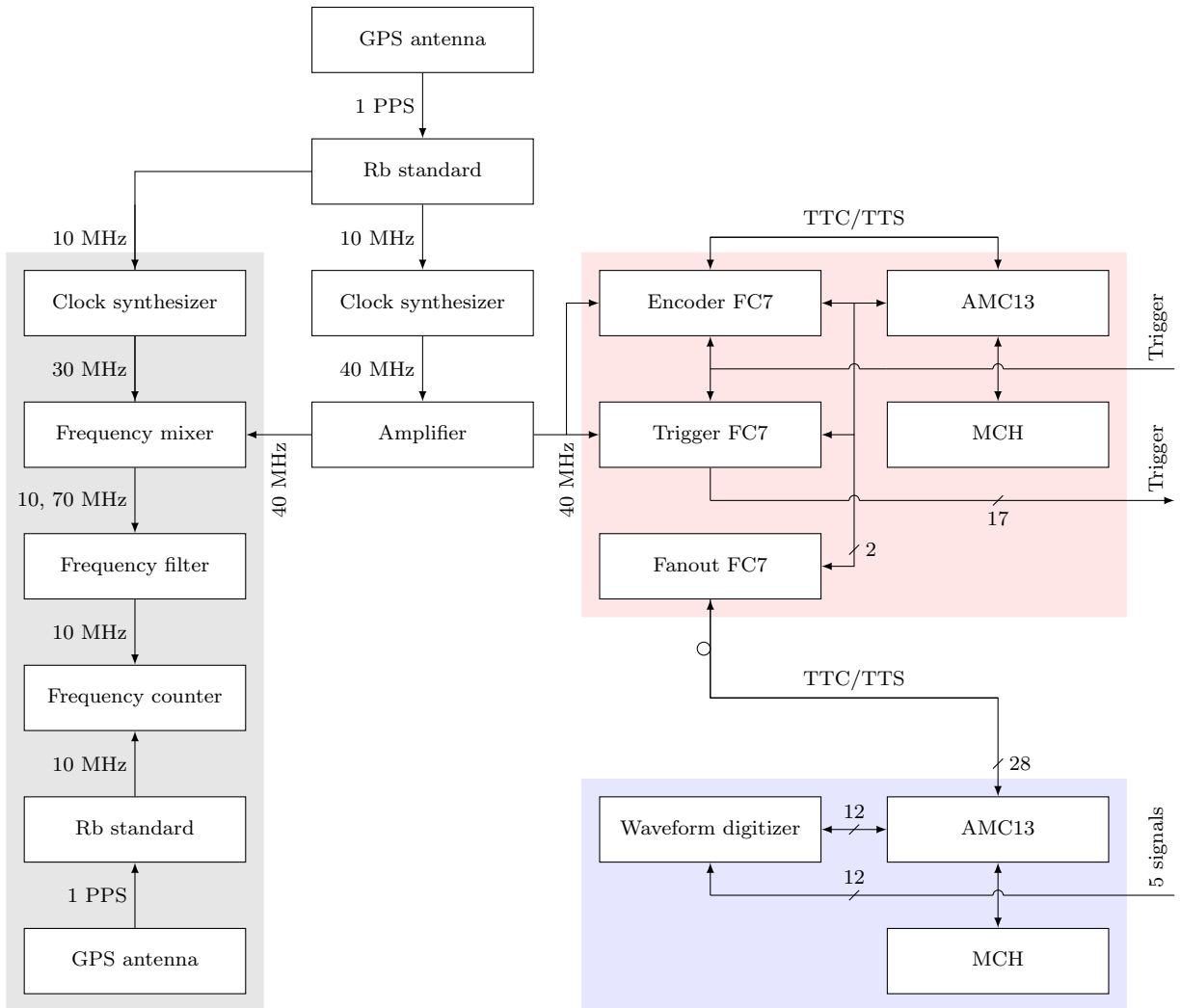
Figure 3.2: A block diagram of the backend electronics. The shaded regions with an AMC13 and MCH represent a MicroTCA chassis, where the MCH is understood to be connected with every AMC. The shaded region on the left composes the clock monitoring system described in Sec. 3.5. To provide it with a copy of the 40-MHz clock, a distribution amplifier (SRS FS370/4) is employed, which does not notably affect the clock's quality. All of these electronics except for those in the waveform digitizers' MicroTCA chassis are located outside of the experimental hall. Note that some of these clock frequencies are purposefully blinded as explained in the text.

field-programmable gate array (FPGA). In general, an FPGA is an integrated circuit consisting of a matrix of logic and input/output blocks. After manufacturing, the interconnects among the blocks can be reconfigured—or programmed—for the functionality detailed in Sec. 3.2.3.

Adding further flexibility, the FC7 also supports up to two mezzanine cards conforming to the
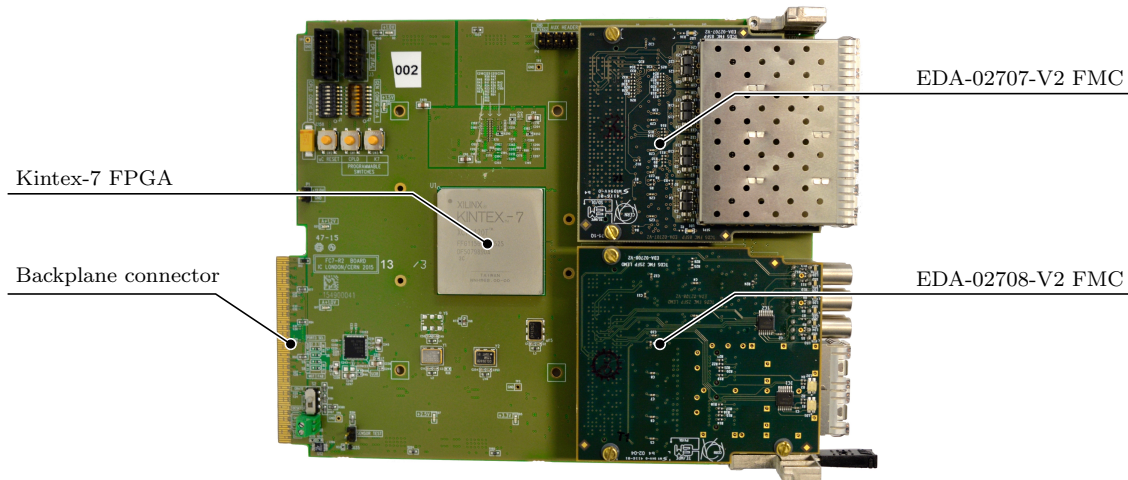
Figure 3.3: The FC7 double full-size AMC. Four of them are used to form the experiment's Clock and Commands Center (CCC). The FC7 pictured is populated with an EDA-02707-V2 and EDA-02708-V2 FMC, making it an Encoder FC7.

FPGA Mezzanine Card (FMC) specification [72]. To receive the primary, 40-MHz clock and the accelerator complex's trigger signal, the CERN EDA-02708-V2 FMC is installed, which is outfitted with six LEMO connectors. While the trigger is connected directly to the FPGA, the clock is converted to a differential signal during transit to protect against any electromagnetic noise. To transmit the TTC signal—and receive the TTS signal—the CERN EDA-02707-V2 FMC is used instead. It has eight sockets for small form-factor pluggable (SFP) modules to which multi-mode optical fibers can be connected. Such an optical distribution system is used to mitigate concerns about pickup and ground loops. It should be noted that the EDA-02708-V2 also has two such sockets, but they are not employed due to a non-optimal pinout in the FMC design. Moreover, the Finisar FTLF1318P3BTL SFP transceiver is exclusively used to convert the TTC signal from optical to copper—and vice versa for the TTS signal.

In addition, the CCC is responsible for distributing square-pulse, electrical triggers to the experiment's systems that are not instrumented in a MicroTCA chassis. They are used, for example, to discharge the kicker plates and fire the calibration system's laser. To do that, a fourth FC7

Figure 3.4: The Clock and Commands Center (CCC), implemented in a MicroTCA chassis with four FC7s. The multi-mode, optical fibers in teal from the two Fanout FC7s are routed through two patch panels to the calorimeter stations and some of the auxiliary detector systems. Note that, in place of the HiTech Global SMA/LVDS FMC, the Trigger FC7 shown is outfited with an FMC custom-designed by Cornell University, which was used after Run 1.

is deployed in the same MicroTCA chassis and was outfitted (for Run 1) with a HiTech Global SMA/LVDS FMC. This mezzanine card provides up to 34 differential-pair outputs—17 of which are routed to an SMA connecter affixed to the front of an adjacent AMC filler panel, as seen in Fig. 3.4. Each SMA output is connected through a Highland Technology V720 electrical-to-optical converter to a multi-mode optical fiber, carrying the trigger signal to the particular system in the experimental hall. This breaks any ground loops that may have formed from using a copper cable, and each system locally converts the optical signal back to electrical.

3.2.2 Implementation

The experiment's clock and commands distribution is largely controlled by the interplay among the four FC7s—each of which plays one of three roles. Each role is fulfilled by one or two FC7s outfitted with a unique pair of the aforenamed FMCs. The one largely directing the others is known as the *Encoder FC7*, which hosts an EDA-02707-V2 and EDA-02708-V2 FMC. This board uses

the primary, 40-MHz clock to encode a TTC signal encompassing the accelerator complex's trigger and the desired commands. As illustrated in Fig. 3.4, the primary, TTC signal is sent from one of the EDA-02707-V2 FMC's SFP ports to the AMC13, where it is distributed across the chassis' backplane to all four FC7s. Of the remaining FC7s, two—known as *Fanout FC7s*—then re-encode the TTC signal and, through two EDA-02707-V2 FMCs, fan it out to each AMC13 in the hall via multi-mode fibers. In the other direction, their TTS signals are received and decoded. After including its own state, the highest-priority status is reported over the chassis' backplane to the AMC13—and ultimately to the Encoder FC7.

The last FC7—known as the *Trigger FC7*—is in charge of issuing the electrical triggers and, during Run 1, had the SMA/LVDS FMC. To receive its own copy of the primary, 40-MHz clock and the accelerator complex's trigger signal, it is also outfitted with an EDA-02708-V2 FMC. This enables the board to sample the incoming trigger signal at 900 MHz—much faster than that in the Encoder FC7 at 40 MHz. Such precision timing, with a configurable delay increment at 450 MHz, is required for the trigger signals discharging the three kicker plates, which must be timed to the muon injection (Sec. 2.3). Moreover, each output trigger's width is configurable at 450 MHz and is used, for instance, to encode different firing patterns of the laser-calibration system. The Trigger FC7 also receives an electrical signal indicating the start of the accelerator's beam cycle from the Encoder FC7 between their EDA-02708-V2 FMCs. As described in Sec. 3.2.3, such a begin-of-cycle signal allows the Encoder FC7 to control when triggers are issued on a per-cycle basis.

3.2.3 Firmware Design

The Encoder, Fanout, and Trigger FC7s' functionalities are defined by their FPGAs. The configuration of each FPGA—generally referred to as *firmware*—is specific to their aforementioned role and version-controlled in the `gm2ccc` repository [73]. The general blocks of the FPGA's firmware
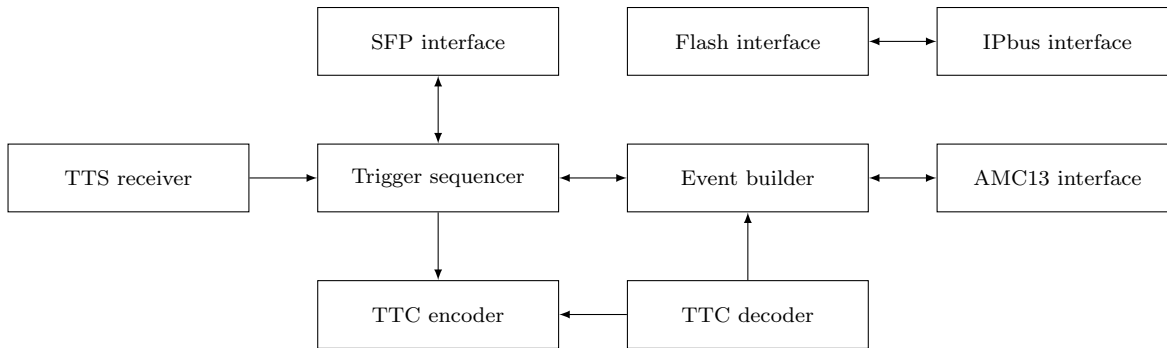
Figure 3.5: A block diagram of the top-level components in the CCC's firmware (Sec. 3.2.3). Note that the "TTC encoder" and "trigger sequencer" blocks are inoperative in the Trigger and Fanout FC7s, respectively. The "AMC13 interface" and "TTC decoder" blocks were provided by Boston University, while the "IPbus interface" and "TTS receiver" blocks were provided by CERN.

are shown in Fig. 3.5. As with the FPGAs in the waveform-digitizer design (Sec. 3.3), the firmware was coded in the Verilog and VHDL hardware description languages using Xilinx's Vivado Design Suite 2014.4. They notably differ from software languages in that one's instructions can be executed in parallel. This is essentially because the logic circuits can be configured independently of each other inside the FPGA.

However, many of the firmware design's blocks are implemented as a collection of finite-state machines—a common technique to *serialize* logic. They are basically groups of different "states" that each drive signals in a specified way, beginning in a default configuration. On a clock's rising (or falling) edge, they can transition between states based on the level of certain signals. In the Encoder-FC7 firmware, for instance, the 40-MHz input clock is upconverted—via Xilinx's mixed-mode clock manager primitive—to 160-MHz. This drives the "TTC encoder" block's finite-state machine that toggles a differential output signal according to the TTC protocol, thereby creating the TTC signal going to the CCC's AMC13 (Sec. 3.2.2).

The kernel of the firmware's design is the "trigger sequencer" block—implemented also as a finite-state machine—which controls when triggers and commands are injected into the Encoder FC7's TTC signal. This starts with the accelerator complex's trigger—more specifically, the $A6
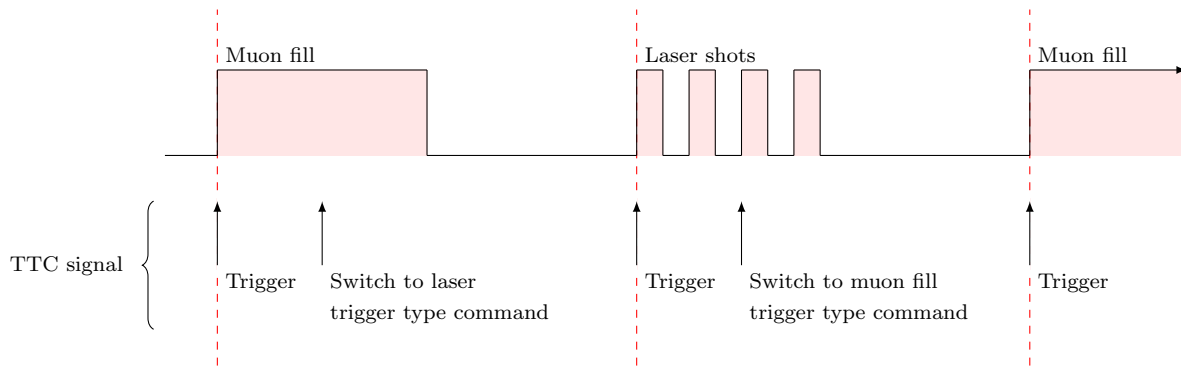
Figure 3.6: An example pattern of triggers and commands that the CCC could encode within the broadcasted TTC signal. The raised levels indicates times when the ADCs' samples are being buffered in a waveform digitizer. The CCC's trigger sequencer allows for triggers to be issued between muon fills to which the digitizers react differently.

RRBS event—indicating that a muon bunch will arrive at the experimental hall in about 400 μs. At closer to 30 μs before the muons' expected injection, a TTC trigger must be broadcasted to each calorimeter's waveform digitizers, initiating the buffering of their digitized signals throughout the fill. In addition, between muon fills, some systems require another TTC trigger for data taking when there are no muons in the storage ring. For example, as depicted in Fig. 3.6, the calibration system's laser must be repeatedly fired into each crystal between fills to track the silicon photomultipliers' gain over time (Sec. 4.6). Moreover, the digitizers' behavior to such laser-firing TTC triggers need not be the same as that for muon fills—as discussed in Sec. 3.3.4.

To meet the experiment's evolving needs of when to trigger each MicroTCA system and how they should respond, the trigger sequencer—as its name suggests—transmits a *sequence* of triggers for every one received from the accelerator complex. The delay between triggers can be configured to be between 0.4 μs and 107 s. This lower bound is because, prior to each trigger, a TTC command is sent to instruct the systems how to respond to subsequent triggers. Defined in Tab. 3.2, these "switch trigger type" commands select among their configured operation modes, and, as they are paired with a trigger by construction, each trigger is typically referred to by its *trigger type*—for example, a muon-fill or laser trigger. Further, some systems have slow operations that must be

8-bit signal	TTC command	MicroTCA module
00000010	Reset the trigger count	AMC13, FC7, WFD5
00101000	Reset the trigger timestamp counter	AMC13, FC7, WFD5
11000000	Start storing asynchronous samples	WFD5
10000000	Stop storing asynchronous samples	WFD5
00001100	Switch to the muon fill trigger type	WFD5
00010100	Switch to the laser trigger type	WFD5
00111100	Switch to the asynchronous readout trigger type	WFD5

Table 3.2: The commands encoded in the "B" channel of the timing, trigger, and control (TTC) protocol by the Encoder FC7. Only some of them are recognized by the AMC13 and FC7. In general, the "switch trigger type" commands are used to select among an AMC's operation modes. To accommodate future functionality, via a web-based configuration system (Sec. 3.4), any 5-bit trigger type can be issued without a firmware upgrade.

executed during the 197- and/or 1063-ms gap in the delivery ring's extraction cycle. A different sequence, therefore, can be configured for each of the cycle's 16 accelerator triggers. However, in Run 1, only eight sequences were configured—one for each trigger in the repeated batch of eight muon injections at 100 Hz (Sec. 2.2).

Accordingly, when the DAQ requests a new run, the trigger sequencer initially aligns itself to the cycle's pattern by waiting until it does not receive an input from the accelerator complex for more than 500 ms. It then loops through the set of preconfigured sequences, checking the run's state in between. When the DAQ requests the run to stop, the sequencer's finite-state machine will exit back to its default, idle state. If a TTS overflow warning is received, the sequencer will skip the next entire set of sequences, giving the system time to recover. However, if an overflow warning is asserted for too many consecutive cycles, the system is deemed unrecoverable, and the run will be stopped immediately. Furthermore, extensive error checking is embedded across the CCC and waveform digitizer's firmware [74, 75]. Any TTS status received with a priority above overflow warning will likewise abort the run, and a new run cannot be started until reset. This helps maximize the experiment's uptime for collecting data useful to the ω_a-measurement.

To also provide such flexible trigger sequencing to the Trigger FC7's output, it has an analogous firmware block. Instead of synchronizing to the cycle's pattern itself, however, it waits for the Encoder FC7's begin-of-cycle signal to start the first sequence. Moreover, as with the other FC7s' firmware, much of it can be configured without reprogramming the FPGA via the IPbus suite of software and firmware [76]. Through the MCH's Ethernet switch, the C++/Python and VHDL IPbus interfaces can modify an FPGA's registers to control the board's behavior—and read them to monitor its status. This versatility was particularly valuable during the experiment's commissioning in 2017 when the accelerator's cycle pattern varied. It is also used by the DAQ to request a run start or stop in the Encoder FC7. While a run is in progress, though, the registers for each FC7 (and waveform digitizer) are locked to prevent any unexpected configuration change.

Additionally, when powered up, the TTC signal output from each SFP transceiver is disabled by the EDA-02707-V2 FMC. Another key block of firmware, therefore, is an interface to control and monitor every transceiver installed. In particular, the FPGA interfaces with each one through up to two I^2C-bus devices—a switch (Texas Instruments PCA9548A) and, for some connections, an input/output expander (NXP Semiconductors PCA8574A). This architecture conveniently allows the FPGA to configure and monitor every SFP transceiver using the I^2C protocol. However, while compact, this FMC design prohibits simultaneous access to the transceivers' information in Tab. 3.3. This information includes, in part, various error/warning conditions as well as vendor information [75]. Whenever the FPGA is programmed, it uses this information for basic connectivity checks. Through a finite-state machine using the firmware's SFP interface, the DAQ can also enable the transmission of any SFP transceiver's TTC signal. When doing so, the vendor information is checked to permit only the Finisar FTLF1318P3BTL model. As these registers are part of the widely adopted multi-source agreement, they should be available in any manufacturer's SFP transceiver. Further, for a trigger-sequencer run to start, every transceiver enabled must be veri-

Signal or register name	Description
Transmission fault	Indicates a fault condition related to laser operation or safety
Transmission disabled	Indicates if data will be transmitted onto the optical fiber
Loss of received signal	Indicates if the optical signal level received is too low
Transceiver absent	Indicates if a transceiver is plugged into the SFP socket
Vendor company number	The transceiver's vendor identification number from IEEE
Vendor part number	The transceiver's part number
Vendor serial number	The transceiver's serial number
Transceiver temperature	The internally measured transceiver temperature
Alarm and warning flags	The status of internally measured sensor values

Table 3.3: The SFP transceiver information available to the FPGA through the EDA-02707-V2 FMC. The alarm and warning flags are specific to the Finisar FTLF1318P3BTL. As detailed in Sec. 3.2.3, these signals/registers are continuously monitored in the FC7's firmware.

fied to have no active alarm or warning. Even though those flags are not continuously monitored, the I^2C switch's interrupt signal alerts the FPGA if any setting changes value. After identifying the change, the run will be aborted out of caution if deemed potentially serious. For diagnostic and monitoring purposes, any SFP transceiver's information can also be read on demand via the IPbus suite, which is used by the monitoring system in Sec. 3.4.

More specifically, in the firmware, the FMC information handling is implemented as a three-layer interface [75]. This organization is designed to allow for generic communication with the I^2C-bus devices while minimizing code duplication. The first layer contains high-level finite-state machines that control the order of operations for reading and writing to the devices. They also communicate among each other—so as to have only one state machine driving the (single) clock–data bus. The second layer consists of firmware modules specific to each I^2C device model on the FMC. Each module handles the communication required for all intermediary I^2C devices between the FPGA and the device of interest. Finally, the third layer of modules performs the I^2C-bus interface to write (read) a single byte of data to (from) the particular device. All together, this architecture can be easily extended for any future FMC—which is likely to have similar chips.

0 0 0 0	AMC	Trigger #		Timestamp		Data length	
0 0 0 0 0 0	XA	e	TT	Timestamp	BT	F	Board ID
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0				Major rev.	Minor rev.	Patch rev.	
Pre-trigger gap		0 0 s x t r a b	TI	SI	FMC ID	FMC ID	
0 0 0 0 0 0 0 0	TTS mask	0 0 0 0 0 0 d f	Trigger width	Trigger width	Overflow warning threshold		
Trigger delay				Trigger delay			
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	RX LOS	0 0 0 0 0 0 0 0	TX fault	0 0 0 0 0 0 0 0	FMC absent		
TTS lock threshold		0 0 0 0 0 0 0 0	TTS lock	0 0 0 0 0 0 0 0	SFP ports		
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	TTS	TTS	TTS	TTS	TTS	TTS	TTS	TTS
TTC delay 8	TTC delay 7	TTC delay 6	TTC delay 5	TTC delay 4	TTC delay 3	TTC delay 2	TTC delay 1
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0							
SFP transceiver serial number 1							
SFP transceiver serial number 1							
. . .							
SFP transceiver serial number 8							
SFP transceiver serial number 8							
CRC		Trigger #	0 0 0 0	Data length			

Figure 3.7: The Encoder FC7's data format sent (in big-endian format) to the AMC13 for each TTC trigger. These data include the FC7's entire status and configuration settings that may be used in the ω_a-measurement. It also includes a cyclic redundancy check (CRC), which is used to ensure data integrity. Details can be found in [77].

In addition, the trigger-sequencer and FMC information per muon fill is crucial for checks of the ω_a-measurement. Therefore, for every TTC trigger the FC7 receives, it returns its current status and configuration settings to the AMC13, which passes it along to the DAQ. As depicted in Fig. 3.7, the packaged data ensures that there is no doubt as to how the calorimeter data was collected for each fill. In turn, it enables such ω_a-consistency checks as against the cycle's bunch number in Ch. 5. Even the serial number of every SFP transceiver is recorded in case any systematic effect can be linked to it. The status information also enables a means by which to monitor the system's health while data are being collected.

3.3 Waveform Digitizer Design

Enabled by the distributed TTC signal, the *waveform digitizer* measures the SiPM's voltage level over time. More specifically, at a given rate, their continuous-time output is sampled, converting

the analog voltage to a digital number. This waveform—or *trace*—must retain the signal's fidelity to infer the positron's energy and time of arrival at the calorimeter sufficiently well in the next chapter. To that end, each waveform digitizer internally upconverts their input clock from 40 to 800 MHz for such sampling with a 12-bit resolution. These improved specifications over those of the previous experiment's digitizers alone reduce the level of pileup three-fold [78]. Further, after being fabricated, the digitizers' operation must be able to adapt to the experiment's evolving needs. Consequently, the waveform digitizer was custom-designed at Cornell University to optimize its performance and cost [79]. This section presents the final hardware and firmware designs, together with the digitizer's characteristics pertinent to the ω_a-measurement.

3.3.1 Hardware Design

A diagram of the waveform digitizer's hardware architecture is given in Fig. 3.8. The baseboard— officially named the *WFD5*—features five independent digitization channels. Pictured in Fig. 3.9, each channel is centered on a 12-bit, 800-Megasamples per second (MSPS) analog-to-digital converter (ADC) (Texas Instruments ADS5401). This chip—discussed further in Sec. 3.3.2—samples the SiPM's level at 800 MHz and outputs a 12-bit value to a Xilinx Kintex-7 FPGA (7K70T-2). When triggered, these samples are buffered in a 1-Gbit DDR3 SDRAM (Micron MT41J64M16), which can hold over nine seconds of data. A sixth "master" Kintex-7 FPGA (7K160T-3) communicates with each of these channel FPGAs over dedicated 5-Gbit/s serial links to provide controls and to read out their buffered data sequentially, which are transferred to the AMC13. By fitting five channels per AMC, only one MicroTCA chassis is needed to hold enough waveform digitizers for each calorimeter's 54 SiPMs. It also allows space for a hot spare to recover quickly from any hardware failure or to use for the experiment's auxiliary detectors.

To produce each channel's 800-MHz sampling clock, an on-board clock synthesizer (Texas In-
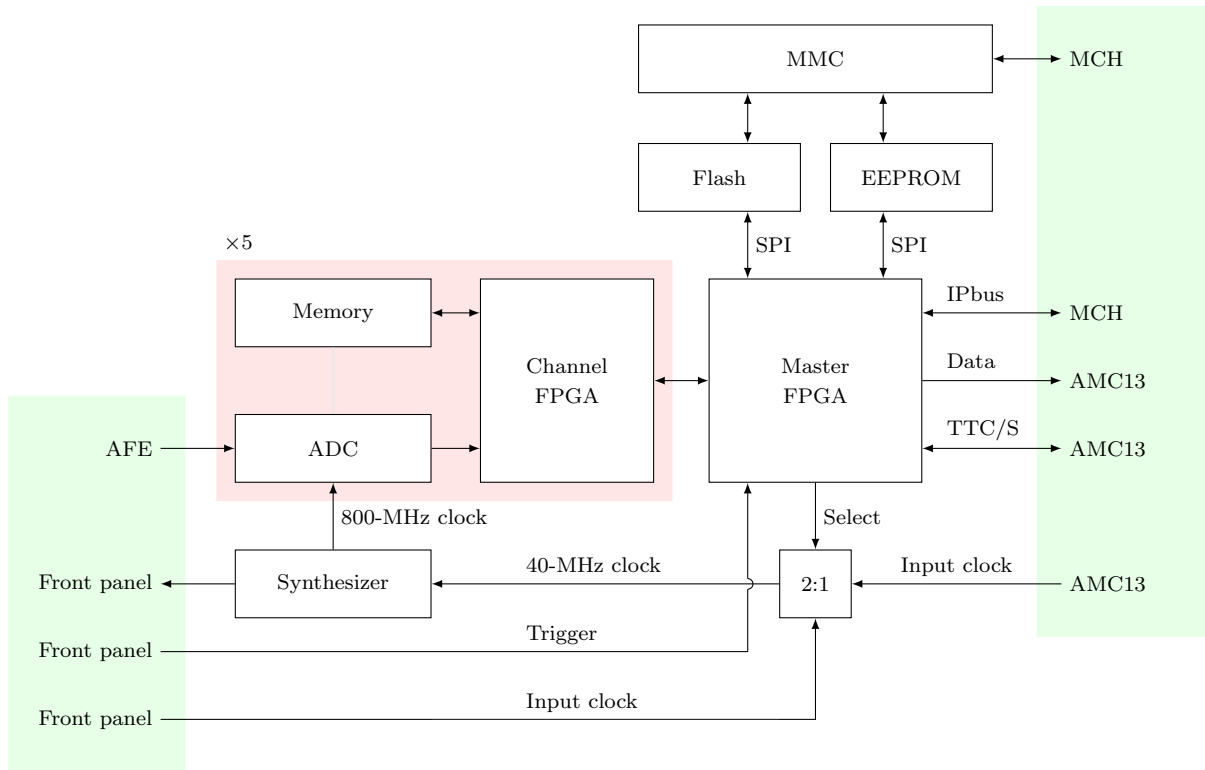
Figure 3.8: A block diagram of the waveform digitizer's hardware architecture. The baseboard feature five independent, digitization channels, which are each controlled by an FPGA. Another FPGA serves as an interface between these channels and the MicroTCA chassis' backplane, and it can optionally employ a trigger and/or clock from SMA connectors on the front panel.

struments LMK04906)—detailed in Sec. 3.3.3—frequency upconverts the 40-MHz clock received

from the AMC13 over the MicroTCA chassis' backplane. Per the MicroTCA standard, the WFD5

also has a micro-controller (Atmel AT32UC3A1) that serves as its MMC, whose design/code is

based on that used by the AMC13 [68]. For Ethernet/IPMI communication with the MCH, the

board's IP and MAC addresses, which encode its serial number, are stored in a 2-kbit EEPROM

(Microchip MCP98243). In addition, the firmware for all six FPGAs is stored in a 256-Mbit flash

memory (Micron NS25Q256), which allows them to be programmed automatically on power-up (or

user request). This includes a golden image of the master FPGA's firmware in case of corruption,

allowing for firmware updates via an SPI interface.

The waveform digitizer also features two custom-designed mezzanine cards. Pictured in
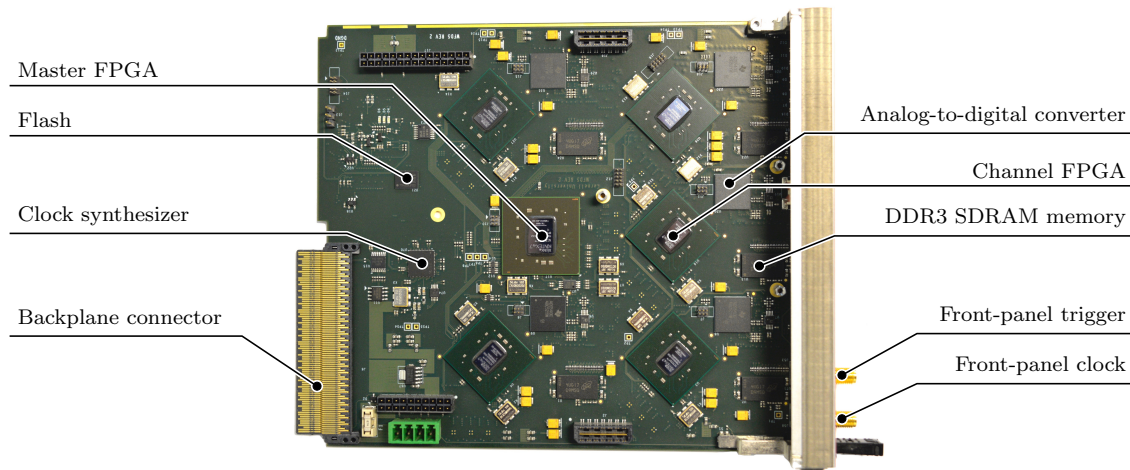
Figure 3.9: The WFD5 double full-size AMC. Eleven of them instrument a calorimeter's 54 SiPM signals. It was custom-designed at Cornell University; the board itself was fabricated by Streamline Circuits and assembled by ABC Assembly. The board's major components are labeled save for the MMC, which is located on the back near where the flash is. The connectors at the top and bottom are used to accept two mezzanine cards, pictured in Fig. 3.10.

Fig. 3.10, the first card creates the baseboard's internal voltage rails from the chassis' 12-V supply, which is controlled by an on-board micro-controller (Atmel ATXMEGA8E5). This allows for easy replacement upon failure and minimizes overall risk. The second card—known as the analog front-end (AFE)—conditions the analog signals from the SiPMs prior to digitization, providing inexpensive customization of the digitizer. Through an ECDP connector, it accepts a differential signal for each channel with a common mode of 2.5 V, which, through a buffer on the AFE and WFD5, is inputted to its ADC with a common mode of 1.9 V—per the ADS5401 specification. To shield the analog signals, they are placed between two ground planes inside the printed circuit boards. Three daisy-chained digital-to-analog converters (Analog Devices AD5666) also supply a programmable offset to each signal in the AFE's buffer, allowing the mostly unipolar SiPM signals to be mapped to the ADC's full ± 0.5-V range. Lastly, the input signals are bandwidth limited in the AFE to 230 MHz to reduce noise and DC-coupled to avoid a rate-dependent pedestal.

In total, 332 waveform digitizers were manufactured, with 318 of them passing an extensive
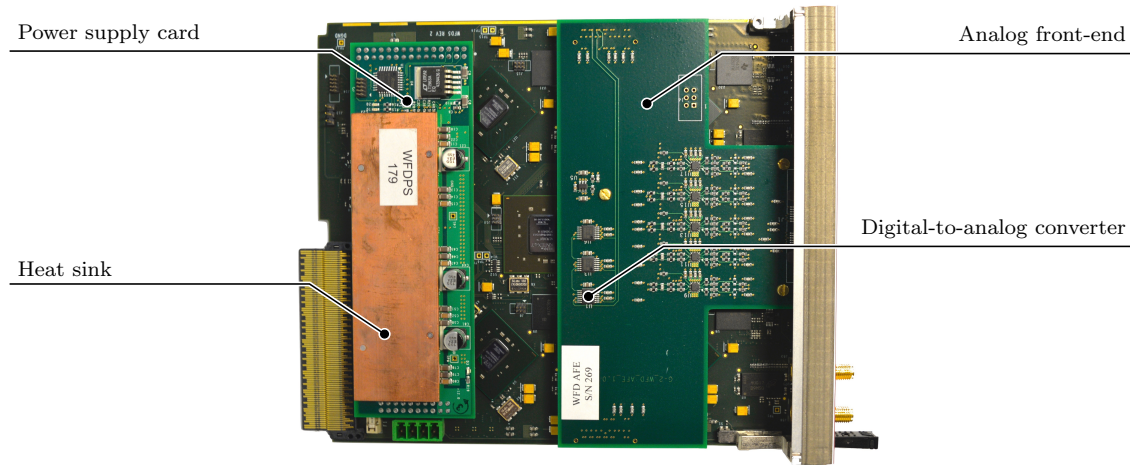
Figure 3.10: The waveform digitizer as deployed, featuring two custom-designed mezzanine cards (Sec. 3.3.1). The copper heat sink attached to the power supply card helps to dissipate heat; the three daisy-chained digital-to-analog converters are used to offset the input signal.

suite of quality control tests for nearly a 96 % yield. Specifically, 14 baseboards (96 % yield), 6 power supply cards (98 % yield), and 2 AFEs (99 % yield) presented anomalies and were not deployed in the experiment's backend electronics. All together, 289 waveform digitizers instrumented the calorimeter and auxiliary systems. Further, the impedance of each signal cable connected to the AFE was measured and only used if within 0.5 of 97 Ω—the AFE's impedance. This mitigates signal reflections at the cable's interfaces that may end up in the data. After the quality control procedures, the health of each board—and the entire MicroTCA setup—was monitored primarily by the firmware's thorough error checking and the system described in Sec. 3.4.

3.3.2 Analog-to-Digital Converter

A crucial component of the waveform digitizer is the ADC. As it must preserve the SiPM's signal fidelity, the ADS5401 chip performance was extensively characterized to identify its ideal configuration registers. Notable ones include 3.5-mA LVDS current strength; 2-kΩ and 100-Ω LVDS termination in the ADC and channel FPGA, respectively; and the high-performance mode 1 en-

abled. This chip also provides a mechanism to synchronize the internal clocks with other ones; however, the baseboard's design does not include connections to the necessary pins. As a result, after reprogramming their registers, the board's ADCs can synchronize to a different edge of the 800-MHz sampling clock. This can cause a 1-clock tick (c.t.) offset among the channels' buffered data. Notwithstanding, any such offset will be automatically removed in the reconstruction's time-correction algorithm (Sec. 4.4).

In addition, an understanding of how the ADC operates directly informs how best to prepare its output for reconstruction. The ADS5401 is composed of two internally interleaved ADCs with a shared track-and-hold circuit. Broadly speaking, this circuit is formed by two capacitors with a switch that toggles between them at half of the 800-MHz sampling rate. At any given time, one capacitor is charged by the input signal, tracking its level. When the switch toggles, the current voltage is held for its ADC's circuitry to resolve (while the other one tracks the input). To first order, this voltage is actually an average over the time the switch transitions from low to high impedance. The sample-to-sample variation in the effective sampling time—often known as the *aperture jitter*—is around 100 fs for the ADS5401 [80].

Further, non-idle FET switches, through channel charge injection, typically drop the held voltage slightly. Since there are two capacitors, this can cause a pedestal and gain difference between adjacent samples. The ADS5401 offers an internal correction circuit for such differences, but it is disabled to avoid any rate-dependent behavior. However, on power-up, the ADS5401 always begins digitizing with the same interleaved ADC. As done in the next chapter, this permits such a pedestal/gain offset to be characterized based on the sample's parity and optionally corrected.

The ADS5401 output is an LVDS pair for each of the 12 bits and the 400-MHz double data rate clock. These lines pass directly to the channel FPGA, which latches each bit's serial data on both clock edges by employing the Xilinx IDDR primitive in firmware. The clock edges, therefore,
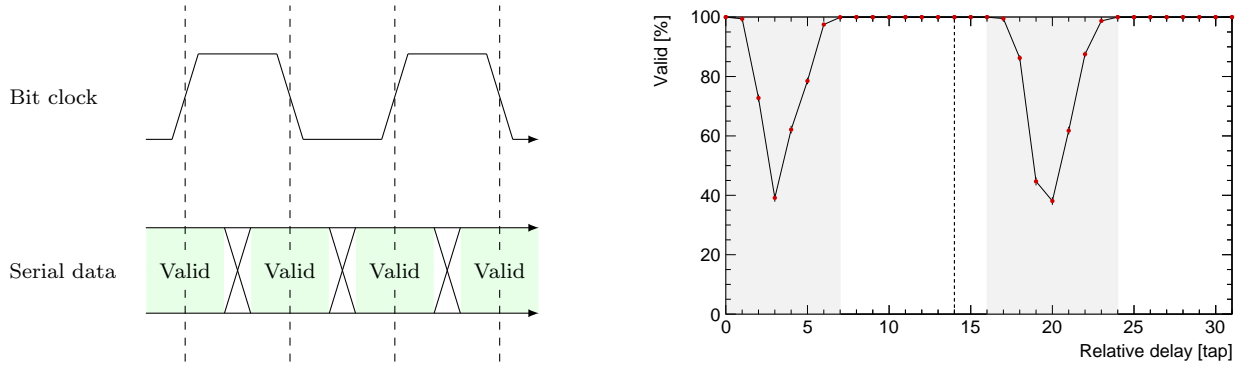
Figure 3.11: Left: A timing diagram, where the serial data's level is latched on both edges of the clock. Stable levels are achieved when the data and clock are aligned with respect to when the data may make a level transition. Right: The percentage of valid data collected when delaying the ADC's output data in steps of 1 tap $\approx$ 74 ps, averaged over 1410 channels. A relative delay of 14 taps for all channels was used in Run 1. In a separate measurement, the validity range was also found to be stable to 1 tap for FPGA temperature changes up to 20 °C.

must be aligned—as shown in Fig. 3.11—to when the data logic levels are stable, i.e., not in transition between levels. To achieve such alignment, the data is internally delayed in each channel FPGA. More specifically, through the Xilinx IDELAY primitive, up to 31 delay taps—where each tap is calibrated to be about 74 ps—can be added to the data path. To identify the optimal delay, each ADC was temporarily configured to output a test pattern of alternating logic levels. A scan over the delay taps was then performed via IPbus, measuring for each one the percentage of samples that were correctly latched. As shown in Fig. 3.11, the range of valid data is about 60 % of the 1.25-ns sampling period, with the middle delay tap of 14 chosen for Run 1.

3.3.3   Clock Synthesizer

Another crucial component of the waveform digitizer is the clock synthesizer, and its performance must be similarly understood for the ω_a-measurement. The LMK04906 chip, in particular, is configured to operate in its dual phase-locked loop (PLL) mode. The first and second PLL circuits employ an external, 25-MHz voltage-controlled crystal oscillator (VCXO) and an integrated, low-noise voltage-controlled oscillator (VCO), respectively. In order to upconvert 40 to 800 MHz, the
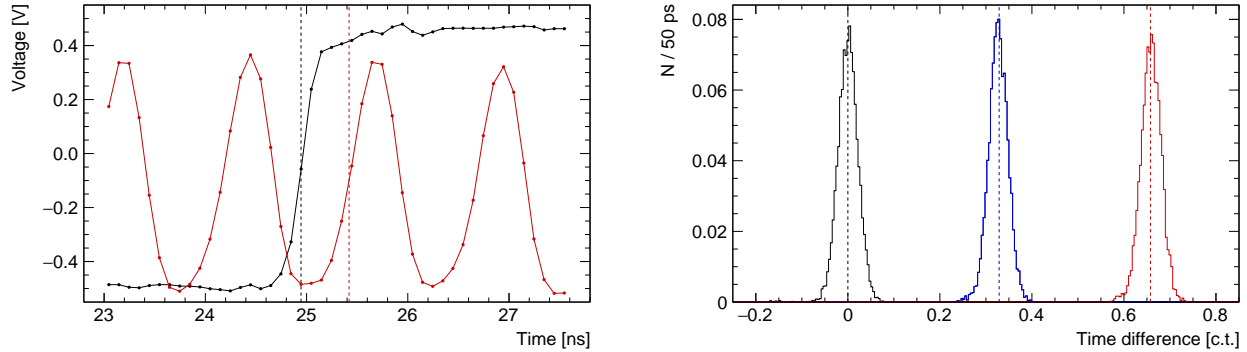
Figure 3.12: Left: A trace of the 40-MHz input (black) and the 800-MHz output (red) clock from a waveform digitizer. The time difference between their 50 % levels on the rising edge was measured after reprogramming its clock synthesizer's registers. Right: The normalized distributions of those measurements, exhibiting the expected 0-, 1/3-, and 2/3-c.t. offsets. Details of these measurements are documented in [81].

internal VCO's frequency is tuned to 2400 MHz, and the output is divided by three. The output 800-MHz clock, however, can be shifted by a multiple of the VCO's period, 1/3 c.t., because the VCXO and VCO are—by construction—asynchronous. As shown in Fig. 3.12, such 0-, 1/3-, and 2/3-c.t. offsets have been verified on the bench [81].

Moreover, as part of the ω_a-measurement's blinding plan in Sec. 5.2, the SG382 clock was in fact set to be $40 + \varepsilon$ MHz, where ε is a small frequency offset. As ε was unknown to the collaboration during data collection, the clock synthesizer's performance had to be previously characterized for varying ε-values. For each synthesizer deployed in Run 1, their dual PLL remained locked—a requirement for its proper behavior—for -5778 Hz $\leq \varepsilon \leq 5197$ Hz. This lock is continuously monitored in the master FPGA's firmware and will cause the CCC to abort the run—via its TTS status—if it should ever become unlocked.

This range was initially checked using two LMK04906 evaluation boards before adding the chip to the baseboard's final design. For *in situ* assessments, the LMK04906's spare, sixth output was routed to a Micro-USB connector on the baseboard's front panel. Most importantly, as demonstrated in Fig. 3.13, the upconversion factor, α, in the output clock's $\alpha \cdot (40 + \varepsilon)$ MHz rate was
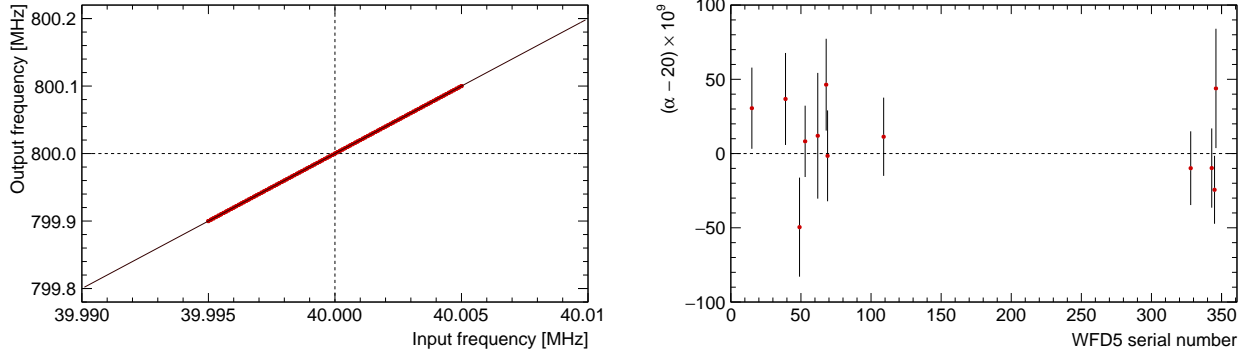
Figure 3.13: Left: A representative scan over the input, 40-MHz clock's ε-offset. The slope α with the output, $\alpha \cdot (40 + \varepsilon)$ MHz clock can be measured to a few ppb within about a ± 5 kHz range. Beyond that, the clock synthesizer's dual PLL becomes unlocked. Each input-frequency measurement was taken using a 10-s gate time and a 50 sample size. Right: The α-value measured for the 12 waveform digitizers instrumenting calorimeter 24.

found to be the nominal value of 20 to ≤ 2.1 ppb for every WFD5 tested. Unsurprisingly, this precision about an order of magnitude larger than that measured on the evaluation board since it has a higher-noise environment. Notwithstanding, when correcting for ε, the variance of α will be added to that of ω_a measured, as discussed in Ch. 6.

3.3.4 Data Acquisition Modes

The master and channel FPGAs' firmware is outlined in Fig. 3.14 and is version-controlled in the `gm2wfdfw-mfpga` and `gm2wfdfw-cfpga` repositories, respectively [82, 83]. It was designed to allow for flexible data acquisition. In the main operation mode, when the master FPGA receives a TTC trigger, its "trigger manager" block broadcasts it to each channel FPGA over a dedicated connection. A set of channels, however, can be disabled via IPbus and will not receive the trigger. Each channel FPGA checks the level of this input on the rising edge of the 400-MHz clock from the ADC. However, inevitable variations on the exact position of the clock edges will occasionally cause a channel to acknowledge the trigger 2.5 ns later than another.

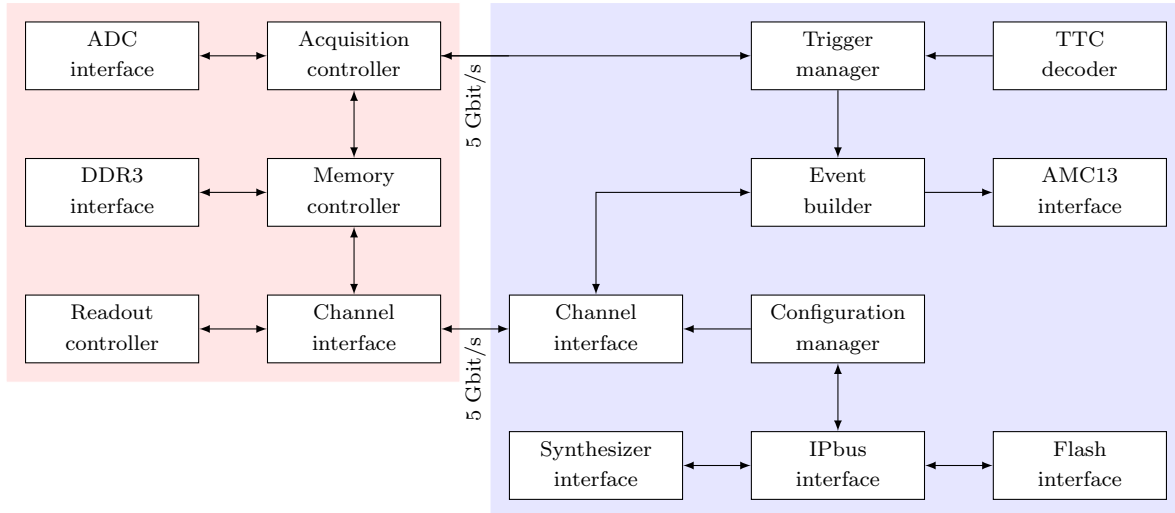When triggered, the channel FPGA will begin buffering the ADC input into the DDR3 SDRAM

Figure 3.14: A block diagram of the top-level components in the waveform digitizer's firmware (Sec. 3.3.4). It is designed primarily for configuring, triggering, data acquisition, and reading out collected data. The "AMC13 interface" and "TTC decoder" blocks were provided by Boston University; the "IPbus interface" block was provided by CERN. The shaded regions in red and blue delineate the channel and master FPGA's firmware, respectively.

according to a regular pattern—defined by the number of consecutive samples to buffer, known as a window; the number of windows; and the gap time between windows. Up to three acquisition patterns can be configured, with the specific one to use specified by the trigger type (also relayed from the master FPGA). In Run 1, every muon fill was buffered by a single window of 560 000 samples, while, to avoid saving samples containing only noise, the laser shots between muon fills were instead buffered by four windows of 800 samples separated by 1600 samples. Afterward, the channel replies to the master FPGA that data acquisition is complete.

Once all enabled channels respond, the "event builder" block sequentially reads out each channel's buffered data—through the Xilinx Aurora 8b/10b core—to package up and send to the AMC13. As illustrated in Fig. 3.15, auxiliary information is also sent to record unambiguously the board's condition/configuration when the TTC trigger arrived. Nonetheless, the data throughput to the AMC13 is maximized at 4 Gbit/s—limited by the 8b/10b encoding over the backplane's 5-Gbit/s link. If a TTC trigger arrives during readout, the firmware's logic to transfer data will

0 0 0 0	AMC			Trigger #			Timestamp		Data length	

CE	1	XA	e	TT	Timestamp		BT	a	Board ID

0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0		Major rev.	Minor rev.	Patch rev.

0 1	XA	Channel tag	Waveform gap	Waveform count	DDR3 address

DDR3 address	Waveform length	TT	Trigger number

Waveform count	DDR3 address	TT	Waveform length

0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	Channel tag	Waveform gap	Waveform index

Waveform 1 ADC data

. . .

Waveform count	DDR3 address	TT	Waveform length

0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	Channel tag	Waveform gap	Waveform index

Waveform N ADC data

Channel checksum

Channel checksum

Data integrity check	Data transfer time

. . .

CRC	Trigger #	0 0 0 0	Data length

Figure 3.15: The waveform digitizer's synchronous-mode data format sent (in big-endian format) to the AMC13 for each TTC trigger. These data include the buffered ADC samples as well as the board's status and configuration settings. It also has a cyclic redundancy check (CRC), which is used to ensure data integrity. Details can be found in [77].

pause until the new data acquisition is finished because the DDR3 memory cannot be read from and written to simultaneously. As such, this mode's data acquisition relies on the predictable trigger pattern, synchronous to the accelerator complex.

Alternatively, the waveform digitizer can operate in an *asynchronous* mode, where ADC samples are continuously stored in a circular ring buffer in the channel FPGA. When the master FPGA receives an electrical trigger through the baseboard's front panel, a configurable number of ADC samples before and after receiving the trigger are buffered in the DDR3 memory. Storing of the asynchronous samples, however, must be enabled first by the CCC via the TTC command given in Tab. 3.2. The buffered data are then read out in their entirety when a TTC trigger is received while the "asynchronous readout" trigger type is configured. In Run 1, this operation mode was used to capture thyratron signals in the kicker system, which helped inform the data selection in the ω_a-measurement, and were read out after each set of eight muon injections. The laser-calibration

system also plans to use it to record future signals from an americium source.

3.3.5 Characterization

The hardware and firmware designs of the waveform digitizer were thoroughly characterized over three revisions from 2014 to 2017. The wide variety of tests included any performance dependence on a magnetic field up to 46 G, a different subset of channels enabled, or a specific bank in the DDR3 SDRAM. They were primarily conducted by analyzing noise, ADC test patterns, and a 10-MHz output from an FS725 frequency standard. In sum, all of the identified performance issues were resolved in the final design deployed in the experiment.

As shown in Fig. 3.16, the board's intrinsic noise level—as measured when the signal input is shorted—was found to be 0.51 ± 0.13 mV among 1385 channels. This noise level is limited by the board design, not chip-to-chip variations. The as-deployed noise level, in contrast, was found to be 0.96 ± 0.35 mV—still less than 0.1 % of the ADS5401 input range of 1 V_{PP}. By fitting laser pulses incident on a SiPM in a similar fashion to that in Ch. 4, an intrinsic timing resolution between channels of less than 22 ps was further found. Moreover, the amplitude and baseline fitted were rate-independent to $\leq$0.2 %. The digitized pulse also scaled linearly with the input, where at most an $\mathcal{O}(10^{-5})$ quadratic term was measured for each board.

Additionally, no statistically significant evidence of crosstalk among the channels has been found. Further, when a $>$1 V_{PP} input voltage is supplied, no recovery time is detected in the first few samples that come back in the ADC's range. They were also heavily stress-tested prior to Run 1, including at the SLAC T-536 test-beam run to evaluate the full calorimeter system using the electron beam at the End Station Test Beam in July 2016 [60]. Throughout Run 1, the waveform digitizers deployed exhibited no faults and fulfilled all of the experiment's requirements. This performance, along with its versatile design, could make it also of interest to future
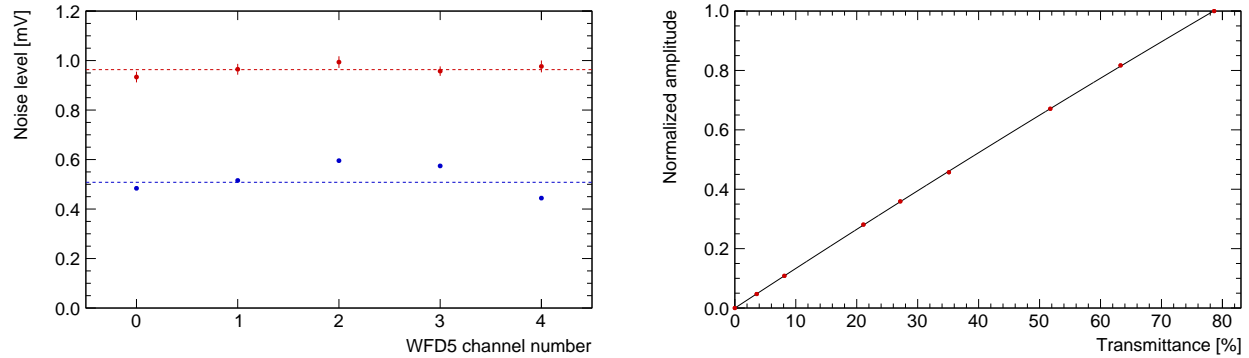
Figure 3.16: Left: The intrinsic (blue) and as-deployed (red) noise level for each channel of the waveform digitizer. They are averaged over 277 and 264 digitizers, respectively. The intrinsic noise was measured by shorting their signal inputs. Right: The fitted amplitude of a laser pulse on a SiPM for different optical-density filters used to diminish the transmittance. The plot shown is from a representative digitizer, exhibiting a quadratic term of $\mathcal{O}(10^{-5})$.

experiments—beyond Muon $g - 2$—in search of a cost-effective waveform digitizer.

3.4   Monitoring and Configuration System

All of the backend electronics are continuously monitored using a system of custom-designed web pages—known as the MicroTCA Crate Monitor (MCM)—which is version-controlled within the `gm2monitor` repository [84]. About every two seconds, this system reads the status registers in each AMC's firmware via IPbus and—without the need for a browser refresh—updates the information on every web page. Such real-time updating is accomplished using the Flask-SocketIO extension to the Flask microframework and written using a mixture of Python, HTML, CSS, and JavaScript. Each MicroTCA system—or *crate*—spawns a web client, which communicates with a central server by sending and receiving events. The web pages themselves were largely designed employing the Bootstrap framework and, for online graphing, the Plot.ly toolkit, allowing for a user-friendly visual of the electronics' status.

The MCM also serves as a key communication link between the DAQ and CCC. When a warning or error flag is found in a register, it will raise an alarm in the DAQ that in turn immediately

stops the DAQ run—before it crashes. As the (asynchronous) register polling only needs a working Ethernet link via the MCH, this scheme is robust against the potential range of faults in the FC7 and the WFD5. To help anticipate some error conditions, the buffer size and FPGA temperature, in particular, are plotted in real-time. For protection, however, the XADC primitive is implemented in every FPGAs' firmware, which is used to automatically power down the chip if its die temperature exceeds a factory set limit of 125 °C. In addition, for a permanent record, the electronics' status is automatically uploaded to an online logbook whenever a run is aborted, allowing for investigations after the DAQ is restarted.

Together, the MCM provides an invaluable tool for the experiment's operations. The main web page, for instance, provides the experiment's operators with a snapshot of the backend electronics' health, which is checked before a new run is started manually. Additional crate-specific web pages provide the electronics experts with detailed information on each of the 294 deployed AMCs, enabling a quick diagnosis of errors. These more detailed pages were especially useful while commissioning the experiment in 2017. Adding to its value, the Encoder and Trigger FC7's trigger sequences can also be easily configured in the MCM. The configured settings can then be exported/imported in the JSON file format. Furthermore, complementing the MCM's real-time view, a program in the DAQ simultaneously polls a few key registers in each AMC each minute, storing them in a PostgreSQL database. They were later used to inform the data selection for the ω_a-measurement in Run 1, as discussed in Sec. 5.1.

3.5   Clock Characterization

As the ω_a-measurement is time-based, the clock signals—as seen by the waveform digitizers—directly influence the measurement result. A careful characterization of the clock, therefore, is required to assess a systematic uncertainty on ω_a from it. In review, the ADC's 800-MHz sampling
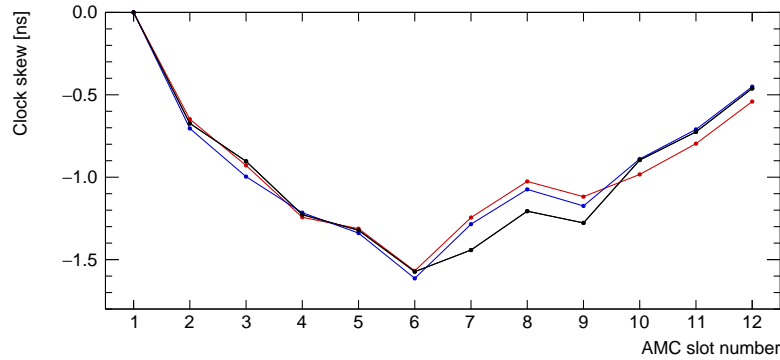
Figure 3.17: The clock skew among a VT892 chassis' 12 AMC slots. The colors denote three different AMC13 modules used, showing that the latency is primarily due to the MicroTCA chassis' backplane and not the AMC13's different clock paths.

clock begins as the SG382 synthesizer's output. Before reaching the digitizer's LMK04906 synthesizer, it is converted up and down twice from 160 MHz inside an FC7 and an AMC13, respectively. As shown in Fig. 3.17, a clock slew also exists up to 1.6 ns among the VT892 chassis' 12 AMC slots; however, the CCC minimizes its effect by placing the Fanout FC7s symmetrically about the MCH. When it reaches the digitizer's synthesizer, it is upconverted once more to the 800-MHz sampling rate and passed through a buffer to the ADC. Aside from scrutinizing the frequency upconversion (Sec. 3.3.3), this sampling clock can be characterized in terms of variations from one clock edge to the next—the *jitter*—and, over a longer time-scale, in the mean frequency—the *stability*.

An uncertainty in the clock's edge, in particular, leads to an uncertainty in when a positron's signal was sampled—on top of the ADC's intrinsic aperture jitter mentioned above. This in turn degrades how well its arrival time at the calorimeter can be extracted from the signal in the reconstruction. Accordingly, the sampling clock's jitter should be as low as possible, and, to help ensure an acceptable level, the CCC's clock equipment was selected to have excellent short- and long-term stability specifications. That said, there remain many jitter sources, including supply noise, input/output crosstalk, and intrinsic jitter in the FPGA.

The TTC-distributed clock, therefore, was carefully characterized both before and after the
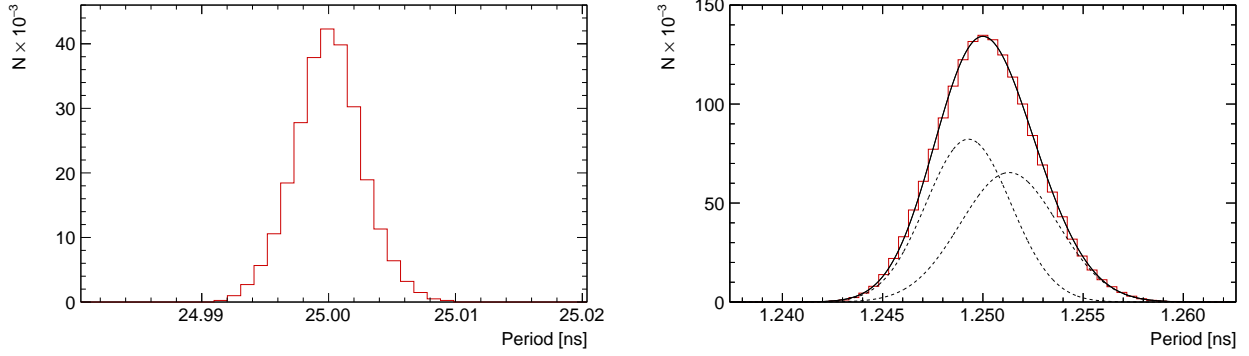
Figure 3.18: The period distribution measured for the 40-MHz (left) and 800-MHz (right) clock. The population standard deviation of 2.67 and 2.49 ps, respectively, can be used as a measure of the clocks' period jitter. These measurements, however, were limited by the LeCroy WaveRunner 6100A oscilloscope used. The 800-MHz clock's distribution was fit to a dual-Dirac jitter model, yielding $Dj(\delta\delta) = 2.030 \pm 0.015$ ps. The fit's two normal distributions are shown as dashed lines, while the sum is overlaid as a solid line.

LMK04906 synthesizer in the clock's path. At the 2017 commissioning's onset—prior to shifting the base frequency by ε—the period jitter, more specifically, was measured. As seen in Fig. 3.18, the distribution of the 40- and 800-MHz clock's period was found with a LeCroy WaveRunner 6100A oscilloscope, and their population standard deviations were measured to be 2.67 and 2.49 ps, respectively. This measurement, however, was limited by the oscilloscope's internal sampling clock and voltage noise floor as well as the input clock signal's slew rate. These jitter values are, thus, upper bounds on those seen by the digitizers.

Further, such jitter can be decomposed into a random and a deterministic component. The unbounded, random interference can be modeled by a normal distribution, while the bounded, deterministic interference—such as from a switching power supply—can be incorporated to first order by convoluting it with two Dirac delta functions. Mathematically, this industry-standard, *dual-Dirac model* reduces to the sum of normal distributions and is needed to describe the sampling clock's period distribution well in Fig. 3.18 [85]. The two delta functions' time separation, $Dj(\delta\delta)$, is 2.030 ± 0.015 ps, indicating a small amount of deterministic jitter in either the sampling

clock or the oscilloscope used. In either case, the period jitter measured, along with the ADC's aperture jitter, was found in simulation to be sufficiently low to achieve satisfactory time resolution of the positron hits, as will be described in Sec. 4.8.

In addition, the ω_a-measurement will be systematically biased if the *mean* frequency of the clock changes over the muon fill. Further, if a single conversion factor of s/c.t. is used to unblind the ε-offset, the sampling clock's mean frequency must also be held stable over the several months in which data are being taken. Such stability can be characterized by the *Allan deviation*—a measure of the mean frequency's fluctuation from one measurement to the next. Formally, it can be estimated from a sample size of N measurements by

$$\sigma_y(\tau) \approx \left[\frac{1}{2(N-1)} \sum_{i=1}^{N-1} (\bar{y}_{i+1} - \bar{y}_i)^2 \right]^{1/2} \tag{3.1}$$

where $\bar{y}_i$ is the i-th measurement of the fractional frequency averaged over τ—typically called the gate time. In contrast to the standard deviation, the Allan deviation does not diverge with N if non-white-noise frequency deviations are present. This makes it, in general, a widely used metric, with a lower Allan deviation indicating better stability over the measurement period.

The Allan deviations of the 40- and 800-MHz clocks were measured with an SRS SR620 frequency counter, whose timebase was externally supplied by the same FS725 frequency standard used in the CCC. As shown in Fig. 3.19, for gate times up to 100 s, both clocks are consistently stable. The sampling clock's Allan deviation is approximately 25 and 0.06 ppb fill-by-fill and cycle-by-cycle, respectively. Note that these values are not limited by the SR620's Allan deviation. Moreover, as also shown in Fig. 3.19, the 40-MHz clock's values were confirmed by alternatively measuring its phase noise for increasing offset frequencies from 40 MHz with a Tektronix TDS5104 oscilloscope. These stability measurements, however, were uncorrelated with the backend electronics' operations. Assuming many muon fills, for an average ω_a-bias, the mean frequency must actually fluctuate *coherently* across the fills. The Allan deviation, therefore, was additionally stud-
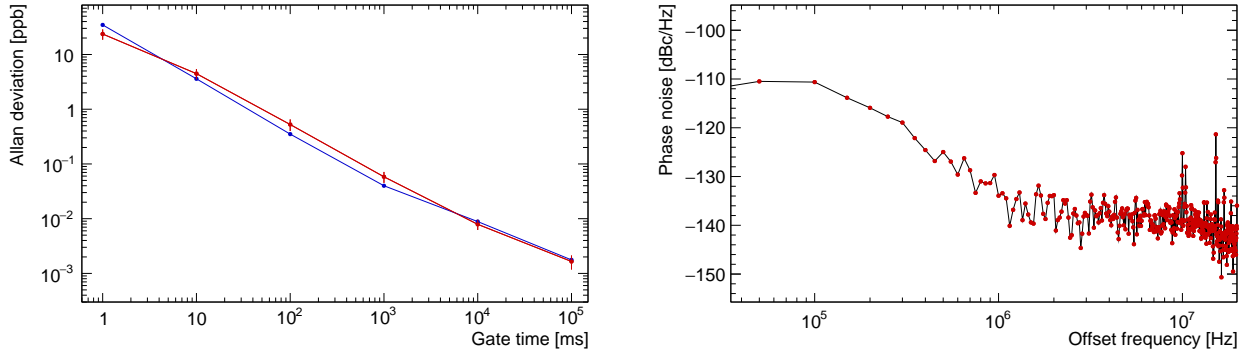
Figure 3.19: The Allan deviation (left) and phase noise (right) of the 40- (blue) and 800-MHz (red) clock, which are consistent with each other. These measure the stability of the sampling clock to about 25 and 0.06 ppb fill-by-fill and cycle-by-cycle, respectively.

ied using a 1-ms gate time—about the duration of a fill—that was timed with muon-fill triggers from the CCC. The resulting $\sigma_y(1 \text{ ms})$ values were found to be consistent with each other within a standard devation—as expected since no fill-dependent load mechanism is known. In principle, the onset of buffering in a channel's DDR3 SDRAM could cause a transient event, but it would be gone well before the muons are injected about 30 μs later.

Such direct stability measurements, however, were prohibited once the 40-MHz clock's frequency was blinded because the ε-offset could be inadvertently revealed. For that reason, a dedicated, clock-monitoring system that is double-blind was designed and deployed alongside the CCC. As diagrammed in Fig. 3.2, a second SG382 signal generator (with the same external timebase) is used to synthesize a 30-MHz clock, which is blinded by a different small offset. The blinded, 30- and 40-MHz clocks are subsequently mixed (Mini-Circuits ZP-1MH+) and filtered (Mini-Circuits BBP-10.7+) to create a 10-MHz clock offset by a third blinding offset—equal to the difference of the original two offsets. Importantly, this clock is sensitive to any relative change in the 40-MHz clock, and ε cannot be deduced from its offset. The primary, 40-MHz clock's stability was, thus, continuously monitored across Run 1 by measuring this mixed clock's frequency using an SR620 with a gate time of 100 s. To keep it independent, the SR620 is supplied with its own external
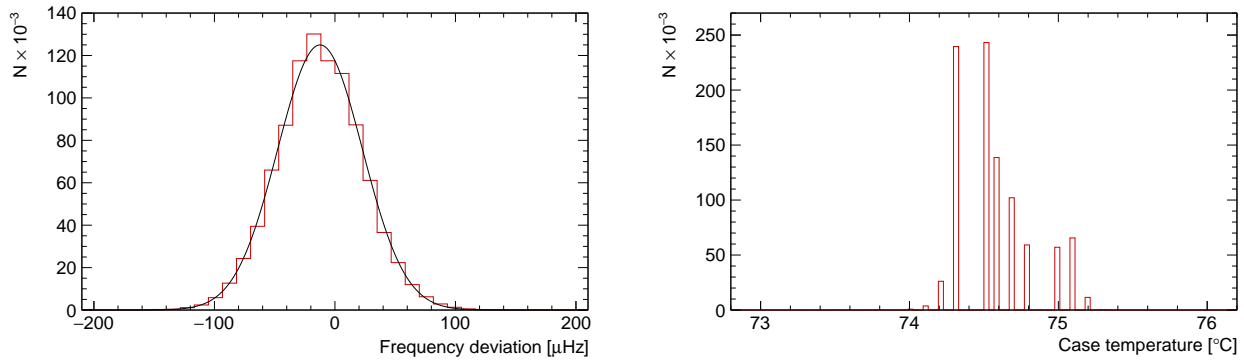
Figure 3.20: Left: The frequency deviation from 9.998888 MHz of the double-blind 10-MHz monitoring clock (Sec. 3.5) across Run 1. Each mean frequency in the histogram was measured using a 100-s gate time. A fit to a normal distribution is overlaid, finding a standard deviation of 35 μHz. The peak-to-peak variation is 377 μHz. Right: The SRS FS725 frequency standard's case temperature across Run 1. This temperature is between the baseplate and lamp temperatures. The discrete readings are due to their quantization.

timebase (SRS FS740/2ABC) with its own GPS antenna (SRS O740ANT2).

As shown in Fig. 3.20, the 10-MHz, monitoring clock's mean frequency throughout Run 1 is well described by a normal distribution. Therefore, its population standard deviation of 35.3 μHz is a reasonable measure of the stability, corresponding to roughly 4×10^{-12}. This is commensurate with the clock equipment's vendor specifications on short-term stability and aging. Further, the peak-to-peak variation of 377.0 μHz corresponds only to 0.04 ppb. This characterization of the clock's stability will be directly used in Ch. 6 to assess a systematic uncertainty to ω_a. To aid in this assessment, the clock equipment's numerous status registers and sensor readings were also monitored throughout Run 1. The FS725 case temperature, for instance, is found to vary by less than 1.2 °C—well within its rated operating regime. All together, the waveform digitizers' clock exhibits excellent characteristics, enabling a precise ω_a-measurement.

CHAPTER 4

RECONSTRUCTION ALGORITHM

Each decay positron that hits a calorimeter produces signals in the waveforms recorded by the backend electronics. To enable the ω_a-measurement, these raw data must be converted into the incident positron's energy and arrival time. The algorithm by which the decay positron's physical quantities, i.e., time, energy, and position, are extracted is referred to as the *reconstruction*. This chapter will thoroughly review the reconstruction algorithm created for the next chapter's ω_a-measurement, followed by its characterization using simulation and data.

The reconstruction was written in C++ within the `art` framework, which has been adopted by the Muon $g - 2$ collaboration [86]. This framework allowed for the reconstruction's steps to be modularized and, thus, more easily maintained. It also allowed for the reconstruction to be configured at run-time via the `FHiCL` language. Much of the code, however, was implemented as object-oriented C++ classes that are independent of `art`. This enabled the reconstruction to be easily deployed in standalone Monte Carlo studies. For reference, the reconstruction algorithm's code is located (and version controlled) in the `gm2reconeast` software package [87].

4.1 Reconstruction Approach

As a positron traverses the PbF_2 calorimeter, an electromagnetic shower develops, generating Čerenkov photons that propagate to the silicon photomultipliers (SiPMs) mounted on the back of the crystals. As seen in Fig. 4.1, around 74 % of these photons arrive at the SiPM within roughly 0.5 ns of each other, while the remaining 26 % arrive much later due to many total internal reflections. Each photon then induces a characteristic electrical response from one of SiPMs' many pixels, as shown in Fig. 4.2. In total, less than about 5 % of these pixels are struck, and, for

(a) $\theta_x = 0$ °.   (b) $\theta_x = 10$ °.
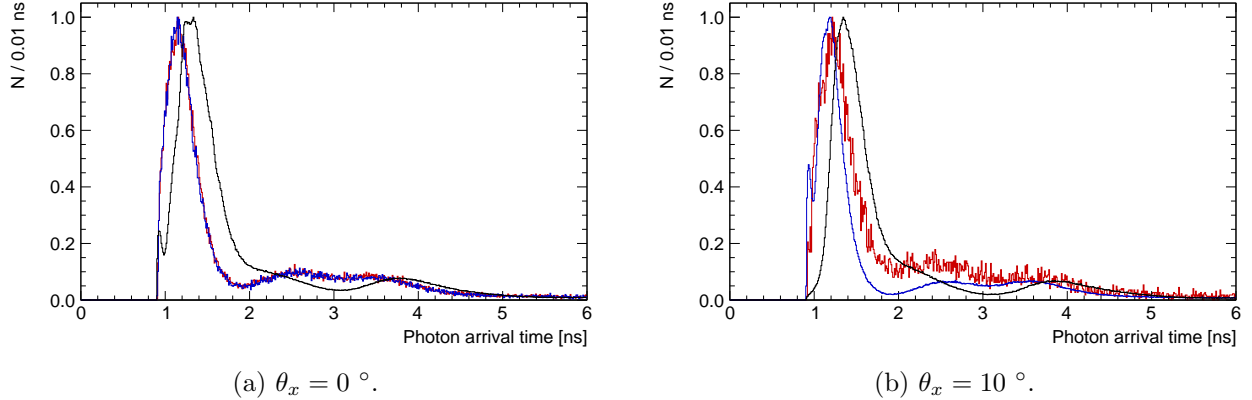
Figure 4.1: The Čerenkov photon arrival time distributions at the SiPM in the central crystal (black), the inner-radius side crystal (red) and the outer-radius side crystal (blue) for a positron striking the central crystal's center at (a) $\theta_x = 0$ °, and at (b) $\theta_x = 10$ ° relative to normal incidence. Each distribution is normalized to its maximum bin content. A positron with $\theta_x > 0$ ° is directed towards an inner radius.

each fired pixel, its response's gain recovers typically on the order of 10 ns [57].

The pixel responses' sum in each SiPM then yields the raw trace recorded by the waveform digitizer. The resulting pulse shape, such as the one shown in Fig. 4.2, is characteristic to each crystal–SiPM pair and scales fairly linearly with the incident positron's energy. This key feature allows each crystal's pulse to be modeled by a *template* pulse shape that scales in amplitude and translates in time. As a result, a reconstruction can be built around fitting each recorded pulse to its crystal's template along with a constant pedestal. The time when the fitted template peaks can be taken as the reconstructed time, and the fitted template's amplitude can be taken—after calibration—as the energy deposited in that crystal.

Further, the positron-induced electromagnetic shower spreads out laterally across the calorimeter, intersecting *multiple* crystals. As the Molière radius of PbF_2 is around 80 % of a crystal width, a single positron typically yields a large pulse in the "central" crystal that it hit and smaller pulses in the "side" crystals adjacent to it. The resulting, spatially isolated pulses are, thus, correlated with each other by the underlying shower. For a positron hit at a given position and angle, the
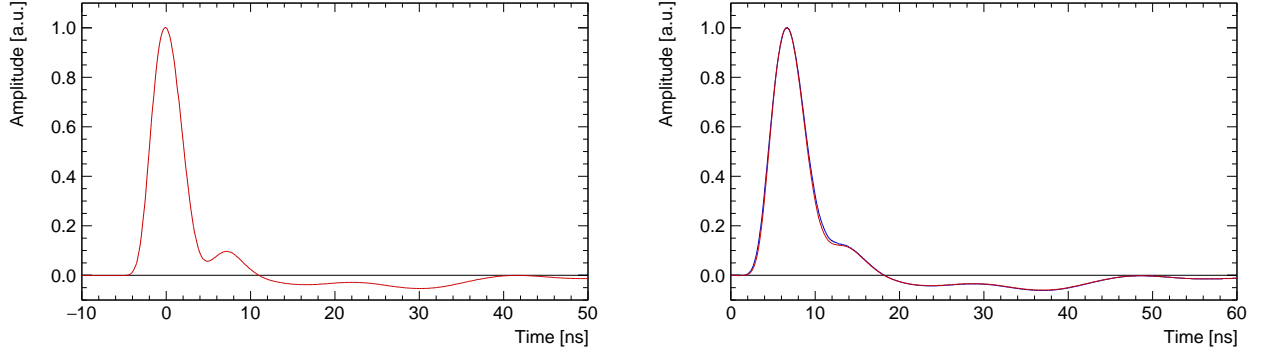
Figure 4.2: Left: The characteristic pulse shape of the pixel response function [87]. Right: The resulting, template pulse shape for the central (red) and side (blue) crystal when a positron is incident normally at central crystal's center.

central crystal's peak time can then be related, on average, to the side crystals' peak times. This further enables the reconstruction to be built around fitting *clusters* of spatially close pulses by incorporating the relative peak-time differences predicted by positron-hit simulations for varying incident positions and angles [88].

For intuition, let's consider that the electromagnetic shower propagates in the calorimeter at the speed of light, c, while its photons created along the way propagate slower at c divided by the refractive index. The Čerenkov photons in a side crystal are, therefore, expected to arrive at the SiPM before those in the central crystal, as indeed observed in Fig. 4.1. Consequently, the side crystal's pulse may be expected to peak before the central crystal's. For normally incident positrons at the crystal's center, however, the side crystal's pulse is found in `Geant4`-simulation to peak about 40 ps *after* the central crystal's pulse, which is reasonably consistent with data [89]. This is due to the substantial tail in the Čerenkov photons' arrival-time distribution, which skews such intuition. Further, as shown in Fig. 4.3, the peak-time difference between adjacent crystals' pulses varies widely with the positron's impact position and, to a larger extent, its angle. For typical positions and angles expected in the Muon $g - 2$ experiment, such peak-time differences can be as large as ± 80 ps—twice the typical peak-time resolution for a single pulse fit.
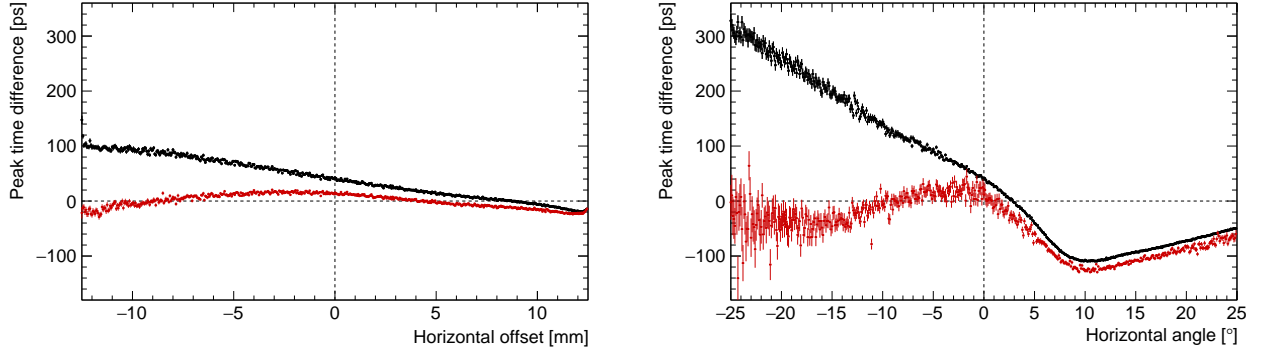
Figure 4.3: The peak-time difference found in simulation between the central crystal and a neighboring crystal as a function of incident horizontal offset from the central crystal's center (left) and incident angle from normal (right). The fitted (true) difference is shown in red (black). A positive difference indicates that the neighboring crystal's pulse peaks afterward. These differences are fairly insensitive to the incident vertical offset from the central crystal's center.

As also seen in Fig. 4.2, the side crystal's characteristic pulse shape differs from the central crystal's pulse shape. The side crystal's pulse tends to be slightly broader because of its unique Čerenkov photon-arrival distribution in Fig. 4.1. Accordingly, the pulse shapes in the central and side crystals are not simple time translations of each other. Therefore, the difference in the times of the peak pulse height extracted from template fits to the central and side crystals' pulses—i.e., the peak-time difference—will not match the true difference for each incident position and angle; the fitted peak time in the side crystal shifts to compensate for the shape distortion. In comparison, as shown in Fig. 4.3, the peak-time differences from such template fits are largely reduced, staying within about ± 30 ps for the largest true differences observed.

From the fitter's perspective, it is, thus, reasonable—at least, as an initial approximation—to claim that the central and side crystal pulses peak at the same time. In general, as long as any inaccuracy remains constant into the fill, there will be no systematic ω_a-bias introduced from it (Sec. 2.6). Advised by these simulation results, this chapter's reconstruction takes a holistic approach by simultaneously fitting multiple crystal traces to their individual template pulse shapes with a shared peak time. As compared to fitting the pulses separately, this novel method changes

the central question from whether there is a pulse *somewhere* in the trace to whether there is a pulse *at this time* in the trace. This is an important difference for identifying some pulses in the side crystals, which can be very small relative to the central crystal's pulse. By combining information among the crystals, this algorithm, which will be expounded in the following sections beginning with its required inputs, can also better estimate an incident positron's time of arrival, as shown in Fig. 4.4, along with other improvements discussed in Sec. 4.8.

4.2 Island Selection

For each muon fill, the waveform digitizer—once triggered—records a continuous trace of $560\,000$ samples (approximately 700 μs) for every crystal composing a calorimeter. As the average fill rate is about 11.4 Hz, however, these data cannot all be written onto disk. If each sample is stored as an object of type short, this would result in roughly a 15.4 GB/s data rate—beyond the resources available. Therefore, while the entire muon fill is read out, we must select only the times of interest for the ω_a-measurement to permanently save. In particular, the data acquisition (DAQ) identifies pulses in the raw trace by looking for samples above a given threshold and saves only the times around them. Such time segments chopped out of the trace are referred to as *islands*, making the underlying process known as the *island chopper* [90].

As the DAQ is optimized for speed by utilizing graphics-processing units, each calorimeter's raw traces are read out and processed in parallel. For each calorimeter, the island chopper begins by estimating every crystal's pedestal, P, once per fill by averaging the first few sample values of its raw trace. The method then marches through the raw traces, comparing each sample value, y, above its crystal's pedestal, i.e., $y - P$, to a threshold of 60 analog-to-digital units (ADU) of the waveform digitizer's analog-to-digital converter (ADC). This threshold, corresponding to roughly 60 MeV, sets the amount of energy that the electromagnetic shower must deposit in a crystal for

the island chopper to be triggered. Whenever this condition is satisfied in any one crystal, all of that calorimeter's 54 raw traces are saved for 8 samples before and 16 samples after the one over the threshold. These samples are referred to as *pre-* and *post*-samples, respectively. Moreover, if two or more islands' samples overlap in time, they are merged into a single island such that all islands are non-intersecting by the end.

The typical island length is 32 samples—about 40 μs. The length of any given island is dynamic, though, due to the island-extension logic above. The mean island length, for example, grows with the rate of positron hits as a pair of them arriving close in time becomes more likely. These time islands form the basic input to the reconstruction algorithm. They deliver a crude pulse separation, providing small chunks of data in which the fit algorithm in Sec. 4.5 can find positrons. As island chopping is an irreversible process, however, its threshold sets a fundamental limitation to the lowest-energy positron likely to be found in temporal isolation. As well, it should be noted that each calorimeter's islands are time ordered—herein tacitly assumed—by precursor software that unpacks them into a form the reconstruction can read.

4.3   Template Construction

With such island data to fit, the reconstruction's next key input is each crystal's template pulse shape, as motivated in Sec. 4.1. As they are vital to this algorithm's approach, let's review how the templates were empirically constructed for Run 1 [91]. The basic procedure is conceptually straightforward. After collecting a large number of pulses for a given crystal, they are normalized to their integral and aligned in time to the pulses' peak. In fine time bins, the sample values are then averaged and interpolated to yield the crystal's continuous template, $S(t)$.

More specifically, the island chopper was temporarily reconfigured to save longer islands with 75 pre- and 450 post-samples, allowing for the entire ringing of the pulse shape's tail to be ob-
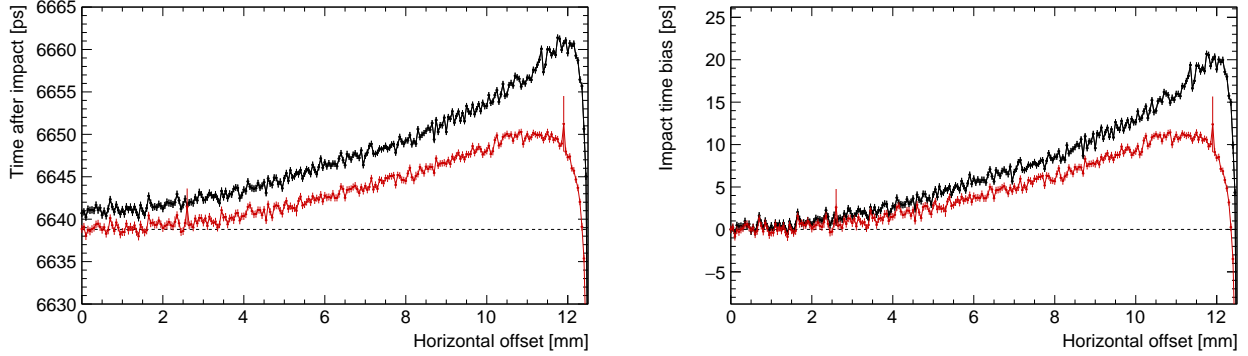
Figure 4.4: The reconstructed peak-time found in simulation by fitting the central crystal alone (black) and, as described in Sec. 4.5, by fitting a 3×3 crystal window about the central crystal (red) for incident horizontal offsets between the central crystal's center to its edge. The times in the right plot are offset by the value when incident at the central crystal's center.

served. Islands were then collected on March 10–14, 2018, where those chopped earlier than 30 μs after injection were ignored to make the chance of pileup occurring sufficiently small. Each island, therefore, will have at most one pulse per crystal. After correcting for the even/odd pedestal difference described in Sec. 4.5, each crystal's pedestal was estimated by the average of the first 20 samples in its trace. To select only large pulses well above the noise floor, any crystal trace with a maximum sample value less than 400 ADU from its pedestal was neglected. In total, about 10^3–10^5 pulses per crystal passed these selection criteria. Each remaining trace was then normalized to have unit area by dividing their sample values above the pedestal by the Riemann sum of the 20 pre- and 180 post-samples about the peak sample [91].

The complexity in the template construction comes from how to suitably align the pulses in time so that the averaged pulse shape is not distorted. Aligning them simply to their peak sample, for example, is inadequate as the waveform digitizer's sampling period of 1.25 ns is an appreciable fraction of the pulses' typical rise/fall time of 2–4 ns. As a consequence, the offset between each pulse's peak sample and true peak time, δt, must be estimated. This can be done by considering not only the maximum sample value, y_i, but also the sample values immediately before and after it, y_{i-1} and y_{i+1}, respectively. As the true peak time varies later in time, for instance, y_{i+1} will in-
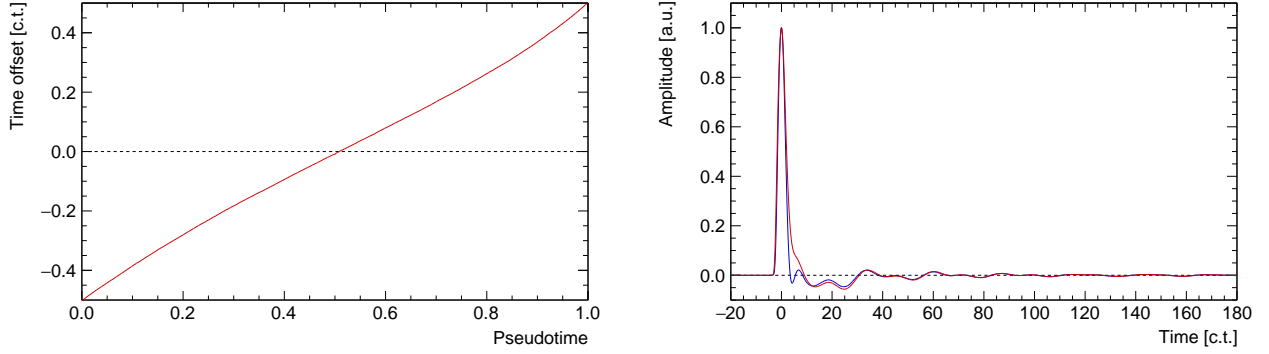
Figure 4.5: Left: The mapping function from pseudotime, t_p, to the offset between the true peak time and the peak sample, δt. Right: The resulting, representative template shape constructed from positron (red) and laser (blue) pulses. For reference, 1 c.t. $\approx$ 1.25 ns.

crease relative to y_{i-1}. The relationship among these sample values can then be encapsulated in a quantity known as the *pseudotime*, formally given by

$$t_p = \frac{2}{\pi} \cdot \arctan \left[\frac{y_i - y_{i-1}}{y_i - y_{i+1}} \right] \tag{4.1}$$

which was originally conceived for the previous experiment's reconstruction [92]. It is important to note that, in Eq. 4.1, the pulse's pedestal and scale cancel out via the samples' subtraction and division, respectively. This yields a proxy for the pulse's true peak time, varying smoothly in the interval $[0, 1]$. The distribution of t_p from the collected pulses, however, is not guaranteed to be uniform—as would be expected of the true peak time relative to the clock boundary. From that expectation, t_p can then be mapped to δt by

$$\delta t(t_p) = \int_0^{t_p} \rho(t_p') \, \mathrm{d}t_p' - \frac{1}{2} \tag{4.2}$$

where $\rho(t_p)$ is the normalized t_p-distribution observed for a given crystal. An example of such a mapping is given in Fig. 4.5, which will always be a monotonically increasing function. Further, it should be noted that, when two adjacent samples have the same maximum value, Eq. 4.2 tacitly defines the true peak to be halfway between their sampled times, i.e., $\delta t(0) = -0.5$ c.t. This may be incorrect for an asymmetric pulse shape. In that case, Eq. 4.2 should be offset by a constant,

redefining $\delta t(0)$. Instead, any such offset is neglected in this procedure since, as will be seen in the next section, it will be included as part of the time-correction method.

After aligning a crystal's collected pulses according to Eq. 4.2, the sample values were averaged in 1/21-c.t. time bins from -20 to 180 c.t., which were then fit to cubic splines using the `UnivariateSpline` routine in the `SciPy` software package [93]. This fit is taken as the crystal's template with its value outside the constructed time range defined to be zero. As compared to a simple interpolation between the finely binned data, such an involved approach was necessary to avoid overfitting statistical fluctuations. Finally, the template function $S(t)$ was normalized to its (approximate) peak value at $t = 0$ so that its scale factor is in ADU—a more intuitive unit than ADU $\cdot$ c.t. A representative pulse template is given in Fig. 4.5.

The assumption underlying this template construction—and, more generally, the fit algorithm in Sec. 4.5—is that the pulse shape is independent of scale, time, and positron-hit rate. This has been verified by constructing each crystal's template under a variety of conditions to probe the shape's stability [91]. As expected from Fig. 4.2, however, the pulse shape in the central and side crystal was found to differ slightly. With that in mind, each template was purposefully formed from pulses without regard to whether or not it was the central crystal. The resulting template, therefore, is the mean pulse shape observed—weighted appropriately by the central and side pulse shapes, which should yield the least time/energy bias. Further, as most of an incident positron's energy is deposited in the central crystal, neglecting such small positron-dependent pulse shape differences is expected to be sufficient for an ω_a-measurement.

4.4   Time Correction

As will be described in Sec. 4.5, the reconstruction is built around a simultaneous fit to pulses in nearby crystals that are modeled to peak in unison. To that end, all of the crystals' traces from

a given calorimeter must be time-aligned to an absolute reference prior to the fit. That is, a time correction for each crystal's trace must be initially found, which may vary from one muon fill to the next. For example, phase advances of the waveform digitizer's 800-MSPS sampling clock can occur between fills, introducing possible 1, 2, 1/3, or 2/3 c.t. offsets (Sec. 3.3). To arrive at the full time correction needed, all of the traces from the same calorimeter will be aligned to each other. Afterward, they will be further aligned to the time that the muons are injected into the storage ring, thereby making all of the traces among calorimeters also time-aligned.

For a given calorimeter, the relative time alignment is accomplished via the laser-calibration system, which fires a laser pulse into each crystal about 25 μs before every fill's muon injection. In particular, each *sync pulse* originates from one of six laser heads, whose light is split and then transmitted to four different calorimeter stations by optical fiber [94]. At each station, the light is diffused nearly uniformly across 54 shorter fibers—ranging in length approximately from 45 to 65 cm—that finally transmit it to each crystal's front face via reflecting prisms [95]. Because a single laser head sources the sync pulse for every crystal in the calorimeter, its digitized waveform can, therefore, serve as a common reference. Aligning the calorimeter's sync pulses thereby aligns its crystals' traces with respect to when its laser head fired up to the different propagation delays due to each crystal's unique fiber length.

As the fiber lengths were not measured before being installed, the differences in their propagation delays were empirically estimated by comparing the times when pulses were recorded in different crystals' traces. More concretely, the peak time of a pulse in crystal i due to the laser shot or a positron hit can be modeled as a reference time plus the various delays:

$$t_i^{\text{laser}} \quad = \hat{t}_i^{\text{laser}} \quad + \delta_i^{\text{electronics}} + \delta_i^{\text{fiber}}, \tag{4.3}$$

$$t_i^{\text{positron}} = \hat{t}_i^{\text{positron}} + \delta_i^{\text{electronics}} + \delta_i^{\text{shower}} \tag{4.4}$$

respectively, where $\delta_i^{\text{electronics}}$, δ_i^{fiber}, and δ_i^{shower} are the delays from the electronics, e.g., clock
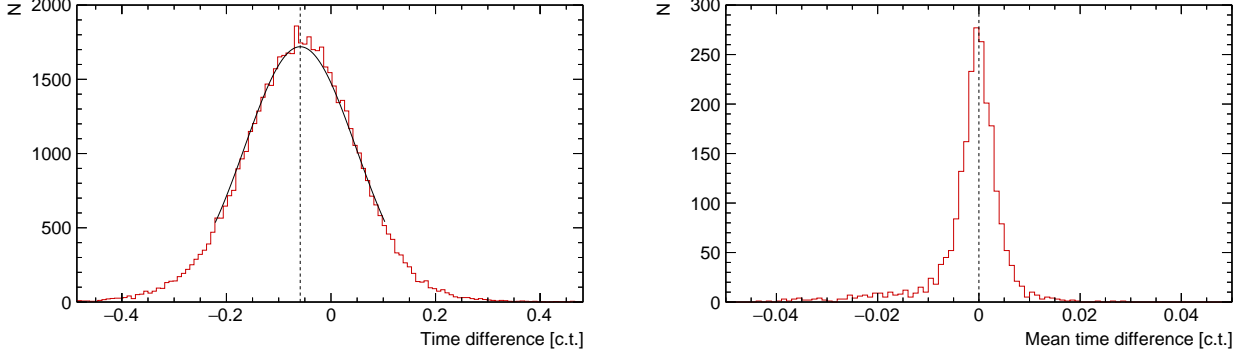
Figure 4.6: Left: A representative distribution of $\Delta_{i,j}$. A normal-distribution fit to ± 1.5 of the population standard deviation about the population mean is overlaid (black). Right: The time difference of the mean $\Delta_{i,j}$ measured for crystals i and j from its $\delta_i - \delta_j$ best-fit values found by solving Eq. 4.7. Both plots were made using the 1a data set (Sec. 5.1).

slips; the light propagation in the fiber from the diffuser to the crystal; and the peak-time differences between neighboring crystals, as shown in Fig. 4.3. The reference times $\hat{t}_i^{\text{laser}}$ and $\hat{t}_i^{\text{positron}}$ can be treated, respectively, as when the trigger signal arrived at the laser-calibration system and at the crystal's waveform digitizer. When the peak times are subtracted, as in

$$t_i = t_i^{\text{positron}} - t_i^{\text{laser}}$$

$$= \hat{t}_i^{\text{positron}} - \hat{t}_i^{\text{laser}} + \delta_i \tag{4.5}$$

where $\delta_i = \delta_i^{\text{shower}} - \delta_i^{\text{fiber}}$, any common delays—such as from the electronics—cancel out. What remains is the difference in the delay from the fiber length and that originating from the electromagnetic shower. Furthermore, for a given positron hit, the difference in the sync-pulse-aligned time for horizontally or vertically adjacent crystals i and j is given by

$$\Delta_{i,j} = t_i - t_j$$

$$= \hat{t}_i - \hat{t}_j + \delta_i - \delta_j$$

$$= \delta_i - \delta_j \tag{4.6}$$

where $\hat{t}_i = \hat{t}_i^{\text{positron}} - \hat{t}_i^{\text{laser}}$. In the last equality in Eq. 4.6, the reference time in both crystals is

imposed to be equal—as expected by definition—which isolates the crystals' delays. With respect to a reference delay of $\delta_0 \equiv 0$, the remaining ones can then be determined.

While many methods have been developed to find the optimal δ_i for each crystal, the most robust way begins by populating a histogram for each of the 93 unique $\Delta_{i,j}$ differences—48 horizontally and 45 vertically [96]. To do that, each pulse's peak time is taken from a template fit to its crystal's trace alone. The fit is performed by minimizing a χ^2 similar to that described in Sec. 4.5 using Newton's method [91]. It should be noted the sync pulse fit used its own "laser" template (Fig. 4.5), which was constructed with the procedure described above for positrons. A representative $\Delta_{i,j}$-distribution is shown in Fig. 4.6. The peak of each $\Delta_{i,j}$-histogram is then fit to a normal distribution to extract the mean $\Delta_{i,j}$ and its uncertainty $\sigma_{i,j}$. Using these best-fit values, a system of equations can be written as

$$
\begin{bmatrix}
-1/\sigma_{0,1} & 0 & \cdots & 0 \\
1/\sigma_{1,2} & -1/\sigma_{1,2} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & -1/\sigma_{44,45}
\end{bmatrix}
\begin{bmatrix}
\delta_1 \\
\delta_2 \\
\vdots \\
\delta_{53}
\end{bmatrix}
=
\begin{bmatrix}
\Delta_{0,1} \ /\sigma_{0,1} \\
\Delta_{1,2} \ /\sigma_{1,2} \\
\vdots \\
\Delta_{44,53}/\sigma_{44,45}
\end{bmatrix}
\tag{4.7}
$$

where each $\Delta_{i,j}$-equation from Eq. 4.6 is weighted by the fit precision. Because this system is overdetermined, it was solved for all 52 δ_i delays via singular value decomposition in the `Eigen` software package [97]. The δ_i values found make the mean reference-time difference among the adjacent crystals as close to zero as possible. As seen in Fig. 4.6, this alignment is achieved for most crystal pairs to within about 6 ps.

A key advantage of this method is that it is independent of the reference crystal chosen. It also yields consistent δ_i values to about 1–2 ps when using different data sets or time cuts. Further, any (much smaller) time differences from the template pulse shapes, as discussed in Sec. 4.3, will be absorbed into the δ_i delays found. Varying amounts of δ_i^{shower} will be absorbed as well, depending
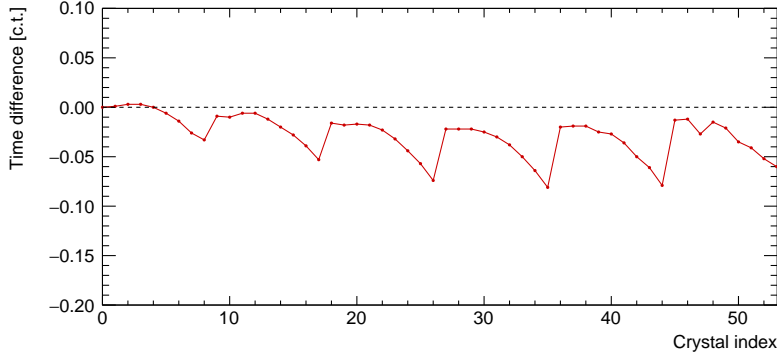
Figure 4.7: The difference in δ_i for crystal i in calorimeter 1, found via Eq. 4.7, between when requiring the pulse amplitudes to be within 10 % of each other and when not. The crystals are indexed from the bottom to top row, where each row is from away to towards the muon storage region. For reference, 1 c.t. $\approx$ 1.25 ns.

on the positron-pulse selection in the $\Delta_{i,j}$ histograms. For example, requiring the pulse amplitudes to be close to each other restricts the allowed position–angle range for the incident positron, thereby imposing a more constant time difference due to the electromagnetic shower. In this way, the δ_i^{shower} contribution to δ_i can be largely removed. As shown in Fig. 4.7, the δ_i-difference when the pulse amplitudes are required to be within 10 % of each other can be as large as -100 ps—in agreement with simulation (Fig. 4.3).

In summary, for a given calorimeter, the trace of crystal i can be time-aligned to that of crystal 0 by subtracting its first sample's index by its sync pulse's time and δ_i. While such alignment is sufficient for the fit algorithm in Sec. 4.5, the traces *among* the calorimeters also need to be time-aligned for the best ω_a-fit to the calorimeter-combined time histogram (Sec. 2.5). Otherwise, the calorimeter's muon-precession signals may deconstructively interfere, worsening the precision on the extracted value of ω_a. This inter-calorimeter time alignment is achieved by aligning them to when the muons were injected into the storage ring for each fill.

Aside from the $\mathcal{O}(1 \text{ ns})$-differences among the crystals, their raw waveforms begin around 31 μs before the muon injection so that their sync pulses can be recorded. Each fill's offset, however, is different due to trigger delays. The accelerator's trigger, for instance, has a precision of 1–2 ns

and can have an internal delay of up to 3 μs. Additional fill-dependent delays arise from the trigger handoff between the accelerator and the Clock and Commands Center (CCC) (Sec. 3.2) and between the CCC and the laser-calibration system. These systems' asynchronous clocks allow for a total delay up to 35 ns.

For a given fill, the injection-time alignment is attained using the "T0" entrance counter—a plastic scintillator placed in front of the inflector magnet [38]. The light produced is detected with a photomultiplier tube, whose signal is recorded by a spare waveform digitizer in calorimeter 22's crate. As illustrated in Fig. 4.8, the T0 counter measures the time distribution of each fill's muons entering the storage ring, whose shape is unique to each of the eight muon bunches in FNAL's accelerator cycle (Sec. 2.2). The muon injection time is then defined as the T0 pulse's mean time. To reference this time to the crystals' traces in calorimeter 22, the same sync pulse is also delivered to the T0 counter via a spare fiber. In an analogous way, the time of the sync pulse and the T0 pulse's mean can then be modeled as a reference time plus the various delays:

$$t_{\text{T0}}^{\text{laser}} = \hat{t}_{\text{T0}}^{\text{laser}} + \delta_{\text{T0}}^{\text{electronics}} + \delta_{\text{T0}}^{\text{fiber}}, \tag{4.8}$$

$$t_{\text{T0}}^{\text{mean}} = \hat{t}_{\text{T0}}^{\text{mean}} + \delta_{\text{T0}}^{\text{electronics}} + \delta_{\text{T0}}^{\text{signal}} \tag{4.9}$$

respectively, where $\delta_{\text{T0}}^{\text{fiber}}$ and $\delta_{\text{T0}}^{\text{signal}}$ are the delays from the light propagation from the diffuser to the T0 counter and the signal propagation from the scintillator to the waveform digitizer. Their fill-independent time difference is then given by

$$t_{\text{T0}} = t_{\text{T0}}^{\text{mean}} - t_{\text{T0}}^{\text{laser}}$$

$$= \hat{t}_{\text{T0}}^{\text{mean}} - \hat{t}_{\text{T0}}^{\text{laser}} + \delta_{\text{T0}}^{\text{signal}} - \delta_{\text{T0}}^{\text{fiber}} \tag{4.10}$$

which is offset by a constant delay. Any shift common to all calorimeters, however, only redefines the anomalous precession's phase—with no ω_a-bias introduced. Accounting for crystal 0's fiber length in calorimeter 22, such a global shift is estimated to be less than about 10 ns. In sum, for
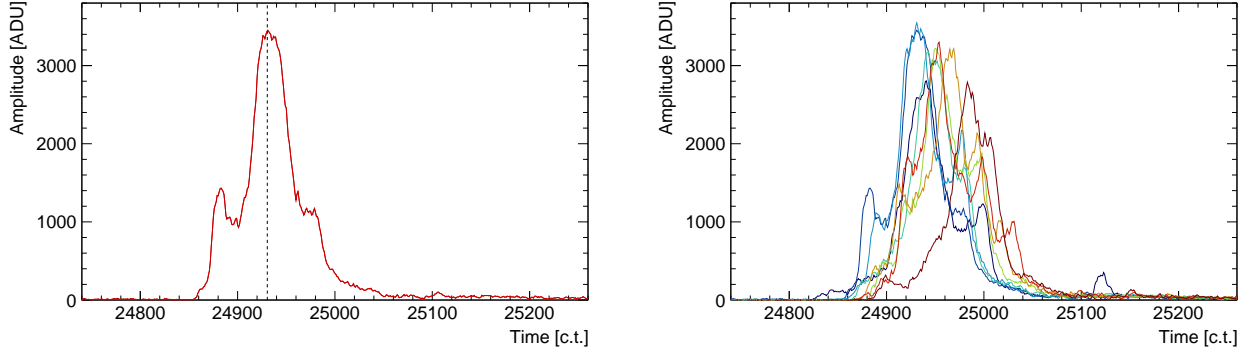
Figure 4.8: Left: An example T0 pulse for bunch 2. The pulse's mean time used for injection-alignment is denoted by a vertical dashed line. Right: Representative T0 pulses for each of the eight bunches in FNAL's accelerator cycle (Sec. 2.2), with each one denoted by a different color from red to blue in order of the rainbow.

each fill, the start of each crystal's trace in calorimeter 22 can be further time-aligned to the muon injection by subtracting its first sample's index by t_{T0}.

Applying the same subtraction to other calorimeters' traces will also time-align them to the muon injection as long as they agree on their reference times in Eqs. 4.3 and 4.4. However, they will disagree due to $\mathcal{O}(1\ \mathrm{ns})$ constant differences among the laser heads' and waveform digitizers' trigger-signal paths. As a final, calorimeter-dependent correction, these time differences were empirically determined by measuring the time of flight of muons passing through adjacent calorimeters. After injection, some of the muons will inevitably exit the storage region—for example, by scattering off an object such as one of the beam collimators. These *lost muons* will often curl inward, leaving a distinct energy signature in a calorimeter as will be discussed further in Sec. 4.6. As the stopping power for $\mathcal{O}(1\ \mathrm{GeV})$ positive muons in $\mathrm{PbF_2}$ is small enough, they will occasionally continue to pass through another two calorimeters. Assuming that the calorimeters are equidistant—which is sufficiently true—the time between the lost muons' calorimeter hits should be the same and can, thus, serve as an absolute reference time for alignment.

To that end, the fit algorithm in Sec. 4.5 was initially performed without an inter-calorimeter time correction. Triple coincidences—i.e., clusters reconstructed in three consecutive calorimeters

within 26 c.t. of each other—were then identified. The hit-time difference between each triple

coincidence's two pairs of adjacent calorimeters i and $i+1$ can be expressed as

$$\Delta_{i,i+1} = t_i - t_{i+1}$$

$$= \hat{t}_i - \hat{t}_{i+1} + \delta_i - \delta_{i+1}$$

$$= \alpha + \delta_i - \delta_{i+1} \tag{4.11}$$

where δ_i is the net delay offset on calorimeter i's reference time. In the last equality above, the

true time difference between every pair of adjacent calorimeters, α, is imposed to be the same.

To measure each δ_i, the $\Delta_{i,i+1}$ values from the triple coincidences identified after 30 μs into the

fill—when the muons have largely debunched—were histogrammed, as shown in Fig. 4.9. Each

$\Delta_{i,i+1}$-distribution was then fit to a normal distribution plus a constant that models the uniform

rate expected for accidental coincidences of positrons. Using the best-fit peak times, a system of

24 $\Delta_{i,i+1}$-equations via Eq. 4.11 can be written and, by defining $\delta_{22} \equiv 0$ as the reference delay, α

and the remaining 23 δ_i-delays can be solved for exactly. The resulting δ_i-values for Run 1 were

reproducible to 0.06 ns among different data sets, and, as seen in Fig. 4.9, α was measured to be

approximately 6.15 ns—roughly as expected [98].

In review, for a given fill, every crystal's trace in each calorimeter can be time-aligned to each

other and approximately to when the muons enter the storage region by subtracting the four types

of offsets discussed. In total, the time correction—to be applied in the next section—for the trace

of crystal c in calorimeter m is given by

$$\varepsilon_{c,m} = -t_c^{\text{laser}} - \delta_c - t_{\text{T0}} - \delta_m \tag{4.12}$$

where δ_c and δ_m are the intra- and inter-calorimeter time corrections found based on Eqs. 4.6 and

4.11, respectively. If the time of the crystal's sync pulse, the T0 counter's sync pulse, or the T0

counter's muon pulse cannot be found for any reason, the time correction is flagged as *invalid*,
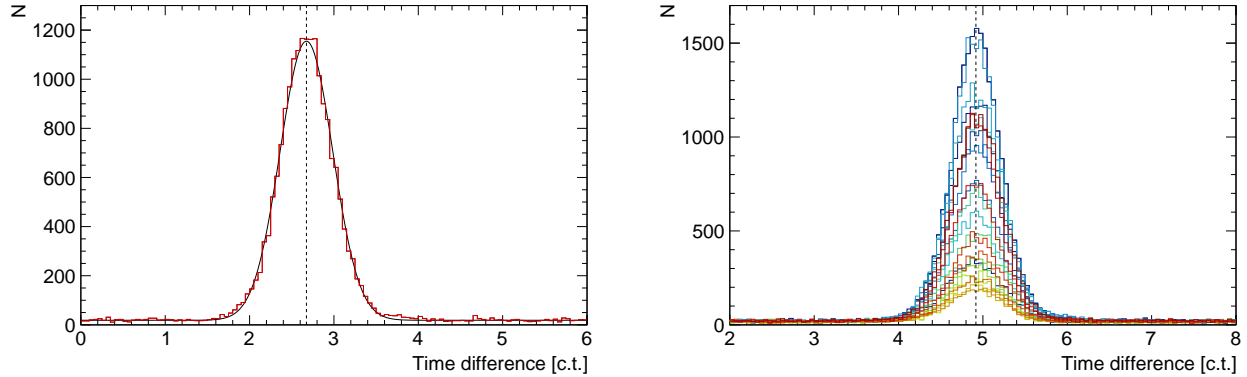
Figure 4.9: Left: A representative $\Delta_{i,i+1}$-distribution before applying an inter-calorimeter time correction. A fit to a normal distribution with a constant background is overlaid (black), with a vertical line at the peak. Right: The $\Delta_{i,i+1}$-distribution from the 1a data set (Sec. 5.1) after time alignment for each calorimeter i, with each one denoted by a different color from red to blue in order of the rainbow [98]. The mean time separation is $\alpha = 4.9171 \pm 0.0091$ c.t.

causing the crystal's trace not to be fit in the next section.

4.5   Fit Algorithm

The kernel of the reconstruction is its fit algorithm, which, with the above pulse templates and time corrections, reduces each island into a list of each positron's time, energy, and position. This algorithm was developed on the simulation introduced in Sec. 4.8 and was then refined on data taken during Run 1. Illustrated in Fig. 4.10, its general flow consists of nine key steps, which will be expounded in turn in the following subsections. Throughout, the default configuration used for Run 1 will be quoted, but much of the algorithm can be reconfigured at run-time, which enabled many of its systematic uncertainty checks (Ch. 6).

Before presenting the minutiae of the fit algorithm, however, a brief summary is in order to provide context. In broad terms, it starts by filtering out any island chopped in Sec. 4.2 because of the laser. Each island's 54 traces, one per crystal, are then prepared by applying the time corrections and removing the waveform-digitizer effects. Next, the algorithm enters a loop to identify and fit for every positron in the island one-by-one in order of decreasing energy. In each iteration,
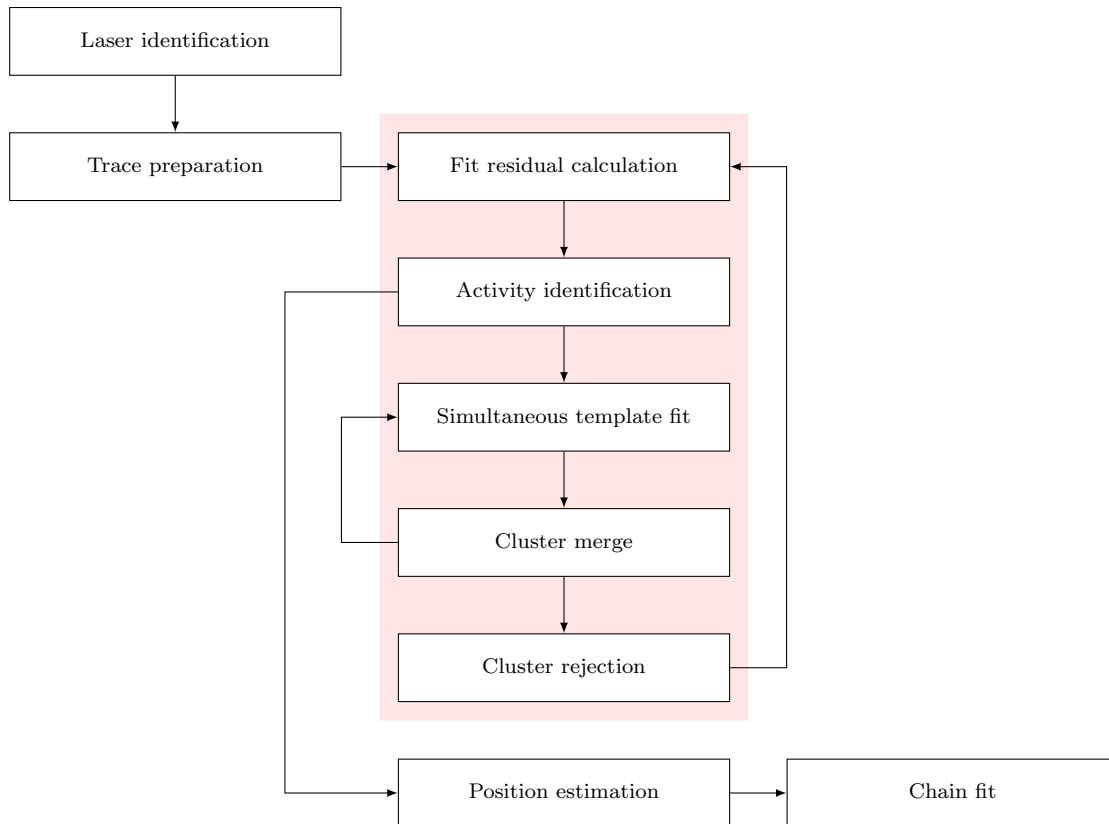
Figure 4.10: A flow diagram of the fit algorithm's key steps, discussed in Sec. 4.5. The crux of the algorithm is the loop between the fit residual-calculation and cluster-rejection steps (red), where clusters are identified and then fit one-by-one in approximately decreasing energy.

the fit residuals are analyzed to decide if there is enough activity in a 3×3 cluster of crystals to indicate another positron, and, if so, they are fit together with a shared peak time. The fit result is then scrutinized to decide whether to merge it into an already found positron's cluster or whether to keep it at all. When no more activity is found in the island, the position of each positron identified is determined. As a final, small refinement step, the positrons are then refit incorporating the pulse templates' tails, if any, from earlier islands.

4.5.1 Laser Identification

In addition to the sync pulse prior to injection, the laser-calibration system also fires a short-duration pulse of light into each crystal at the end of every fill. As seen in Fig. 4.11, such laser
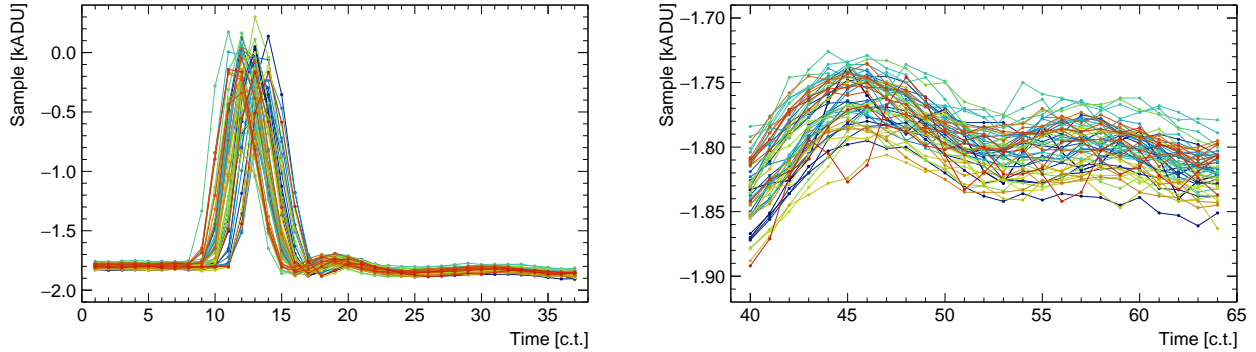
Figure 4.11: Each crystal's trace in two subsequent islands chopped at the end of a fill due to the laser firing, with each crystal denoted by a different color from red to blue in order of the rainbow. The traces have not been time corrected.

firing causes an island to be chopped with a large signal—typically half of the full scale—in each crystal. In some cases, a laser pulse's tail is also large enough to trigger a second island shortly thereafter. This reconstruction algorithm, however, is not intended to fit such laser pulses. As each crystal is optically isolated due to their wrapping, these laser pulses are largely uncorrelated among each other, making a multi-crystal fit to them, in practice, disadvantageous. The fit algorithm's initial step, therefore, is to identify and filter out these laser-triggered islands.

To do that, for each crystal trace, the maximum sample value, y_d, is subtracted by the average sample value—an estimate for the trace's pedestal, P. This difference, $y_d - P$, is treated as the crystal's signal. To ascertain how significant it is, the signal is compared to that expected if the trace had only been composed of normally distributed noise, i.e., no laser pulse. In such a noise-only trace, the maximum sample value from the pedestal, y_n, and its uncertainty, σ_n, can be computed via normal order statistics based on the trace length and the previously characterized noise level for the crystal, illustrated in Fig. 4.12 [99]. The comparison is done by subtracting y_n from the signal and then dividing by σ_n to yield the significance of the signal over the noise. A crystal is then defined to be "over threshold" if

$$\frac{(y_d - P) - y_n}{\sigma_n} \geq N_n \tag{4.13}$$
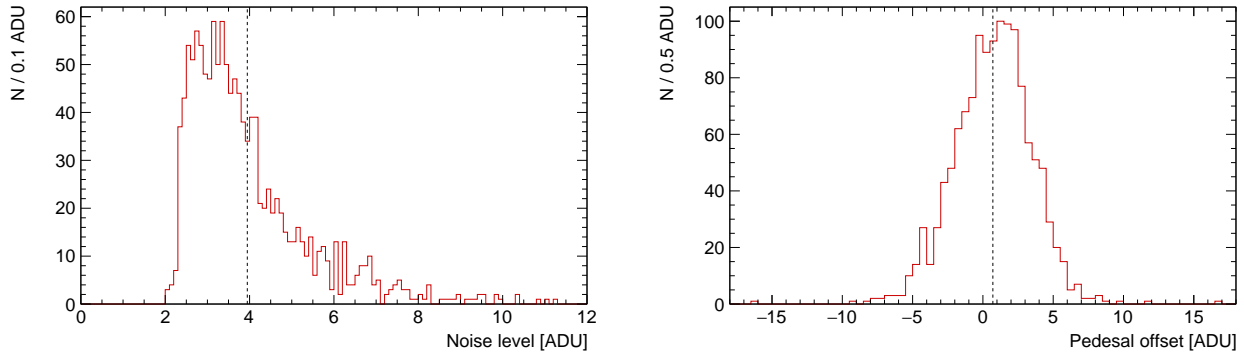
Figure 4.12: The noise level (left) and pedestal difference (right) for each waveform digitizer deployed in Run 1, which were found from the first 200 samples of out-of-fill islands (Sec. 4.6) in the 1a data set (Sec. 5.1) [87]. The noise level is defined as the sample values' population standard deviation; the pedestal difference is defined as the pairwise sample differences' mean.

where $N_n = 2$, by default. If the number of over-threshold crystals is ≥ 44, by default, the island is inferred to include laser pulses and is, therefore, skipped. This procedure has been found to be 100 % and slightly less than 100 % efficient at filtering out laser and laser-tail islands, respectively. It must also be noted that, through this procedure, *any* activity across over 80 % of the calorimeter will cause the island to be skipped. However, as seen in Fig. 4.13, ordinary positron-triggered islands have never been found to yield such large-scale activity.

4.5.2 Trace Preparation

The every crystal's trace in the island is next prepared for the subsequent steps by first applying the crystal's time correction that was calculated for the fill (Sec. 4.4). To do that, the trace's first sample index—the number of the first sample in the island, counting from the start of the raw digitized waveform—is added to Eq. 4.12. This aligns it in absolute time, allowing for it to be sensibly fit together with neighboring crystals. If, however, the crystal's time correction is invalid, this required time alignment cannot be achieved, and, therefore, its trace is removed. The primary consequence of a crystal's trace being removed is a change in the calorimeter's acceptance, which can impact the observed muon dynamics (Sec. 5.6). Further, if every crystal's time correction in
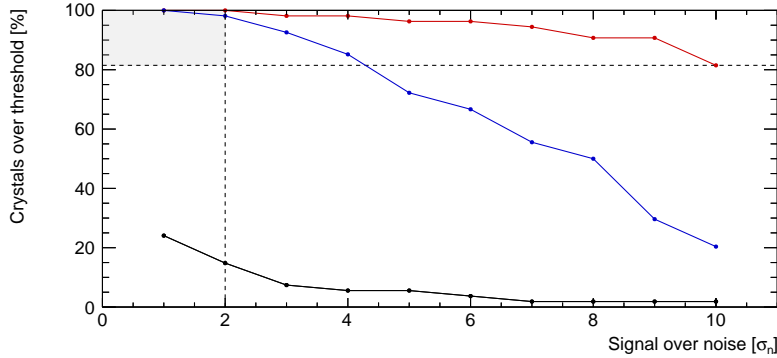
Figure 4.13: The percentage of crystals passing an N_n-threshold in Eq. 4.13 for two laser-tail islands (red, blue) and a representative positron-hit island (black). By default, the islands in the shaded region are inferred to include laser pulses and, thus, skipped.

the calorimeter is invalid, the island is skipped altogether.

Afterward, the trace's sample values are modified to account for effects of the waveform digitizer, starting with saturated samples. In general, for a N-bit ADC in two's complement format, the possible integers represented are from -2^{N-1} to $2^{N-1} - 1$. Thus, given the WFD5's 12-bit ADC (Sec. 3.3), the minimum and maximum sample value that should be observed is -2048 and 2047 ADU, respectively. When the input signal is larger than its 1-V_{pp} range, the ADC is said to have *saturated*, outputting one of these extreme sample values as shown in Fig. 4.14. As such saturated samples do not hold physical meaning, they are removed from the trace. However, in practice, some traces have been found to saturate at 2046 ADU—one lower than expected. To be careful, therefore, any samples beyond ± 2045 ADU are also removed.

In addition, the WFD5's ADC is composed of two internally interleaved ADCs that share a track-and-hold circuit (Sec. 3.3.2). Within this circuit, non-ideal FET switches can cause channel charge injection, resulting in a small pedestal and gain difference between the odd- and even-numbered samples. The odd/even pedestal difference—shown in Fig. 4.12—roughly follows a normal distribution about 0.709(76) ADU with a width of 2.76 ADU. The largest difference found is -25.12 ADU for crystal 13 in calorimeter 22, which is evident in its traces, as seen in Fig. 4.14.
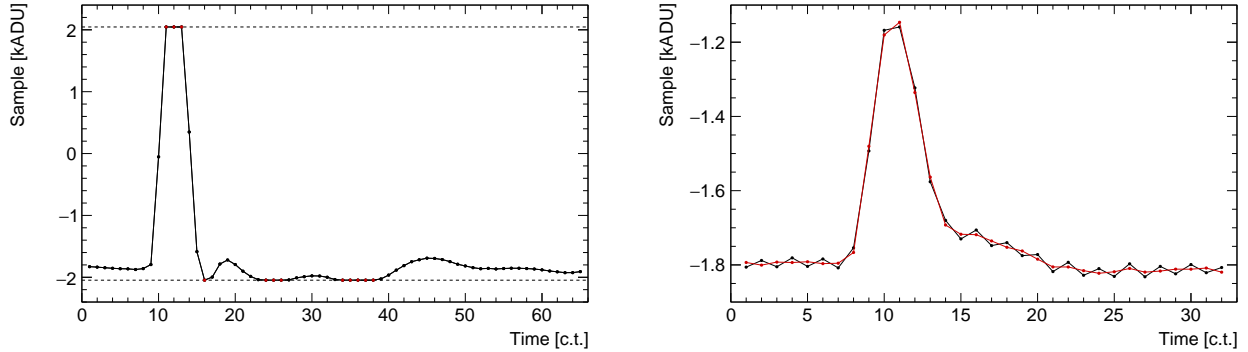
Figure 4.14: Left: An example trace from data where its samples in red are saturated and, thus, will not be fit. The bounds of possible values are denoted by dashed lines. Right: A representative trace from crystal 13 in calorimeter 22, which has the largest even/odd pedestal difference of -25.12 ADU. The uncorrected (corrected) trace is shown in black (red).

Monte Carlo studies, as illustrated in Fig. 4.15, show that such an effect will significantly distort the fitted time's resolution but not bias its mean value by more than 2 ps—a negligible time shift. However, even if this bias was large, no ω_a-bias is expected as the odd/even pedestal difference should remain unchanged throughout the fill. Nonetheless, the odd/even pedestal difference was characterized for each crystal from data taken between muon fills. The odd- and even-numbered samples—based on the trace's uncorrected first sample index—are then shifted in value by half of the crystal's measured difference, as shown in Fig. 4.14. This removes most, if not all, of the odd/even pedestal difference in each crystal's trace.

The odd/even gain difference, on the other hand, has been previously studied by inputting sine waves and SiPM pulses. The largest odd/even gain ratio found differs from unity by only 10^{-3}, which similar Monte Carlo studies indicate are entirely negligible. The traces are not adjusted, therefore, for such an odd/even gain difference in this trace-preparation step.

4.5.3   Fit Residual Calculation

The algorithm now enters a loop to identify and then fit each positron in the island one-by-one in order of decreasing energy. It starts out with the hypothesis of no positrons, where the model
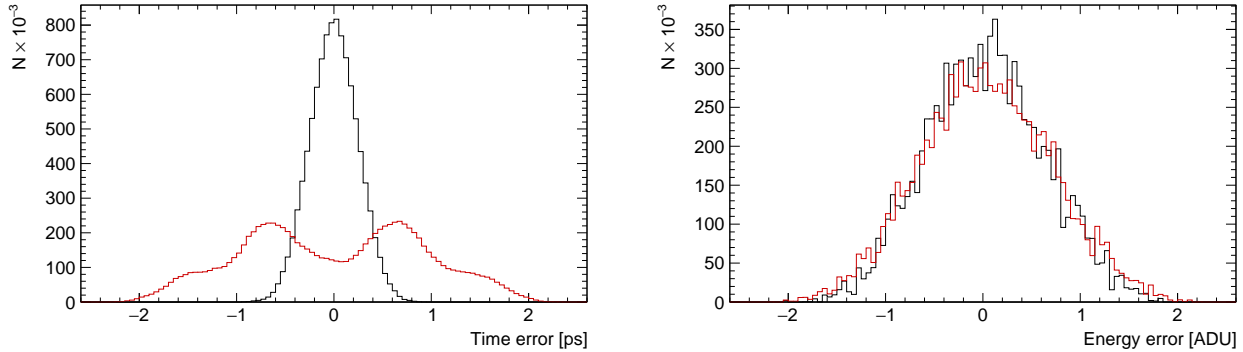
Figure 4.15: The distribution of the time (left) and energy (right) error from template fits to a single trace with a simulated 2-GeV pulse and an even/odd pedestal difference of 0 (black) and 10 ADU (red). The differences in their means are $-0.53(30)$ ps and $0.17(26)$ ADU.

for each crystal's trace is a constant pedestal P—initially estimated as the mean sample value. As positrons are added in the subsequent steps, the model for some of these crystals will be updated to include the fitted pulses described by their template shapes (Sec. 4.3).

In each iteration, the residuals $r_c(t_s)$ from the best-fit model, $y_c(t)$, are first computed for each sample s in every crystal c's trace as given by

$$r_c(t_s) = y_{c,s} - y_c(t_s) \tag{4.14}$$

where $y_{c,s}$ is the sample's value and t_s is the sample's time. As shown in Fig. 4.16, the shape of residuals from a single positron hit is consistent with the broadening of the pulse shape expected in Fig. 4.2. The largest positive residuals typically occur in the samples adjacent to the fitted peak time, being as much as 20 ADU for a 3-GeV positron hit, and approximately scales with the positron's energy. To prevent misidentifying these residuals as energy from another positron, the residuals within δt about every fitted pulse's time are *masked* by setting them as zero.

Further, the residuals' size and spread about the fitted pulse's peak time increases when a second pulse is present from another positron hit. For reasons discussed in the following steps, the fit algorithm imposes an *artificial dead time* (ADT)—defined as the time separation below which only one positron is allowed. To also prevent adding a second positron below the ADT, the mask
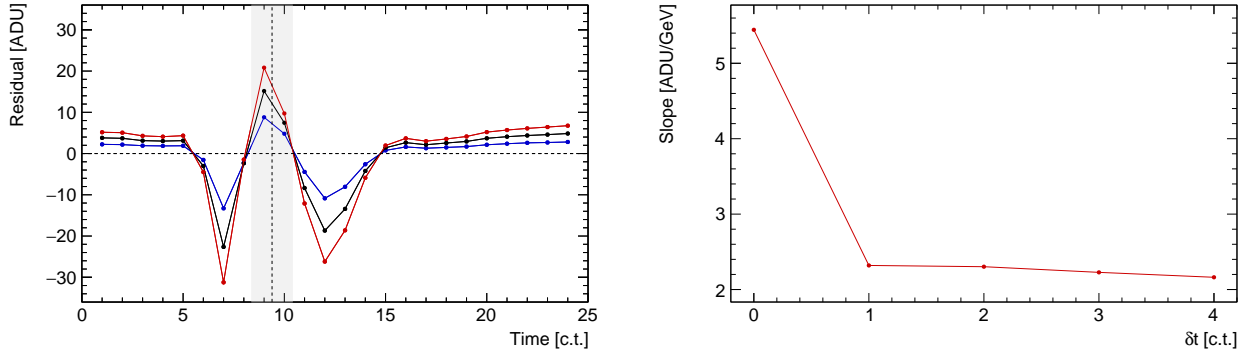
Figure 4.16: Left: The average residuals for 1- (blue), 2- (black), and 3-GeV (red) positron hits. Those falling in the shaded region about the average best-fit time, denoted by the dashed line, will be masked. Right: The slope of a linear fit to the post-fit metric sum's energy dependence, as in Fig. 4.17, for varying values of the residual-masking range δt. Both plots were found in `Geant4`-simulation using the representative distributions in Fig. 4.29.

window, δt, is set equal to the ADT, which is 1 c.t. by default.

4.5.4 Activity Identification

Each crystal's residual trace should now only contain signals from any unaccounted positrons in the island. The next step is to analyze these residuals to decide whether another positron should be added to the model hypothesis. In other words, the algorithm must estimate if there is enough activity in a spatial region of the calorimeter to warrant a fit attempt. To do that, the general approach is to assign a metric for how much energy is in each crystal and then locate the set of crystals that maximizes the sum of their metrics.

An optimal *crystal metric* should have a strong dependence on the amount of energy deposited in the crystal before the fit, thereby being sensitive to a positron hit. At the same time, for only a single positron hit, it should have a weak dependence on the energy after the fit, preventing the algorithm from triggering a second time on the same positron's residuals. The possibly simplest metric satisfying these criteria is the maximum residual in the trace—the pulse height strongly scales with energy and, as seen in Fig. 4.16, the largest residual is less than 0.7 % of the pulse's

original height for a 3-GeV positron. Each crystal, however, has a different noise level, which causes such a metric to be position-dependent. To avoid that, the maximum sample value expected from a noise-only trace, y_n, is subtracted to give a metric for crystal c of

$$m_c \equiv \max\{r_c\} - y_n \qquad (4.15)$$

where y_n is found using normal order statistics as in Eq. 4.13, accounting for the residuals that have been masked. This crystal metric can be thought of as the amount of signal above the noise floor, and the crystal's characteristic noise level used needs only to be approximate to remove the majority of any potential position dependence across the calorimeter.

More elaborate metric definitions were also tested, but, compared to Eq. 4.15, each one suffers in one aspect or another. For example, the absolute value of the number of positive residuals minus the number of negative residuals was tried. This metric could be reasonable since the pulses are unipolar, producing an imbalance in the number of positive and negative residuals. It would also be largely immune to noise level differences because any noise will, on average, produce the same number of samples above and below the pedestal. This leads to a very small energy dependence after the fit—but it also has a small energy dependence before the fit. Alternative metrics investigated include the maximum minus minimum residual, the sum of the largest two or three residuals, and the sum of three or five residuals weighted by the template shape.

With a metric for every crystal, the next step is to find the $N \times M$ crystal region that maximizes the sum of their metrics. From Geant4-simulation, about 95 % of the positron's deposited energy is in a 3×3 window about the central crystal. The pulse heights from the energy deposited in the remaining crystals are on the same order as the typical noise level. A 3×3 *crystal window* is, therefore, taken by default as the spatial region identified with a positron hit—although it can be extended in a later cluster-merge step. When describing the merging step, it will become clear that a larger nominal window would only increase the amount of pileup.
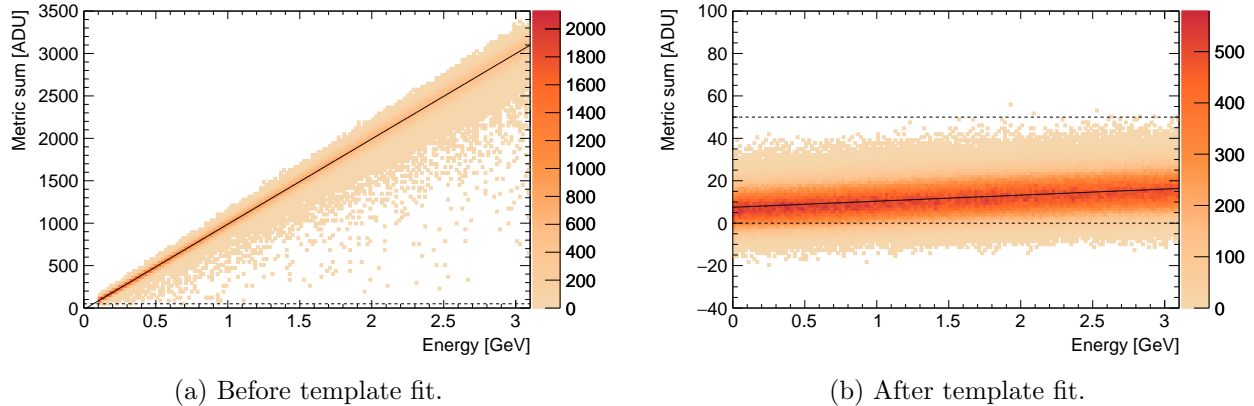
(a) Before template fit.

(b) After template fit.

Figure 4.17: The energy dependence of the metric sum (a) before and (b) after the fit, with the default residual masking of $\delta t = 1$ c.t, which is used to decide if there is enough activity in the 3×3 crystal region to suggest another positron hit. The default metric-sum threshold of 50 ADU is denoted by a horizontal line. These data are from simulated single positron hits with nominal incidence and with positions uniformly spread across the crystal's front face.

The metric sum is then computed for each possible window location in the calorimeter, and the crystal window with largest value is selected. This maximum is guaranteed as there is a finite number of window locations. It is in this procedure that the fit algorithm operates in order of decreasing energy. Next, the window selected is shifted to be about the central crystal—the crystal with the largest metric. If the central crystal is on the boundary of the calorimeter, then the window size will be truncated. The metric sum in the shifted window is then computed, and, if it is more than 50 ADU, by default, the crystal window becomes a *cluster* that is passed to the subsequent step to be simultaneously fit. Each cluster is typically equated to a positron hit but could also be from a lost muon or, as discussed in Sec. 4.9, from noise. If the window's metric sum is lower than the threshold, then all clusters in the island are considered to have been identified, and the fit algorithm exits the loop to the position-estimation step.

In general, this metric-sum threshold sets the minimum positron energy required to be found and should be as low as reasonably possible. If set too low, though, the rate of false triggers on the same positron's residuals becomes high. As seen in Fig. 4.17, the 50-ADU threshold strikes a

balance between identifying positrons above about 66 MeV while rarely triggering falsely at all observable energies. The latter is only achieved by masking the residuals that are within $\delta t = 1$ c.t. of the fitted peak time in the previous step (Fig. 4.16). The metric sum's energy dependence after the fit is only slightly further reduced for larger values of δt. Moreover, for crystal windows truncated at the calorimeter's boundary, this procedure will result in a relatively slightly higher energy threshold because it includes only part of what would have been the full electromagnetic shower. The exact threshold increase, however, will depend on the incident angle and position of the positron striking the calorimeter.

4.5.5   Simultaneous Template Fit

All of the clusters that have been identified so far are then simultaneously fit to determine their times and energies. For every cluster, each of its crystals' traces is modeled by its template pulse shape, $S_c(t)$, previously found for that crystal c (Sec. 4.3). This template's amplitude is allowed to float, and the sum of the amplitudes is taken as the cluster's (uncalibrated) energy. It can also be translated in time but must share the same peak time with the templates in the cluster's other crystals. Finally, each crystal's pedestal is modeled by a constant.

The fit is performed by minimizing the χ^2-function—formed over the residual samples s and clusters r for each of their crystals c—given by

$$\chi^2 = \sum_c \sum_s \sigma_{c,s}^{-2} \left[y_{c,s} - \sum_r A_{c,r} \cdot S_c(t_s - t_r) - P_c \right]^2 \tag{4.16}$$

where $\sigma_{c,s}$ is the sample value $y_{c,s}$'s uncertainty; $A_{c,r}$ and t_r are the template's amplitude and the cluster's peak time, respectively; and P_c is the pedestal. The fit parameters are $A_{c,r}$, t_r, and P_c, and $\sigma_{c,s}$ is taken as the crystal's previously characterized noise level.

In general, $\sigma_{c,s}$ allows for each measurement, $y_{c,s}$, to be weighted by its uncertainty, which is assumed to be normally distributed. A more realistic weight, therefore, would be to incorporate

the clock's period jitter—the uncertainty of the sampling time t_s. Both the y- and t-uncertainty can be properly accounted for with orthogonal-distance regression. Even better is to use a covariance matrix among the samples, incorporating the photostatistics and pulse shape. These more realistic weights, however, are not necessarily the best ones to use for an ω_a-measurement. Most notably, Monte Carlo studies have concluded that, out of those considered, flat weights—such as the crystal's noise level—cause the smallest systematic uncertainty from so-called unseen pileup, which will be introduced and assessed in Sec. 6.2.

The χ^2-minimization method used is motivated by a key observation that collapses the problem into an optimization over only the time parameters [91]. For each crystal, if the peak times $\boldsymbol{t} = \{t_r\}$ are held fixed, $\nabla\chi^2$ is linear in all parameters. Therefore, given the times $\boldsymbol{t}$, the remaining parameters, $\{A_{c,r}\}$ and $\{P_c\}$, can be solved for exactly in a system of $N + 1$ linear equations, where N is the number of clusters for that crystal. These equations can be written as a matrix equation of the form $TT^T X = TY$, with M residual samples, where

$$T = \begin{bmatrix} 1/\sigma_{c,1} & 1/\sigma_{c,2} & \cdots & 1/\sigma_{c,M} \\ S_c(t_0 - t_1)/\sigma_{c,1} & S_c(t_1 - t_1)/\sigma_{c,2} & \cdots & S_c(t_M - t_1)/\sigma_{c,M} \\ \vdots & \vdots & \ddots & \vdots \\ S_c(t_0 - t_N)/\sigma_{c,1} & S_c(t_1 - t_N)/\sigma_{c,2} & \cdots & S_c(t_M - t_N)/\sigma_{c,M} \end{bmatrix}, \tag{4.17a}$$

$$X = \begin{bmatrix} P_c \\ A_{c,1} \\ \vdots \\ A_{c,N} \end{bmatrix}, \text{ and } Y = \begin{bmatrix} y_{c,1}/\sigma_{c,1} \\ y_{c,2}/\sigma_{c,2} \\ \vdots \\ y_{c,M}/\sigma_{c,M} \end{bmatrix} \tag{4.17b}$$

which is guaranteed a unique solution if T is nonsingular. The χ^2-minimization can, therefore, be performed numerically over only the time-parameter subspace, where, at every search iteration, the pulse amplitudes and crystal pedestals are analytically calculated using Eq. 4.17. In practice,

the `Eigen3` linear algebra library is employed to solve Eq. 4.17 via a LDLT decomposition, and its results have been found to agree with those from other software libraries [97].

The cluster-time minimization, on the other hand, is carried out with the `MnMigrad` routine in the `Minuit2` software package [100]. This routine has been thoroughly tested over time by a broad user base and, as such, has been widely accepted in the physics community. It employs a variation of the Davidon–Fletcher–Powell (DFP) gradient method—a quasi-Newton method, motivated by the classic Newton's method for root finding. In Newton's method applied to $\nabla\chi^2(\boldsymbol{t}) = 0$, the next iteration point is given by $\boldsymbol{t}_{i+1} - \boldsymbol{t}_i = -\boldsymbol{H}^{-1} \cdot \nabla\chi^2(\boldsymbol{t}_i)$, where $\boldsymbol{H} \equiv \nabla^2\chi^2(\boldsymbol{t})$ is the Hessian matrix. In contrast, the quasi-Newton method uses only first-order information and approximates the second-order at each iteration so that $\lim_{i\to\infty} \boldsymbol{H}_i = \boldsymbol{H}$. If far from the root, it turns out that such approximations can actually be better than using the true Hessian [101]. The quasi-Newton's next iteration analog then becomes

$$\boldsymbol{t}_{i+1} - \boldsymbol{t}_i = \boldsymbol{H}_{i+1}^{-1} \cdot [\nabla\chi^2(\boldsymbol{t}_{i+1}) - \nabla\chi^2(\boldsymbol{t}_i)] \tag{4.18}$$

where $\boldsymbol{H}_{i+1}^{-1}$ is an updating formula that must be based on the function values and gradients from previous iterations, e.g., $\boldsymbol{X}_i \equiv \boldsymbol{t}_{i+1} - \boldsymbol{t}_i$ and $\boldsymbol{F}_i \equiv \nabla\chi^2(\boldsymbol{t}_{i+1}) - \nabla\chi^2(\boldsymbol{t}_i)$. The DFP gradient method is then defined as using the updating formula given by

$$\boldsymbol{H}_{i+1}^{-1} = \boldsymbol{H}_i^{-1} + \frac{\boldsymbol{X}_i \cdot \boldsymbol{X}_i^T}{\boldsymbol{X}_i \cdot \boldsymbol{F}_i} - \frac{[\boldsymbol{H}_i^{-1} \cdot \boldsymbol{F}_i][\boldsymbol{H}_i^{-1} \cdot \boldsymbol{F}_i]^T}{\boldsymbol{F}_i \cdot \boldsymbol{H}_i^{-1} \cdot \boldsymbol{F}_i}. \tag{4.19}$$

The `MnMigrad` routine continues to iterate according to Eq. 4.18 until one of two convergence criteria is satisfied. The first such stopping condition is when the difference between the χ^2 at the current iteration point $\boldsymbol{t}_i$ and the minimum of a quadratic with $\boldsymbol{H}_i$, defined as

$$\rho = \frac{\nabla\chi^2(\boldsymbol{t_i})^T \cdot \boldsymbol{H}_i^{-1} \cdot \nabla\chi^2(\boldsymbol{t_i})}{2}, \tag{4.20}$$

is below 0.001 times the (default) tolerance value of 0.1. This scale-free quantity is the expected vertical distance to the minimum, assuming χ^2 is locally quadratic, defining how close to the min-

imum is close enough. It should be noted that the routine's working precision is set, by default, to be machine precision. The second such stopping condition is when the (approximate) number of function calls exceeds a maximum threshold, which by default is

$$N_f = 200 + 100 \cdot N_t + 5 \cdot N_t^2 \tag{4.21}$$

where N_t is the number of clusters in the island. This criterion allows the minimization to stop if the ρ-threshold is never reached. As the χ^2-space needed to be explored grows with the number of time parameters, Eq. 4.21 was defined to scale quadratically with N_t. When either of these convergence criteria has been reached, the best-fit cluster times that are returned are saved while their uncertainties are ignored as they are not used in the ω_a-measurement.

As well, the gradient $\nabla \chi^2$ is numerically estimated at each iteration point $\boldsymbol{t}_i$ by the `MnMigrad` routine by sampling the χ^2 value at, by default, four points about each time parameter using a step-size tolerance of 0.5. This corresponds to the lowest "strategy level" of zero, which has been found to be sufficient. As compared to the highest strategy level of two, it is also about 1.5 times faster. Further, these gradient values agree with those numerically estimated by hand.

Moreover, Eq. 4.19 tacitly assumes that $\boldsymbol{H}_i$ is positive definite. A non-positive definite matrix means that, at least, one eigenvector points uphill, indicating either a local maximum or a saddle point. There are a number of ways detailed in literature to handle encountering such an iteration point, including adding a correction matrix to $\boldsymbol{H}_i$ to force positive definiteness [102]. In this case, the `MnMigrad` routine—as found in its `MnPosDef` class—implements such a correction matrix by applying the so-called Levenberg–Marquard rule:

$$\boldsymbol{H}_i \rightarrow \boldsymbol{H}_i + \lambda \cdot \boldsymbol{I} \tag{4.22}$$

where $\boldsymbol{I}$ is the identify matrix and λ is a scalar given by

$$\lambda = 0.5 + \max\{10^{-6}, 2\sqrt{\epsilon}\} - \min\{(\boldsymbol{H}_i)_{jj}\} \tag{4.23}$$

with ϵ as the machine epsilon and the last term as the smallest diagonal element of $\boldsymbol{H}_i$. This procedure can loosely be thought of as "kicking" the time parameters away to where $\boldsymbol{H}_i$ is known to be positive definite. A potential problem, however, is that the new iteration point $\boldsymbol{t}_i$ may be far away from samples' times, making the templates' influence in Eq. 4.16 very small. This can lead to the numerically calculated χ^2-gradient being nearly zero, prompting the `MnMigrad` routine to stop and flag the fit result as *invalid*.

In general, the `MnMigrad` routine flags a given fit result as invalid if it stopped with ρ above the threshold, it encountered a problem with the covariance matrix, or it exceeded the maximum number of function calls. These status flags are saved along with the fit result to allow for related ω_a-systematic studies. Thorough understanding of the `MnMigrad` routine, however, heavily influenced the remaining aspects of the fit procedure, maximizing the number of successful fit results. In the 1a data set (Sec. 5.1), for example, only 0.0005 % of the clusters reconstructed were flagged as invalid. A naïve fit procedure, in comparison, is known to result in up to 0.17 % of clusters fits being invalid. The invalid fit results remaining have been extensively examined in simulation and in data and are found to have reasonable best-fit values. They almost always occur due to a ρ over-threshold when there are multiple clusters in the island; from simulation, a two-cluster fit, on average, stops with a ρ about 1.41 times larger than that for a one-cluster fit, even though they both have comparable reduced χ^2 values. As shown in Fig. 4.18, this is confirmed in the 1a data set where all of the invalid fit results were due to a ρ over-threshold and follow the two-positron pileup rate. Out of over 10^9 total fits performed, remarkably no χ^2-minimization reached its function call limit, and only five needed the $\boldsymbol{H}_i$ matrix to be forced positive definite via Eq. 4.22. The `MnMigrad` routine also never returned a "covariance status" of zero—the only invalid-fit flag found in simulation to yield a bad best-fit result. As a result, this fit-algorithm step accepts all fits as successful, and no cut on clusters flagged as invalid is used for the ω_a-measurement.
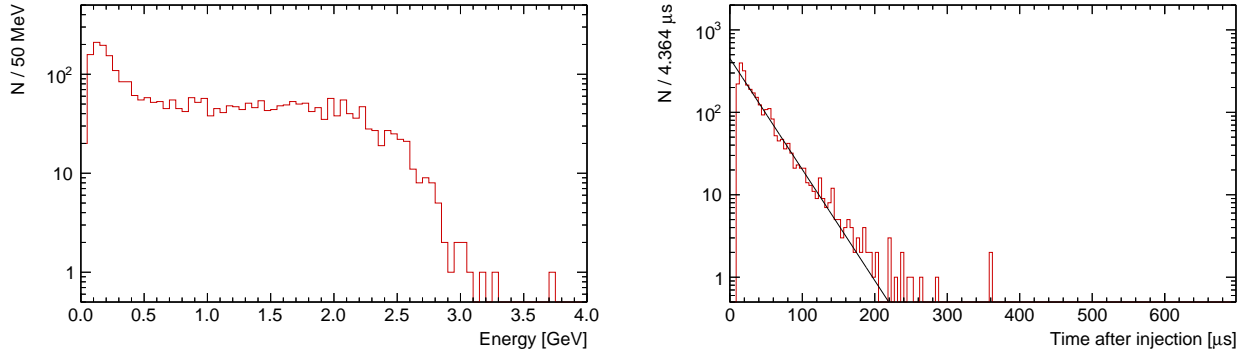
Figure 4.18: The energy (left) and time (right) distributions of clusters from fits flagged by the `MnMigrad` routine as invalid in the 1a data set (Sec. 5.1). In total, these clusters constitute 0.0005 % of the approximately 10^9 clusters reconstructed. It should be noted that the energy distribution closely resembles that of the full data set and that the time distribution closely follows the two-positron pileup rate—as shown by a fit to $Ne^{-t/32.22\,\mu s}$ (black).

It should be noted that other χ^2-minimization routines from various software packages were also tested, including Newton's method, a Newton's method augmented with Armijo's rule, and a Nelder–Mead implementation of the downhill-simplex method [100, 103]. Each routine strikes a different balance between computation time and robustness. Newton's method, for instance, is fast but is not guaranteed to converge if the χ^2 function is not convex. The downhill-simplex method, on the other hand, is exceptionally slow but also quite robust. All in all, the `MnMigrad` routine is found in simulation to perform as well or better than all of the other routines tested, while being sufficiently fast—as discussed further in the last step (Sec. 4.5.9).

With the χ^2-minimization method laid out, the procedures used to help ensure the robustness of its best-fit result can now be discussed. First off, the `MnMigrad` routine requires an estimate of t_0, known as the *seed*, to make its initial ($i = 0$) iteration in Eq. 4.18. As such, this estimate of each cluster's time needs to be sufficiently close to the true time but need not be perfect. For the newly added cluster, its time seed is initially taken as the time of its central crystal's maximum residual found when calculating its metric in Eq. 4.15. It is then refined within a clock tick by using its adjacent samples, if available, to calculate its pseudotime according to Eq. 4.1 and mapped

onto real time by using Eq. 4.2. In contrast, the time seed for each previously fit cluster is taken as its current best-fit value. The only exception is when its best-fit value is more than the ADT—1 c.t. by default—away from its previous seed. This is known to occur for some two-positron hits that result in overlapping pulses with resolved peaks of commensurate amplitude. Such, in effect, swapping of clusters poses an issue with the sample fitting range that will be discussed next. In this case, the previous seed is used again for the current χ^2-minimization. It should also be mentioned that the previous fit's parameter errors are not carried over to the current fit. This is to prevent any bad fit—for example, because not all of the clusters have been accounted for yet—from propagating forward. Instead, their tolerances are reset to the default value of 0.1.

In addition, for a given crystal, not every sample is always included in the χ^2's sum over s in Eq. 4.16. Relative to the time of the new cluster's maximum residual in the central crystal, only the previous 8 and the next 16 samples, by default, are also used—akin to the island chopper's pre-/post-sample concept (Sec. 4.2). This avoids any possible systematic ω_a-bias from the island length, which, on average, shrinks at the pileup rate. If not all pre- or post-samples exist because, say, the peak is towards the end of the island, as many samples as available are used. The island chopper, however, always saves 8 pre- and 16 post-samples about every sample that exceeds its threshold, guaranteeing the fit algorithm enough samples for all clusters with a pulse over-threshold. In simulation, as long as there are more than 2 pre- and 3 post-samples included in the χ^2, the best-fit time and energy remain unchanged. Also, this range of samples used in the χ^2 does not update in subsequent fits, partly motivating the seed-reset procedure above.

Furthermore, the amplitude parameter for each crystal in a cluster is allowed to be negative, i.e., it is unbounded in the fit. This can typically occur if the true signal is below the noise floor because a pulse's amplitude from a noise-only trace's fit is equally likely to be either positive or negative. A negative amplitude, though, is unphysical if related by a calibration constant to the

energy deposited in a crystal. As it is the *sum* of these amplitudes that is taken as the cluster's (uncalibrated) energy—the value of interest for the ω_a-measurement—it was found best to prioritize this sum as the physical quantity. Put another way, the pulse amplitude in a given crystal is allowed to become slightly negative in order to give a better fit of the cluster's physical time and energy, which will be calibrated in Sec. 4.6.

Each time parameter, on the other hand, is bounded in the fit to be within its cluster's aforenamed sample range, i.e., $t_r \in [T_1, T_2]$ where $T_1 = t_{\max} - 8$ c.t. and $T_2 = t_{\max} + 16$ c.t. with $t_{\max}$ being the time of central crystal's peak residual. This imposes that each pulse must peak inside the sample range used in the χ^2, which is necessary for the amplitude(s) to be stable and reasonable. For example, a noise fluctuation or a pulse below the island chopper's threshold can occur in the last few samples of the island. If unbounded, an associated cluster's peak time may then move outside the island and grossly inflate its amplitude(s) so as to fit the data to only its rising edge. In such cases, there are simply not enough data to sensibly constrain the amplitude. Assuming a reasonable threshold, however, the island chopper would have extended the island if there were any significant activity. Therefore, the only consequence to bounding the time parameters is the possibility of missing *very* low-energy positrons (Sec. 6.2).

Moreover, as the quasi-Newton method applies to unconstrained optimization, the `MnMigrad` routine—inside its `SinParameterTransformation` class—maps the bounded time t_r to an internally used unbounded time, t_r', according to

$$t_r = T_1 + \frac{T_2 - T_1}{2} \left(\sin t_r' + 1 \right). \tag{4.24}$$

As t_r' by construction barely changes near $\pm\pi/2$, the expected value of t_r should ideally not be close to its bounds for the χ^2-minimization to perform best. Fortunately, this condition is largely satisfied by the above procedure and by the island chopper always saving pre- and post-samples, thereby guaranteeing a buffer from the island's bounds.

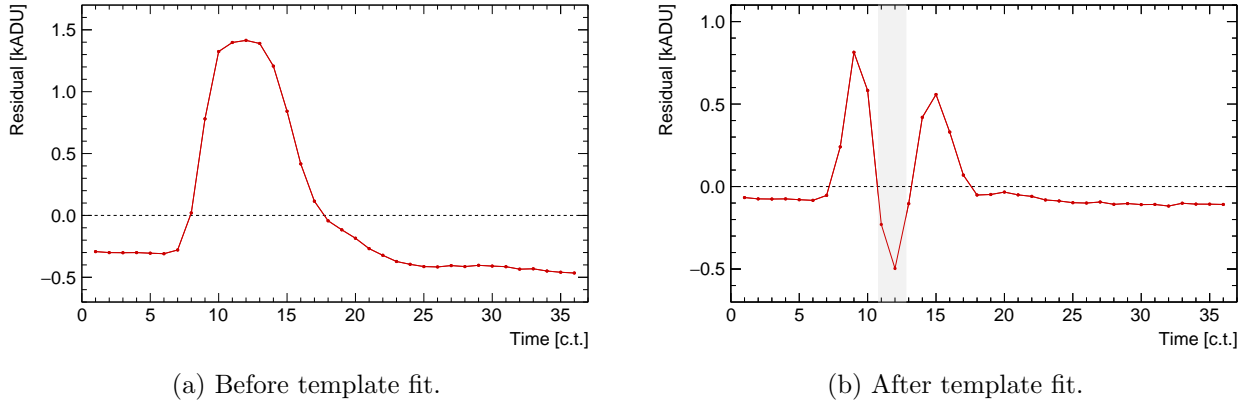(a) Before template fit.

(b) After template fit.

Figure 4.19: The fit residuals (a) before and (b) after the template fit for a cluster's central crystal, formed by 3.05- and 2.04-GeV positrons striking the crystal with a time separation of 4.14 ns in `Geant4`-simulation. This time separation is roughly the maximum such that there is initially only one peak resolved. The fit residuals before the template fit are calculated using a model only of the crystals' pedestals. After the fit, the residuals in the shaded region will be masked.

Finally, if multiple clusters' pulses overlap in time and/or space, their seeds may not be sufficiently close to their true values for the χ^2-minimization. One of the most egregious situations involving two positrons is depicted in Fig. 4.19. With initially only one broad peak, the two positrons' time seeds are difficult to estimate accurately, resulting in a poor fit—even if flagged as valid. The remedy used is to tweak the time-parameter seeds and repeat the χ^2-minimization, looking for a better result. In practice, whenever there are two or more clusters with overlapping crystal windows, the fit is always repeated, by default, four more times with the new cluster's seed tweaked by ± 1 and ± 2 c.t. This is also performed if the `MnMigrad` routine flags the fit result as invalid. The best result is then selected based on its χ^2 value. Specifically, a tweaked-seed result is deemed better if it is flagged as valid with a χ^2 lower than the original seed's χ^2 by more than (by default) 0.1. If all of the fit results are flagged as invalid, then the one with the lowest χ^2 is selected. In general, this procedure allows for the `MnMigrad` routine to escape from local minima and/or saddle points. It is also very important for avoiding a χ^2-minimization that encounters a non-positive definite $\boldsymbol{H}_i$, which can cause it to stop prematurely as discussed.

4.5.6   Cluster Merge

In the next two steps, the fit result with this new cluster is evaluated. To begin, the time/space relationship of the clusters is assessed to ensure the fit stability in any overlapping crystal. If two clusters' times are very close to each other and share a crystal, for example, the fit has two free amplitude parameters that are effectively degenerate. In these cases, the best-fit amplitudes will sometimes be one very negative and one very positive, where the sum is the physical amplitude observed in the waveform digitizer's trace. Furthermore, there is a degeneracy with how to divide the physical amplitude between the two pulses' amplitudes. For these reasons, this step enforces the fit's aforementioned ADT—defined as the smallest time separation allowed between any two clusters. In general, the ADT should be as low as reasonably possible to reduce the occurrence of pileup and, enabled by the last step's robust χ^2-minimization, is set, by default, to 1 c.t. with sufficient behavior, as shown in Sec. 4.8.

To that end, if any two clusters share at least one (but not every) crystal between their windows and are within an ADT of each other, they are *merged* together. If a cluster satisfies these criteria with two or more different clusters, then the clusters closest to each other in time are the ones merged. When two clusters are so-called merged, the more recently added cluster is removed from the fit model, and the unshared crystals are added to the other cluster's window to form a spatially larger cluster. This procedure is repeated until no additional pair of clusters needs to be merged. Afterward, if any clusters were merged, the fit algorithm returns to the previous step where the updated set of clusters is then simultaneously fit.

In this way, a cluster's window can be extended beyond its initial 3×3 crystal size. This is beneficial, for example, when a positron occasionally hits a calorimeter at a large angle, depositing its energy mostly across *four* crystals horizontally. For such a case in simulation, the pulse outside

the initial 3×3 window will peak together with the other pulses up to $\mathcal{O}(100 \text{ ps})$—much less than the ADT. If this energy leakage is large enough, a second cluster will be found and then merged into the first, resulting in a 3×4 crystal window and a better energy resolution.

4.5.7 Cluster Rejection

The (possibly merged) clusters are next evaluated to decide whether to accept the fit result with the new cluster added in the activity-identification step. In particular, there are three cases that cause rejection. If any of these rejection criteria are satisfied, the best-fit result is tossed out, and the previous best-fit result is restored.

First, the fit result is rejected if any two clusters have the identical crystal window and are within an ADT of each other. In this case, the more recently added cluster is interpreted as one that was mistakenly triggered by the first cluster's fit residuals or by noise. Second, the result is rejected when any of the clusters' energies is too small to discern from that of a fit to simply noise-only traces. That is, it is rejected if, for any cluster r,

$$\sum_c A_{c,r} < \sum_c y_{c,n} \tag{4.25}$$

using the notation in Eq. 4.16, where $y_{c,n}$ is the peak sample value expected from a noise-only trace—found via normal order statistics using the cluster's number of samples in the χ^2. This removes the vast majority of clusters formed from oscillatory noise in the traces, which will be discussed in Sec. 4.9. Third, the fit result is rejected if the new cluster's time is within (by default) 0.5 c.t. of its bounds. This protects against the issues raised in regard to Eq. 4.24. It also improves the best-fit amplitude's stability by ensuring that the pulse's peak can be resolved in only the sample range used in the χ^2—even by the pseudotime-based method.

If the fit result is rejected, the central crystal of the now-rejected new cluster is *masked* by fixing its crystal metric to be zero hereafter for the island. This prevents the same cluster from being

found during the algorithm's next iteration. However, as the clusters are added in approximately decreasing energy, this step will also prevent any lower-energy clusters from being found near the masked region of the calorimeter. Lower-energy clusters in other spatial regions, though, can still be identified. After this step, the fit algorithm then proceeds back to the fit residual-calculation step to prepare to look for another cluster, beginning its next iteration.

4.5.8 Position Estimation

After all of the clusters are identified and fitted, a better estimate of each cluster's position is ascertained. While not strictly necessary for an ω_a-measurement, a well-known position, $\boldsymbol{x}$, for each positron enables a wide-range of ω_a-consistency and -systematic studies. In general, there have been a number of position-reconstruction methods developed, ranging in complexity, for laterally segmented calorimeters. One of the simpler methods is to compute the classic "center of gravity" of the electromagnetic shower, which is generically given by

$$\boldsymbol{x} = \frac{\sum_c w_c \cdot \boldsymbol{x}_c}{\sum_c w_c} \tag{4.26}$$

where $\boldsymbol{x}_c$ is the center position of crystal c and w_c is a weight assigned to it. Out of the best-fit quantities available, the most obvious weight to use is the (uncalibrated) energy deposited in the crystal, i.e., $w_c = A_c$. However, as seen in Fig. 4.20, the position reconstructed using such linear weighting is systematically biased towards the crystal center. This nonlinear behavior, though, is well-known in literature and is characterized by

$$\boldsymbol{x} = \boldsymbol{x}_c + b \cdot \mathrm{arcsinh}\left[\frac{|\hat{\boldsymbol{x}} - \boldsymbol{x}_c|}{\Delta} \cdot \sinh\frac{\Delta}{b}\right] \tag{4.27}$$

where b is a free parameter related to the shower's profile, Δ is half a crystal's width (c.w.), and $\hat{\boldsymbol{x}}$ is the true position [104]. While an excellent model to Geant4-simulated data, the best-fit b depends on the precise geometry and the positron's energy and incident angle. Instead of making

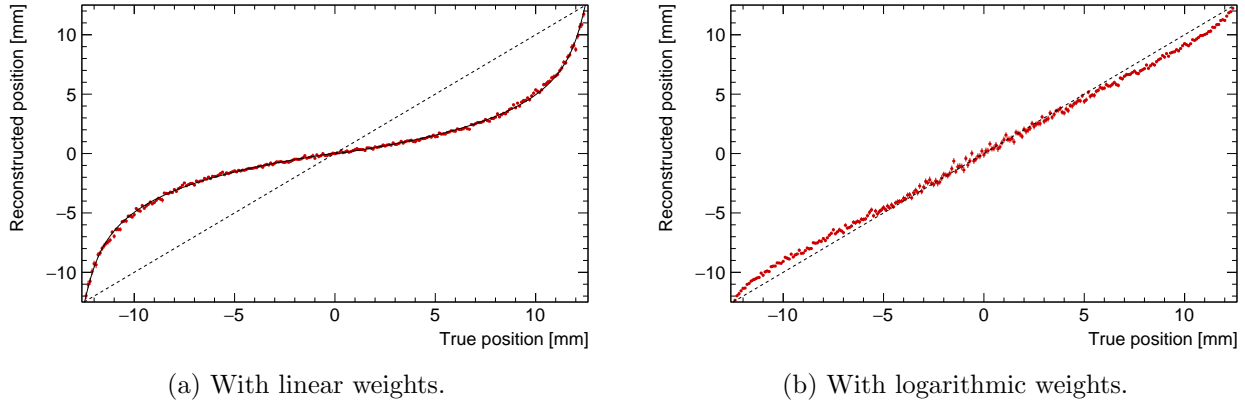(a) With linear weights.

(b) With logarithmic weights.

Figure 4.20: The reconstructed position as a function of the true position using the center-of-gravity method in Eq. 4.26 with (a) linear weights, $w_c = A_c$, and (b) logarithmic weights given by Eq. 4.29, where $W_0 = 3.2$ based on Fig. 4.22. These data are from Geant4-simulation of a normally incident 2-GeV positron. Any deviation from the one-to-one dashed line indicates the position reconstruction's bias. A fit to Eq. 4.27 in (a) has been overlaid (black).

a more complex model, much smaller biases are achievable without a correction by using a more informed weighting scheme based on the shower's lateral distribution.

As shown in Fig. 4.21, the energy deposited laterally across the calorimeter from a positron hit falls off exponentially. The best model, in fact, is known to be the sum of two exponentials, integrated over the cross-section of crystal c, as given by

$$A_c = \int_{x_c-0.5 \text{ c.w.}}^{x_c+0.5 \text{ c.w.}} \int_{y_c-0.5 \text{ c.w.}}^{y_c+0.5 \text{ c.w.}} \left[E_1 \cdot e^{-\sqrt{x^2+y^2}/\sigma_1} + E_2 \cdot e^{-\sqrt{x^2+y^2}/\sigma_2} \right] \mathrm{d}x \, \mathrm{d}y \qquad (4.28)$$

where, from Geant4-simulation, $E_2/E_1 \approx 26$, $\sigma_1 \approx 0.79$ c.w., and $\sigma_2 \approx 0.16$ c.w. [104]. This exponential behavior can be taken into account in w_c by instead weighting by the *logarithm* of the fraction of energy deposited, according to

$$w_c = \max \left\{ 0, W_0 + \log \frac{A_c}{\sum_{c'} A_{c'}} \right\} \qquad (4.29)$$

where W_0 is a free parameter [105]. In essence, W_0 sets the relative importance of the electromagnetic shower's tail, defining the energy threshold required for a crystal to be included in the calculation. To exploit any of the crystals' small energies, A_c is in fact summed over for each col-
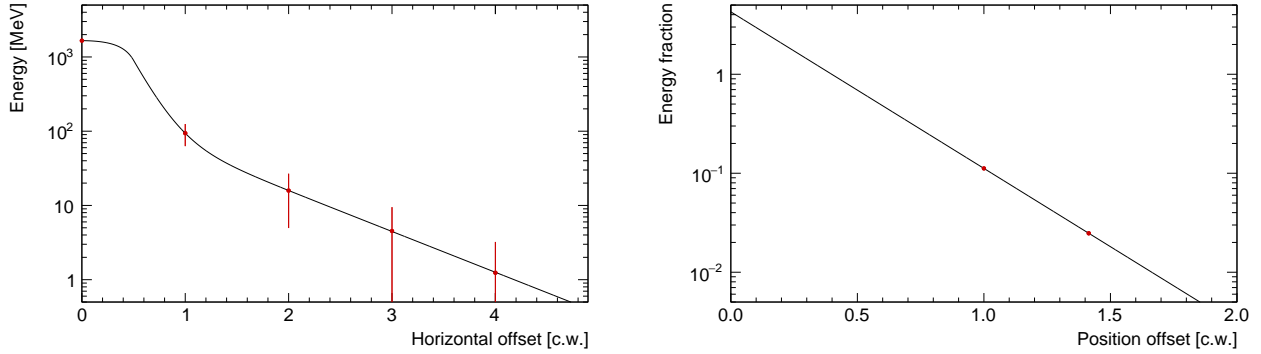
Figure 4.21: Left: The lateral distribution of energy from a Geant4-simulated, 2-GeV positron striking the crystal's center. A fit to the two-exponential model in Eq. 4.28 is given in black with $E_2/E_1 \approx 26$, $\sigma_1 \approx 0.79$ c.w., and $\sigma_2 \approx 0.16$ c.w. Right: The side and corner crystals' energies reconstructed from data in a 3×3 window, normalized by the central crystal's energy. The position offset is defined by the distance between the crystal centers. A one-exponential fit with a characteristic distance of 0.2740(77) c.w. is overlaid in black.

umn (row) when calculating the horizontal (vertical) position in Eqs. 4.26 and 4.29. Even for a reasonable choice of W_0, however, occasionally none of the side crystals' amplitude sums will be large enough for a non-zero weight using Eq. 4.29. In such cases, the position reconstructed via Eq. 4.26 will be the central crystal's center.

Based on Fig. 4.22, W_0 was chosen to be 3.5, which minimizes the position resolution measured using normally incident, 3-GeV electrons at SLAC. The optimal W_0, however, is known to be slightly linearly dependent on energy [105]. Accordingly, this value agrees well with the optimal value of 3.2 found from a Geant4-simulation of normally incident, 2-GeV positrons. As the resolution valley is fairly shallow, though, the W_0 parameter need not be finely tuned for a sufficient result. Notwithstanding, as shown in Fig. 4.20, the reconstructed position's systematic bias with true position is greatly reduced when using the logarithmic weights in Eq. 4.29. Thus, no correction needs to be applied afterward. From Geant4-simulation, the reconstructed position has also been found to be unbiased for modest angles up to about 15 ° from normal—roughly the largest angle expected for positrons with energy above 1.3 GeV.
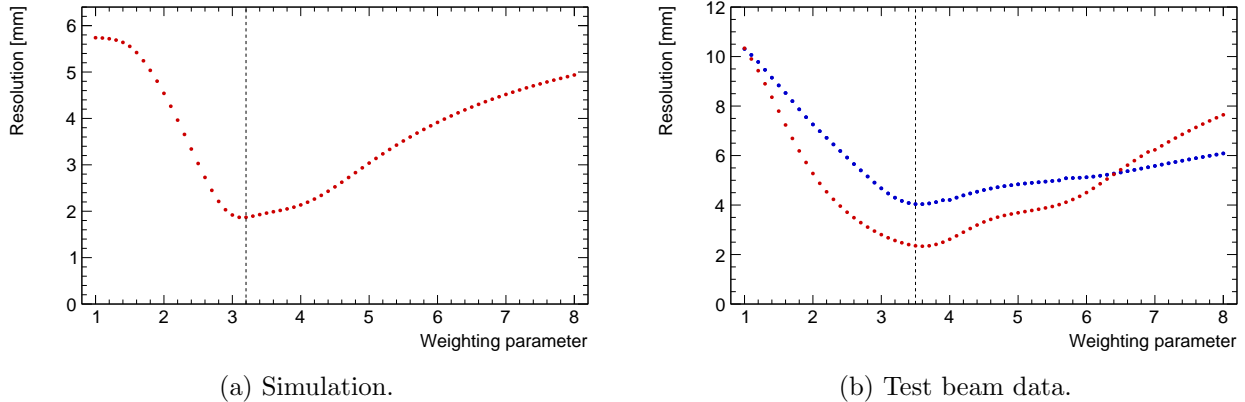
(a) Simulation.

(b) Test beam data.

Figure 4.22: The horizontal (red) and vertical (blue) position resolution, found using the center-of-gravity method with the logarithmic weights in Eq. 4.29, as a function of the weighting parameter W_0. The data are from (a) `Geant4`-simulation of 2-GeV positron hits across the crystal face and from (b) 3-GeV electrons observed at the SLAC T-536 test beam in July 2016 [60]. Note that the resolutions from data are enhanced due to the electron beam's spread.

The center-of-gravity method breaks down, however, when the true position is in a boundary row or column of the calorimeter since there is no opposite crystal to balance the weighting. In such a case—defined by the position of a cluster's central crystal—the position is estimated using a so-called adjacent-crystal method instead, formally given by

$$\boldsymbol{x} = \boldsymbol{x}_c \pm \Delta \mp W_1 \cdot \log\left[\frac{1}{2}\left(\frac{A_c}{A_{c\pm 1}} + 1\right)\right] \tag{4.30}$$

where $\Delta = 0.5$ c.w. and $W_1 = 0.2$ c.w.—a free parameter that was likewise optimized with simulations [104]. Here, A_c and $A_{c\pm 1}$ denote the sum of the cluster's amplitudes in the boundary and adjacent column (row) for calculating the horizontal (vertical) position, respectively.

Each position-reconstruction method was further evaluated for varying incident positions on the crystal's front face, as shown in Fig. 4.23. As expected, the center-of-gravity method with linear weights has the largest absolute error near the crystal's edge but also has the smallest resolution near the crystal's center. Aside from weighting with $w_c = \sqrt{A_c}$, the other methods tested all have similar characteristics, exhibiting a minimal bias across the crystal and a best resolution near the crystal's edge. A summary of their resolutions and absolute errors, from a population
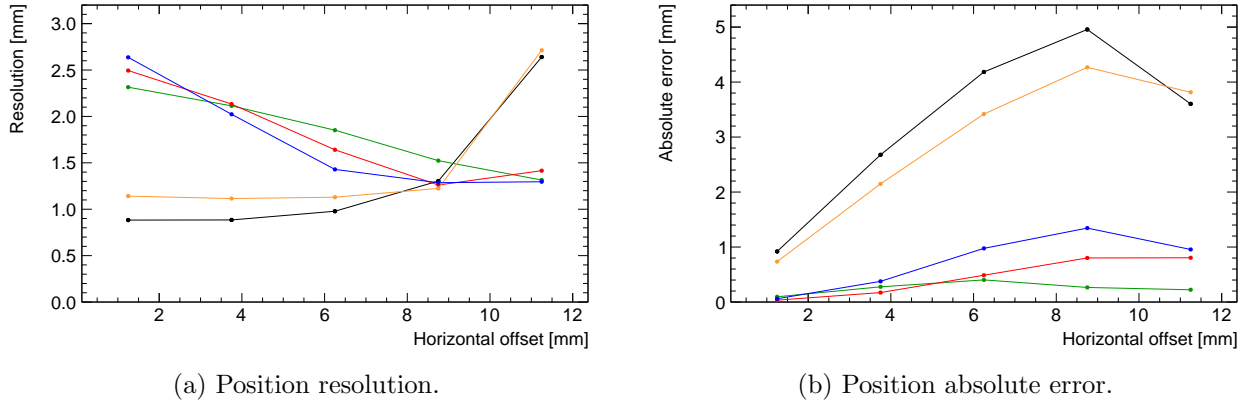
(a) Position resolution.

(b) Position absolute error.

Figure 4.23: The reconstructed position (a) resolution and (b) absolute error for horizontal off-sets from the crystal center across half of a crystal width—found from `Geant4`-simulated, 2-GeV positrons. The methods tested are the center-of-gravity (COG) with linear weights (black), corrected COG with linear weights (green), COG with logarithmic weights, COG with square-root weights of $w_c = \sqrt{A_c}$ (orange), and the adjacent crystal (blue).

of positrons hitting across a crystal's front face, is given in Tab. 4.1. As can be seen, the two position-reconstruction methods employed in this fit algorithm via Eqs. 4.29 and 4.30 both have roughly a 2-mm resolution and a bias of less than 0.03 c.w.

Moreover, each cluster's position is recomputed whenever its energy is calibrated, corrected, and/or partitioned, as done in the reconstruction's following stages. In such cases, the identical equations are used save for an updated definition of A_c after converting it to a calibrated energy. It should also be noted again that, in this step, the electromagnetic shower's position is reconstructed—more specifically, the position at the shower's mean depth of roughly $7.13 \cdot X_0$, where $X_0 = 9.3$ mm is the radiation length in PbF_2. To map this position to the hit position on the crystal's front face, the positron's incident angle $\boldsymbol{\theta}$ must be known. Attempts have been made to estimate $\boldsymbol{\theta}$ from the shower's moments, central and standardized, as well as from data-driven shower models, but they all exhibit poor resolution [106]. As a crystal width is comparable with the Molière radius of PbF_2, few crystals have appreciable energy deposits, and their distribution suffers from an $\boldsymbol{x}$–$\boldsymbol{\theta}$ degeneracy. The best approach identified is to estimate $\boldsymbol{\theta}$ from the cluster's

Position-reconstruction method	Resolution [mm]	Absolute error [mm]
Center-of-gravity with linear weights	3.83	3.27
Corrected center-of-gravity with linear weights	1.83	0.25
Center-of-gravity with logarithmic weights	1.88	0.46
Center-of-gravity with square-root weights	3.51	2.88
Adjacent-crystal	2.00	0.74

Table 4.1: A summary of the reconstruction position's resolution and absolute error for a variety of methods tested. These values are found from a population of normally incident, 2-GeV positrons in a `Geant4`-simulation, hitting across a crystal's front face. A crystal's width is 25 mm.

reconstructed position and/or energy, as depicted in Fig. 4.24. As such, a cluster's $\boldsymbol{\theta}$-estimate is left to be calculated post-reconstruction, as needed for a given study.

4.5.9   Chain Fit

The fit algorithm's final step, which is used by default, refines the best-fit result by accounting for any cluster's pulse coupling this island with one chopped before it. Because the typical island length is only 16 % of the template length, if two islands are chopped close in time, the tails from the first positron hit will extend into the second. Further, as seen in Fig. 4.25, the ringing of a pulse's tail can even be large enough to chop another island—without there actually being another positron hit. As each island is processed separately in the foregoing steps, though, the influence from such tails are completely neglected in the χ^2. To remedy it, this step refits the current island's clusters with any earlier ones that have at least one template overlapping in what is known as a *chain fit* in that it chains together islands.

Specifically, each island's χ^2-function in Eq. 4.16 is summed together to form a grand χ^2 that is minimized analogously to in the simultaneous template-fit step above. The fit parameters are held fixed, however, for any cluster whose χ^2-contribution will remain unchanged when including the earlier island's clusters. Moreover, in the grand χ^2, each island's crystal by default still has
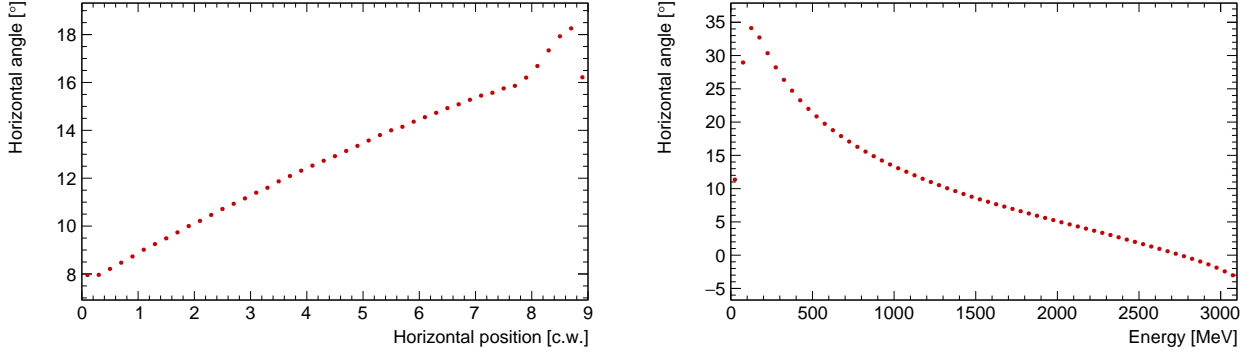
Figure 4.24: The mean horizontal angle of positron hits as a function of horizontal position (left) and energy (right), found using the realistic Geant4-simulation of the experiment discussed in Sec. 4.8. Note that lower positions are closer to the muon storage region.

its own pedestal, which is analytically fit by extending the matrices in Eq. 4.17. This accounts for a possible time-dependent change of the baseline. Afterward, the best-fit result is evaluated in the cluster-merge step and a variant of the cluster-rejection step. In particular, if a cluster is too close in time with the identical window as another or if its amplitude sum is below threshold, it is removed (in that order) from the model. After a cluster is removed, this step then goes back to the grand χ^2-minimization and cluster-merging steps before removing another one. This removal process allows for clusters formed from the tail of an earlier island's cluster to be rectified. When finished, each cluster's position is updated per the previous step's procedure.

In addition, the fit algorithm's computationally slowest part is the MnMigrad routine, which gets increasingly slow with the number of clusters, i.e., fit parameters in the model. After 10 μs into the fill, typically the number of clusters in one island is at most six. The chain fit, however, can cause as many as 26 clusters to be fit at once. As seen in Fig. 4.26, the number of clusters in a chain fit—fit or fixed—spikes below roughly 7 μs after injection, when the particle flux at the calorimeters is much higher. Therefore, to speed it up, the fit algorithm only reconstructs islands by default whose first sample's time is after 9000 c.t.—about 11.25 μs. As the ω_a-measurement's fit starts closer to 30 μs (Sec. 5.7), there is no physics impact from this lower bound.
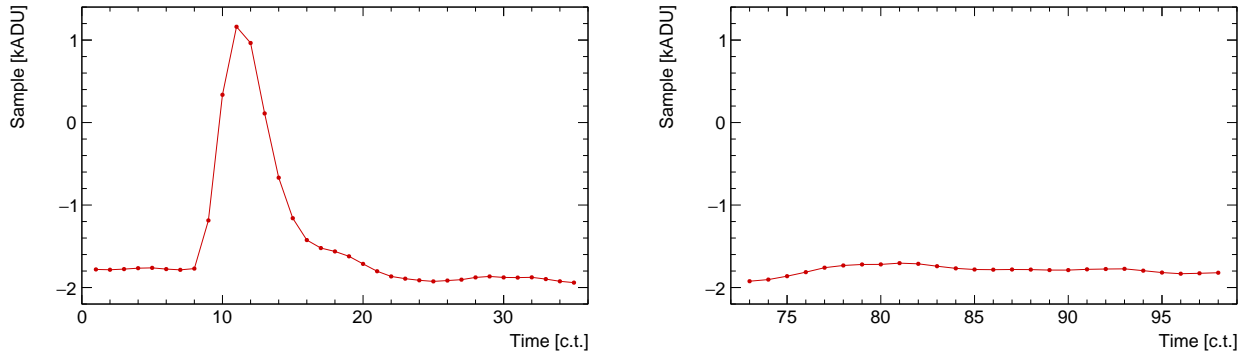
Figure 4.25: An example trace from data of a high-energy cluster's central crystal (left) which exhibited a tail large enough to cause a second island (right) to be saved shortly thereafter. Any clusters reconstructed from a pulse's tail are rectified in the fit algorithm's chain-fit step.

4.6   Energy Calibration

After all of the cluster fits have been refined, their energy—taken thus far to be the sum of the pulse amplitudes—is calibrated from ADU to MeV. To do that, a calibration constant is used for each crystal–SiPM pair by which to multiply its best-fit amplitude. This energy scale for the pair is primarily set by the SiPM's gain and, in turn, its bias voltage and programmable gain amplifier's set point. By tuning these settings, the SiPMs' gains were initially equalized prior to the start of Run 1 to about 10 % of each other by measuring how a laser pulse's amplitude changes when the incident light is attenuated via a neutral-density filter [91]. Further, the absolute energy scale was roughly set so that two piled-up, high-energy decay positrons would not saturate the ADC. This energy-calibration technique cannot, however, account for differences in the crystals' optical properties, e.g., transmittance, which also affect the crystal–SiPM pair's energy scale. To refine the hardware calibration, software calibration constants were found by measuring instead the distinct energy signature of lost muons traversing each crystal [107].

To that end, the basic approach was to align a well-defined, physical feature in each crystal's observed energy distribution. As their stopping power in PbF_2 is nearly energy-independent from
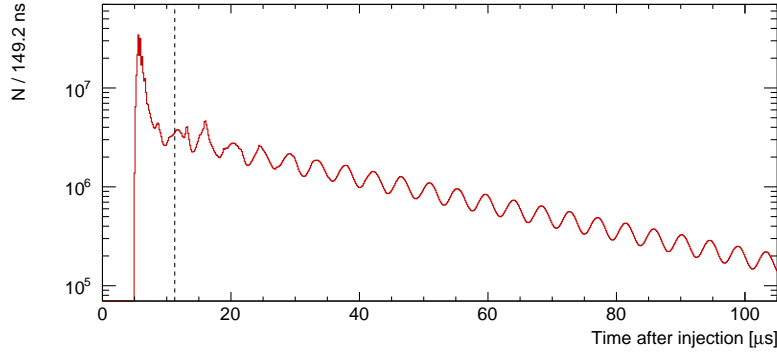
Figure 4.26: The number of clusters reconstructed in islands that were part of a chain fit, found from muon fills in the 1a data set (Sec. 5.1). The dashed line indicates the time below which the fit algorithm normally does not attempt to reconstruct to save computation time.

0.1 to 10 GeV, the energy deposited by a lost muon is governed almost solely by its path length through the calorimeter [22]. Further, if a muon traverses multiple calorimeters as it is lost, the range of possible path lengths is tightly constrained by geometry. Together, this leads to a sharp low-energy peak in each crystal's observed distribution that corresponds to the same amount of deposited energy. These peaks can, therefore, be aligned to each other for a *relative* energy calibration among the crystal–SiPM pairs. Since the SiPM detects Čerenkov photons, for an absolute calibration, the photon yield for the same amount of deposited energy between a muon's ionization and a positron's electromagnetic shower must be related. This extra factor will be estimated as part of the ω_a-measurement in Sec. 5.3.

More specifically, lost muons were identified as triple coincidences from clusters at any time after injection within 1 c.t. windows about their mean time-of-flight, α, found in Sec. 4.4. Further, all coincident clusters were required to have an (uncalibrated) energy less than about 800 ADU, mitigating the accidental background. As shown by Fig. 4.27, the (uncalibrated) energy of the first calorimeter's cluster was then coarsely histogrammed for each crystal—as defined by the reconstructed position. In general, this energy spectrum should follow a Landau function that is somewhat less skewed due to the calorimeter's thickness [22]. Lacking a precise model, its mode
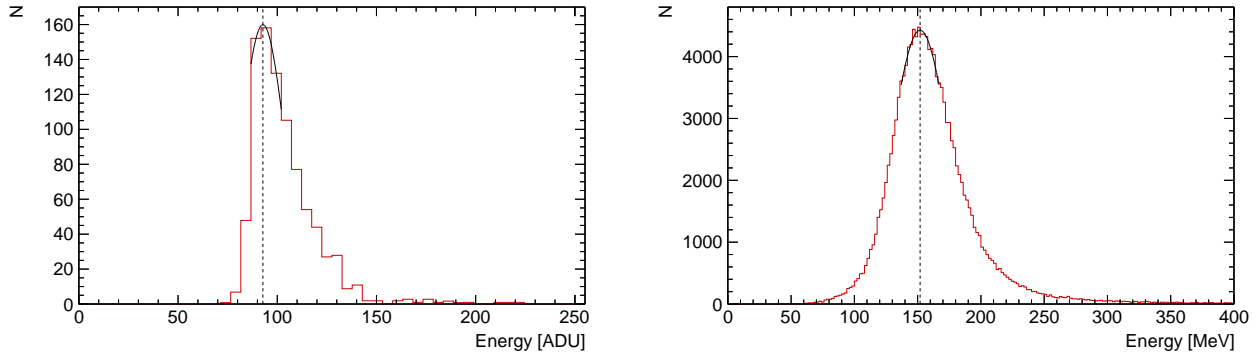
Figure 4.27: The energy distribution of lost muons before (left) and after (right) relative calibration for crystal 16 in calorimeter 17 and all crystals, respectively. The peaks are locally fit to a parabola (black), whose maxima are denoted by dashed lines.

and two adjacent bins were analytically fit to a parabola, whose maximum is taken as the lost muon's characteristic energy, A_c^{peak}, observed—in ADU—in crystal c.

Each crystal's calibration constant, C_c, is then defined by

$$C_c = E/A_c^{\mathrm{peak}} \tag{4.31}$$

to align all of the peaks to E in MeV—a loose estimate of the muon's true amount of deposited energy. For Run 1, E was chosen to be about 152 MeV, as can be seen in Fig. 4.27 [107]. Iterations on this procedure do not significantly change C_c for any crystal c. A baseline energy scale is then equalized across all of the crystals by multiplying their pulses' best-fit amplitudes by its calibration constant in Eq. 4.31. Slow environmental changes, however, will eventually introduce deviations from this baseline, primarily due to the temperature dependence of each SiPM's gain. For that reason, the preceding baseline calibration was carried out on only half an hour of data collected during a special running condition where material was inserted into the storage region to enhance the lost-muon rate [107].

To maintain a stable energy scale over Run 1, any drifts in a SiPM's gain response after the baseline calibration had to be measured and corrected. This was done with the laser-calibration system by firing laser pulses—analogous to the sync pulses—into each crystal when there were

no muons present in the ring. The underlying idea was to compare the size of these *out-of-fill laser pulses* to those observed close in time to when the baseline calibration data was collected. A difference in their amplitudes was then taken to be proportional to the change in the SiPM's gain—an assumption that is correct up to the stability of the laser pulses' intensity. To monitor the laser head's fluctuations shot-by-shot, the response of each SiPM was referenced to that of a suitably stable, independent detector, called the *local monitor*. Specifically, the laser heads' light was partly split and transmitted to each calorimeter station, where its intensity was measured by a photomultiplier tube with a resolution of approximately 2 % [108].

To mitigate the laser head's statistical fluctuations, each of the out-of-fill laser pulses' best-fit amplitudes were averaged over a so-called subrun—about five seconds of data collection. The "long-term" gain reduction for crystal c was then found by

$$g_c^{\text{LT}} = \frac{\langle A(t) \rangle}{\langle A(0) \rangle} \frac{\langle M(0) \rangle}{\langle M(t) \rangle} \tag{4.32}$$

where $\langle A(t) \rangle$ and $\langle M(t) \rangle$ are the average response of the SiPM and local monitor, respectively, to the laser pulse over the subrun. The analogous quantities with $t = 0$ are those averaged over the reference time when the baseline calibration data was taken. To revert back to the baseline gain, each reconstructed cluster's pulse amplitudes must be divided by Eq. 4.32.

Importantly, note that an incorrect long-term gain correction does not directly pose a bias to the ω_a-measurement. As the time scale for environmental changes is much longer than the duration of a fill, it will not be correlated with the oscillations at ω_a beyond a change in statistics from having an energy threshold. This broadly applies to any gain effect that stays constant throughout each 700-μs measurement period, including the intrinsic nonlinearity in the SiPM's response [57]. To reach the experiment's systematic uncertainty goal, changes in each crystal–SiPM pair's energy scale must, therefore, be stable to better than 0.1 % over every fill [38]. This requires two further gain corrections for the so-called in-fill and short-term double-pulse effects.
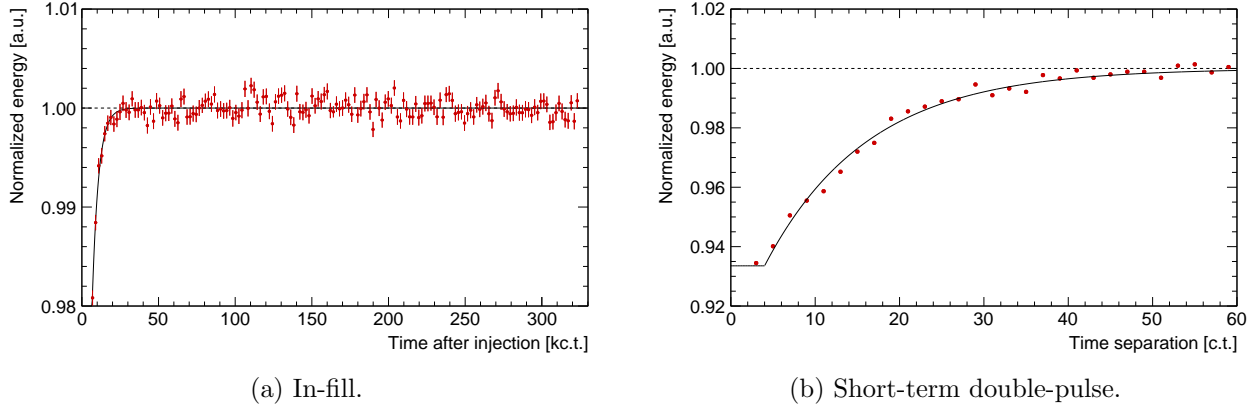
(a) In-fill.

(b) Short-term double-pulse.

Figure 4.28: Representative examples of the (a) in-fill and (b) short-term double-pulse gain recoveries found for crystal 0 in calorimeter 1 [108, 109]. The best-fit gain functions in Eqs. 4.33 and 4.34, respectively, are overlaid (black).

The *in-fill* effect describes the SiPM's bias-voltage recovery from the transient load as the instantaneous hit rate at the calorimeter drops by more than four orders of magnitude over the fill [57]. This leads to an initial gain sag that recovers exponentially with an $\mathcal{O}(1\ \mu s)$ time constant. Each crystal–SiPM pair's exact recovery model was measured *in situ* using the laser-calibration system in an analogously way to Eq. 4.32. In about 1 of every 10 fills, laser pulses—separated by 200 μs—were fired into all crystals at known times after injection. After averaging the statistical fluctuations, their relative pulse heights directly reveal a gain drop, sampled at those times [108]. By systematically cycling the times over the 700-μs measurement period, the crystal–SiPM pair's recovery curve can be mapped out. After normalizing to the local monitors' signals, these data were fit to an exponential-recovery model,

$$g_c^{\mathrm{IF}}(t) = 1 - A_c^{\mathrm{IF}} \cdot e^{-t/\tau_c^{\mathrm{IF}}} \tag{4.33}$$

for each crystal c, as seen in Fig. 4.28a. Typical values of A_c^{IF} and τ_c^{IF} are 4–7 % and 2–4 μs, respectively. To equalize the energy scale then across the fill, every reconstructed cluster's pulse amplitudes is divided by Eq. 4.33. Further, to avoid any ω_a-bias, clusters from the roughly 10 % of fills collected with these laser pulses are excluded from the ω_a-measurement.

The *short-term double-pulse* (STDP) effect arises when two pulses are close in time to each other at a given SiPM. When the SiPM's pixels are fired creating the first pulse, the charge that is delivered must be replenished. This leads to a bias-voltage drop that, based on the SiPM board design, should recover exponentially with a (resistor-capacitor) time constant of about 6 ns [57]. However, in practice, the recovery should be slower due to parasitic elements in the realized circuit. The amplitude of the second pulse, therefore, will be reduced slightly as compared to that from an unloaded SiPM. To characterize this effect, after Run 1, a dedicated running mode was used where two laser pulses were fired into each crystal at a known time separation between 2 and 60 c.t. [109]. Scanning over this time-separation range in 2 c.t. steps, the second pulse's amplitude normalized to that of the first was mapped out (Fig. 4.28b). These data were then fit to the same exponential-recovery model as in Eq. 4.33. Further, the light of the two laser pulses were attenuated by varying amounts using neutral-density filters, revealing a linear dependence of the exponential's scale on the first pulse's height [109]. Therefore, the SiPM-gain reduction between two pulses in crystal c—separated by $4 \leq \Delta t \leq 60$ c.t.—is modeled as

$$g_c^{\mathrm{STDP}}(\Delta t) = 1 - E_1 \cdot A_c^{\mathrm{STDP}} \cdot e^{-\Delta t/\tau_c^{\mathrm{STDP}}} \tag{4.34}$$

where E_1 is the reconstructed amplitude of the first pulse that has been corrected for temperature changes by Eq. 4.32. For time separations below 4 c.t., fitting the two overlapping laser pulses became inaccurate. For that reason, it was decided that the model's value would be fixed to $g_c^{\mathrm{STDP}}(4 \text{ c.t.})$ for $\Delta t < 4$ c.t. Typical values of A_c^{STDP} and τ_c^{STDP} are 3–7 %/GeV and 13–18 ns, respectively. To further equalize the energy scale across the fill, each reconstructed cluster's pulse amplitudes is then be divided by Eq. 4.34 if within 60 c.t. of a spatially overlapping, earlier cluster. Note that, in the case of three or more pulses close enough together, the correction is applied in time order. Also note that the STDP effect was corrected for the in-fill laser pulses used in Eq. 4.33 that closely followed a positron hit.

In summary, in this reconstruction step, each cluster's energy is relatively calibrated to each other by applying a predetermined calibration constant, found from the distinct energy signature of lost muons. This baseline energy scale is subsequently held stable over time by applying a long-term, in-fill, and short-term double-pulse gain correction, as given by Eqs. 4.32, 4.33, and 4.34. Together, a given cluster's energy, E, is calibrated according to

$$E = \sum_c A_c \cdot C_c \cdot \frac{1}{g_c^{\mathrm{LT}}} \cdot \frac{1}{g_c^{\mathrm{IF}}} \cdot \frac{1}{g_c^{\mathrm{STDP}}} \tag{4.35}$$

where the summation is over the crystals in the cluster's window. With this calibration, residual gain changes throughout each muon fill are held sufficiently small for the ω_a-measurement. Some of these gain corrections have also been independently cross-checked, such as by comparing the cluster's reconstructed energy to its momentum measured by the straw tracker [110]. Finally, as detailed in Sec. 4.5, each cluster's position is subsequently recomputed using the newly calibrated energy reconstructed in each of its crystals.

4.7 Cluster Partition

As part of the fit's evaluation in Sec. 4.5, when two partially spatially overlapping clusters are separated by less than an ADT, they are merged into one larger cluster and refit. While necessary for the fit's stability, as a result, the amount of pileup is enhanced. Even for fairly obvious, two-positron cases, e.g., when their 3×3 windows only share a corner crystal, one merged cluster will be reconstructed. To resolve such clear pileup situations, if deemed necessary, a merged cluster can be partitioned back into its originally identified clusters by splitting the shared crystals' energy among them according to a model. Such a *cluster partition* procedure makes up the reconstruction algorithm's final—but optional—step.

For each cluster that was merged, this step begins by relabeling the "central crystal" of the original, un-merged 3×3 windows. When a cluster is first formed, its central crystal is defined as

that with the largest metric and, as such, is assumed to be that with the most energy deposited. However, while infrequent, the crystals' energy calibration can change which crystal has the most energy after all. If that occurred for one of these crystal windows, therefore, the central crystal is reassigned to the one with the largest calibrated energy. Afterward, any crystals that are now outside of a 3×3 window about the original cluster's new central crystal are discarded from its "crystal window" hereunder used.

Further, if two of the central crystals are separated by a crystal, their clusters are very likely from distinct positron hits. For that reason, a cluster merged from N clusters is partitioned into $M \leq N$ clusters if the central crystals are not contained within each other's crystal window, i.e., not the same or adjacent. The energy of each crystal that some of the new clusters share is then split in proportion to the central crystals' energies and spatial separation via an isotropic, exponential energy-deposition model [111]. Specifically, the energy found in crystal i assigned to the partitioned cluster, with central crystal j, is given by

$$E_{i,j} = E_i \cdot \frac{E_j \cdot e^{-|\boldsymbol{x}_{i,j}|/\xi}}{\sum_k E_k \cdot e^{-|\boldsymbol{x}_{i,k}|/\xi}} \tag{4.36}$$

where E_i is the merged cluster's calibrated, best-fit energy in crystal i. The summation in the denominator is over the partitioned clusters whose crystal windows share crystal i. As shown in Fig. 4.21, the characteristic distance of the deposited energy's lateral distribution is empirically estimated to be $\xi = 0.2740$ c.w. This simple model splits the crystal's energy among the clusters correctly on average and is adequate given the experiment's modest requirement on energy resolution for an ω_a-measurement [38].

For the same reasons discussed in Sec. 4.5, the partitioned clusters are not refit. Instead, its time is assigned to be the same as that of the merged cluster, and its energy is assigned to be the sum of the split-up energy for each crystal in its window. Lastly, as done in Sec. 4.5, each new cluster's position is computed using each of its crystals' energy. For a reasonable ADT, this
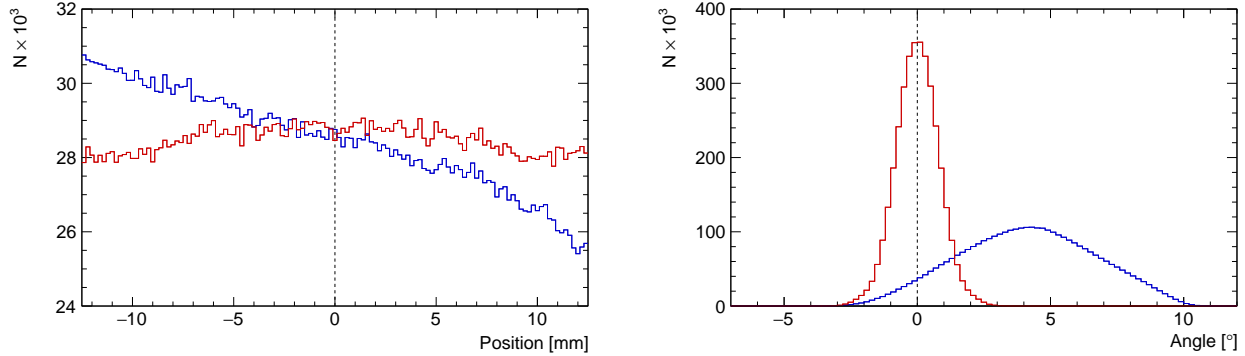
Figure 4.29: The distribution of `gm2ringsim`-simulated, incident positrons' horizontal (blue) and vertical (red) positions relative to the crystal center (right) and angles normal to the crystal face (left). The horizontal position's negative slope arises from the higher-energy positrons having a smaller radius of curvature, resulting in more of them hitting closer to the storage region. Most of the decay positrons also hit with a slight horizontal angle away from the storage region but, on average, no vertical angle.

cluster-partition step should not introduce a net time or energy bias as long as the positron hits' spatial distribution remains constant across the fill. Its key advantage is to reduce the amount of pileup by about a factor of a half [112]. However, when using some pileup-correction methods, it can inflate the total systematic uncertainty on ω_a (Sec. 5.4). In such cases, it may actually be better to allow more pileup but a better pileup correction. Nonetheless, the cluster-partition step serves as an important systematic uncertainty check of the reconstruction together with the pileup-correction method used.

4.8   Simulation Characterization

Aside from its corrections, the reconstruction has been benchmarked using a full-scale simulation of the experiment, developed by many in the collaboration. The simulation, largely subsumed in the `gm2ringsim` software package, is built on the widely used `Geant4` toolkit, interfaced with `art` through the `artg4` framework [113, 114, 115]. Specifically, the decay vertices of stored muons are predicted by closed-form expressions of the muon dynamics, and their decay positrons are tracked
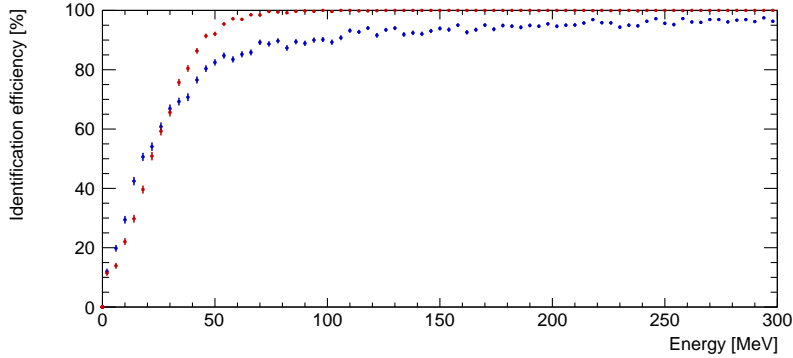
Figure 4.30: The reconstruction's efficiency at identifying single positrons (red) as a function of the incident positron's energy, found from the `Geant4`-simulation. For this study, a fit is defined as successful if the reconstructed time is within 1 c.t. of the true value. For comparison, the same metric is shown when fitting only the central crystal using Newton's method (blue).

via `Geant4` as they curl radially inward. For those entering a calorimeter, their electromagnetic shower is then simulated by `Geant4` with the Čerenkov photons' arrival-time distribution at the SiPMs approximated via a lookup table [116]. Aggregating each detected photon's pixel-response function in Fig. 4.2, raw waveforms are made, modeling such realistic effects as the pixel-gain recovery, sampling-clock jitter, and noise [116]. Each calorimeter's waveforms are next chopped into islands using the same algorithm detailed in Sec. 4.2. All together, this allows for the above reconstruction to be applied to data mimicking those from the real experiment.

As discussed in Sec. 4.1, the characteristic pulse shape is notably dependent on the incident positron's position and angle at the crystal's front face. Such a `Geant4`-simulation is, thus, necessary to provide the expected distribution of impact positions/angles. To that end, the following characterization of the reconstruction was performed on nearly 9×10^6 decay positrons that hit a calorimeter, simulated using the experiment's 2017 commissioning configuration, as shown in Fig. 4.29. For reference, this corresponds to about 10 % of the 1a data set (Sec. 5.1).

To start, for each positron hit, the reconstruction's metric-sum threshold and artificial dead time were tweaked to force a single fit. Labeling the fit as successful if the reconstructed time is within 1 c.t. of the true value, the identification efficiency was characterized as a function of the
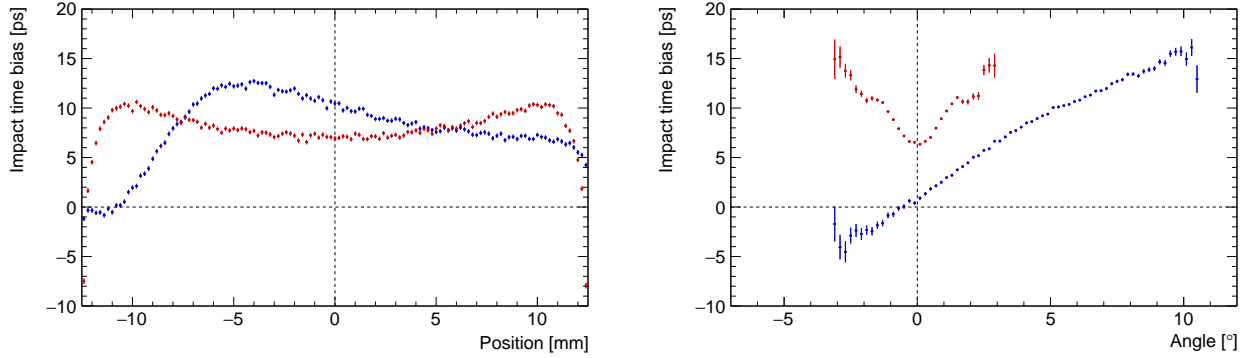
Figure 4.31: The mean time bias for `Geant4`-simulated, ≥1.7-GeV positron hits reconstructed as a function of the incident positron's horizontal (blue) and vertical (red) position relative to the crystal center (right) and angle normal to the crystal face (left). The time is calculated with respect to a fixed reference. The asymmetry in the horizontal position arises from that in the horizontal distribution of angles in Fig. 4.29.
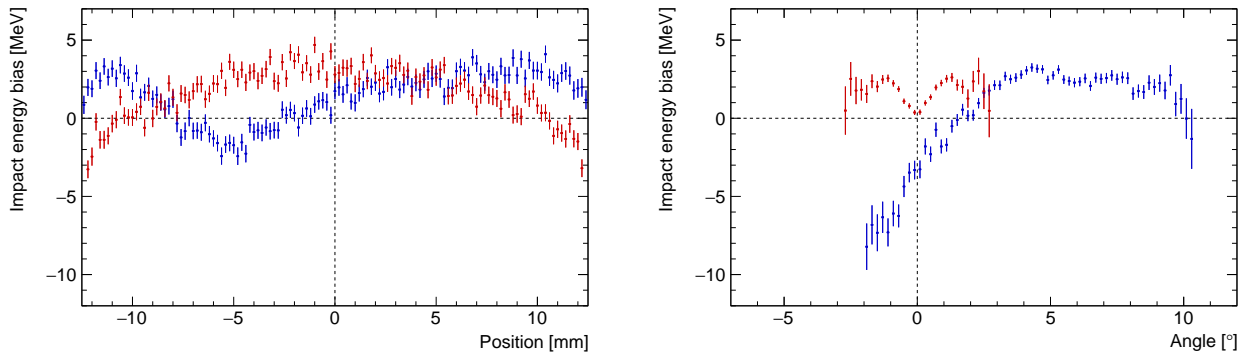

Figure 4.32: The mean energy bias for `Geant4`-simulated, ≥1.7-GeV positron hits reconstructed as a function of the incident positron's horizontal (blue) and vertical (red) position relative to the crystal center (right) and angle normal to the crystal face (left). The asymmetry in horizontal position arises from that in the horizontal distribution of angles in Fig. 4.29.

incident positron's energy. As shown in Fig. 4.30, the reconstruction should always identify any positron with energy above around 80 MeV. Also, simultaneously fitting to the side crystals with the sophisticated χ^2-minimization in Sec. 4.5.5 greatly improves this efficiency. In contrast, perfect efficiency is not achieved until nearly 500 MeV when fitting to only the central crystal's trace using Newton's method. This metric, however, will vary slightly with the different energy scale of each crystal–SiPM pair in the experiment.

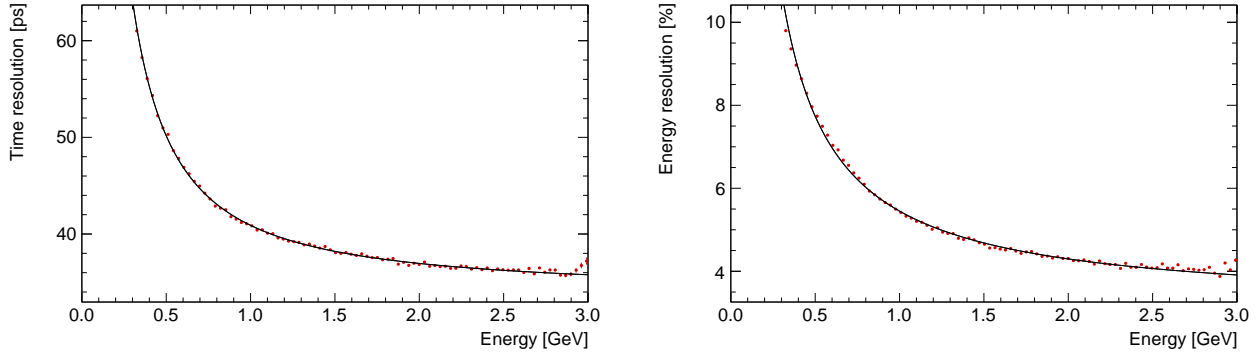Furthermore, the simulation can make known any time and/or energy bias of the positrons re-

Figure 4.33: The time (left) and energy (right) resolution for a single positron hit as a function of the incident positron's energy, found from the `Geant4`-simulation, where their fits—given by Eqs. 4.37 and 4.38, respectively—have been overlaid (black).

constructed in the algorithm. Most notably, for single positron hits above 1.7 GeV, the time and energy biases as a function of the hit position and angle are given in Figs. 4.31 and 4.32. As can be seen, the bias remains below 20 ps and 10 MeV for such high-energy positrons of interest to a typical ω_a-measurement. Also, as expected from Fig. 4.3, the bias seems to be more sensitive to the incident positron's angle rather than position. Further, the mean time and energy biases are fractions of their resolutions. As shown in Fig. 4.33, the time and energy resolution for a single positron hit is well described by the quadratic sum of three terms:

$$\sigma_t = \left[33.7 \text{ ps} \right] \oplus \left[\frac{19.6 \text{ ps}}{\sqrt{E}} \right] \oplus \left[\frac{12.4 \text{ ps}}{E} \right], \tag{4.37}$$

$$\frac{\sigma_E}{E} = \left[3.1 \text{ \%} \right] \oplus \left[\frac{3.8 \text{ \%}}{\sqrt{E}} \right] \oplus \left[\frac{2.3 \text{ \%}}{E} \right] \tag{4.38}$$

respectively. The first of these terms is a constant term, describing the electromagnetic shower's containment; the second is a stochastic term, describing fluctuations related to the shower's development; the third is a noise term, describing the waveforms' electronic noise. These biases cannot bias the ω_a-measurement, though, unless coupled to an early-to-late effect. To identify any such ω_a-bias, a Monte Carlo simulation was conducted with 2×10^{11} positrons, whose times and energies were biased according to Figs. 4.31 and 4.32. For different position oscillations as large as a crystal width, no ω_a-biases were detected [117].
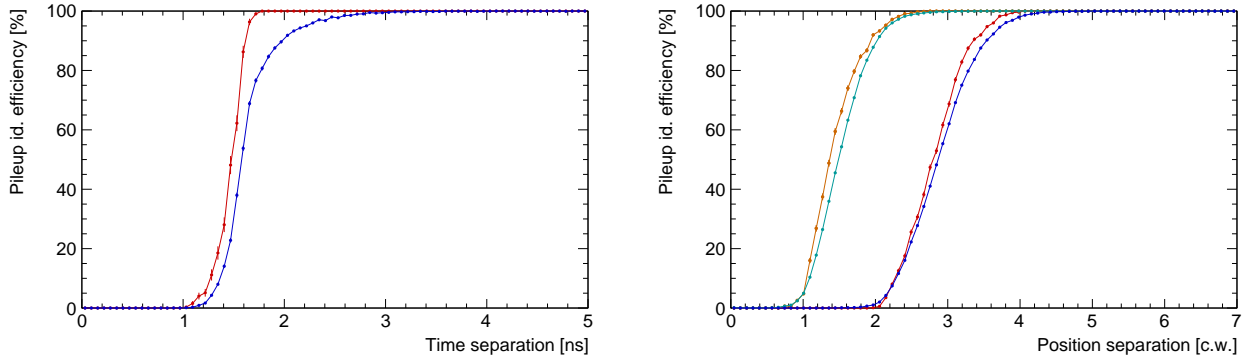
Figure 4.34: The reconstruction's efficiency at identifying two positrons in the same island as a function of their time (left) and position (right) separation, found from the `Geant4`-simulation. This was found without cluster partitioning, where both positrons are of any energy (blue) or above 1.7 GeV (red), and with cluster partitioning, where they are of any energy (cyan) or above 1.7 GeV (orange). Note that, as a function of the time separation, the results with and without cluster partitioning are identical.
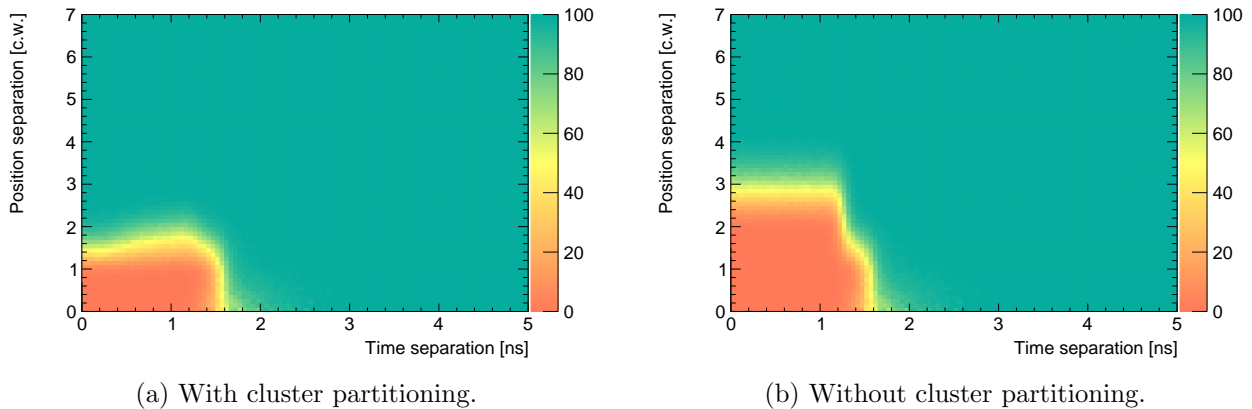

(a) With cluster partitioning.

(b) Without cluster partitioning.

Figure 4.35: The reconstruction's efficiency (as a percentage) at identifying two positrons in the same island as a function of their time and position separations (a) with and (b) without cluster partitioning. These results were found from the `Geant4`-simulation, where both positrons are of any energy and follow the expected position and angle distributions in Fig 4.29.

Moreover, the reconstruction's efficiency at identifying two or more positrons in an island is expected to be a function of their time and position separations from each other. As shown in Fig. 4.34, two positrons separated by ≥ 3 ns or ≥ 4 c.w. are always found. For two high-energy positrons, they will also always be found when separated by only 2 ns. Further, as expected, two positrons can be identified at closer spatial separations if the cluster-partition step (Sec. 4.7) is
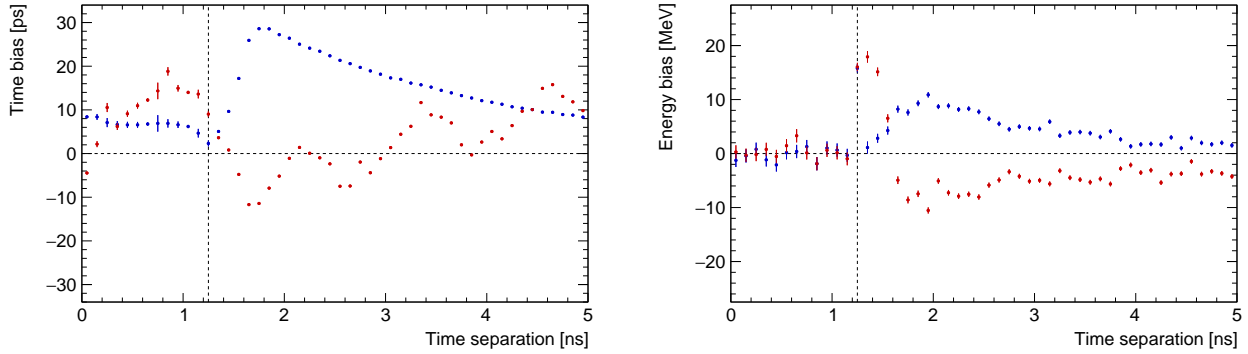
Figure 4.36: The mean time (left) and energy (right) bias of the first (blue) and second (red) positron in time reconstructed in two-positron pileup events. The time is defined with respect to a fixed, reference in the `Geant4`-simulation. Note that, below the artificial dead time of 1.25 ns, it becomes difficult to accurately map the reconstructed to the true positrons.

used. In summary, the *effective* dead time and position is about 1.5 ns and 3 c.w.—or 1.5 c.w. if clusters are partitioned—respectively. However, as seen in Fig. 4.35, a correlation exists between the time and position separations in terms of how well the reconstruction can find nearby positron hits. And, as further seen in Fig. 4.36, the mean time and energy bias of the two reconstructed positrons remains below 30 ps and 20 MeV, respectively—even for time separations close to the artificial dead time in the fit of 1.25 ns.

4.9 Noise Mitigation

Not modeled in this `Geant4`-simulation are noise signals other than white Gaussian noise. There are oscillatory signals, for example, observed throughout the experimental hall around 160 MHz, which pass the SiPM board's frequency filter. They, however, do not appear to be correlated with accelerator or experiment operations. As illustrated in Fig. 4.37, such noise signals can be large enough for an island to be chopped and for a cluster to be reconstructed with energies near 45 MeV. Notwithstanding, the fit algorithm in Sec. 4.5 naturally prevents such clusters from being formed through its thresholds and masking procedure. In total, it mitigates nearly 99 % of these single-frequency wave packets. This efficiency crystal-by-crystal, though, depends on its energy
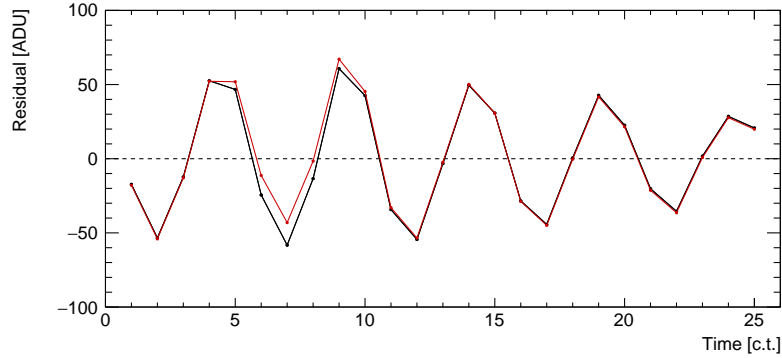
Figure 4.37: An example of a crystal's noise residuals before (black) and after (red) being fit, including those that would be masked out in Sec. 4.5 about the best-fit time. This 160-MHz wave packet is representative of the omnipresent, oscillatory noise found in every calorimeter.

scale, with the worst case being about 91 %.

In review, a new cluster must first pass a metric-sum threshold to verify that it has enough activity to warrant a fit attempt. As shown in Fig. 4.38, approximately 35 % of such possible noise clusters are initially rejected because their peak residuals are collectively too low. This was estimated from clusters reconstructed at late times into the fill—when noise clusters outnumber the decay positrons—for several crystals known to be very noisy. Furthermore, after passing this threshold, the best-fit cluster must also pass an amplitude-sum threshold to determine if it is distinguishable from a fit to only white Gaussian noise. As shown in Fig. 4.38, around 98 % of the noise clusters are then rejected because their best-fit amplitudes are collectively too low. Such a high efficiency is expected as the template pulse shapes do not resemble a single-frequency wave packet. For that reason, the residuals are typically not altered appreciably, as seen in Fig. 4.37, causing about 60 % of the accepted noise clusters being refit. Multiple noise clusters, however, are prevented because the central crystal's metric will become masked. The only consequence of such masking is that positron clusters below about 45 MeV will be missed in that region of the calorimeter, which is acceptable for the ω_a-measurement.

While robust against these wave packets, the fit algorithm is less so against another character-
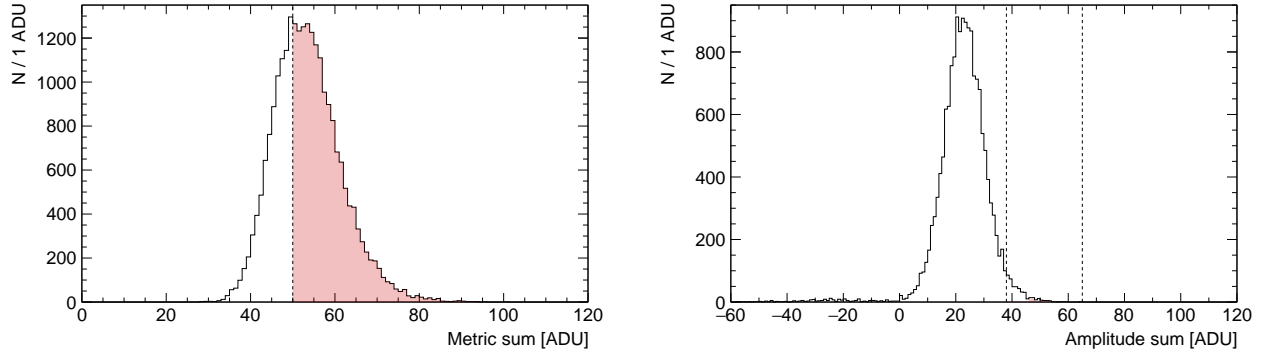
Figure 4.38: Representative distributions of the metric (left) and amplitude (right) sum for noise clusters found in several crystals. Clusters in the shaded regions pass the fit algorithm's threshold in the activity-identification and cluster-rejection steps, respectively, which together amounts to only about 1 %. Note that the cluster-rejection step's threshold depends on the crystal noise levels, being typically between the dashed lines shown.

istic noise signal, created by the periodic polling of the SiPM's temperature. This noise signal's distinctive shape—illustrated in Fig. 4.39—has a width around its peak residual close to that of a positron's pulse. This tends to cause one or two higher-energy clusters being formed and accepted. To flag islands triggered by such characteristic noise, the reconstruction above was augmented to identify the majority of them based on their rapid voltage swing, which is unphysical for positron pulses. In particular, islands are first selected if they were not part of a chain fit, preventing any confusion with an earlier pulse's tail. For each cluster's central crystal, the maximum and minimum sample value in the first 12 samples is found with respect to the pedestal—estimated as the first 6 samples' average. As shown in Fig. 4.39, the cluster is flagged if the positive and negative peaks are below $40y_n$ and $-10y_n$, respectively, where y_n is the maximum value that is expected from only white Gaussian noise, found via normal order statistics.

Analyzing the time of the flagged clusters, as illustrated in Fig. 4.40, they occur at 30-s intervals, which matches the polling frequency of the SiPM's temperature. Within a fill during these noisy times, multiple noise clusters are often found with distinct time separations ranging up to 300 μs—attributed to the electrical communication on the SiPM board. Further, the mean energy
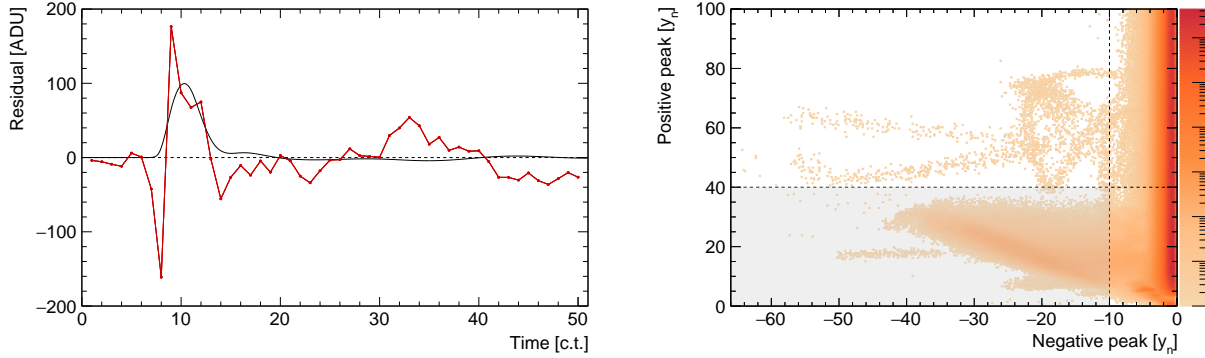
Figure 4.39: Left: The mean, characteristic pulse shape (red) observed from many crystals, with a positron's template (black) overlaid for comparison. Right: The distribution of central-crystal traces' maximum and minimum values in the first 12 samples, in units of that expected from only white Gaussian noise. The clusters falling in the shaded region are flagged by the reconstruction as noise. The dense, vertical band on the left are positron pulses, while the remaining light bands are noise pulses for particular crystal–SiPM pairs.
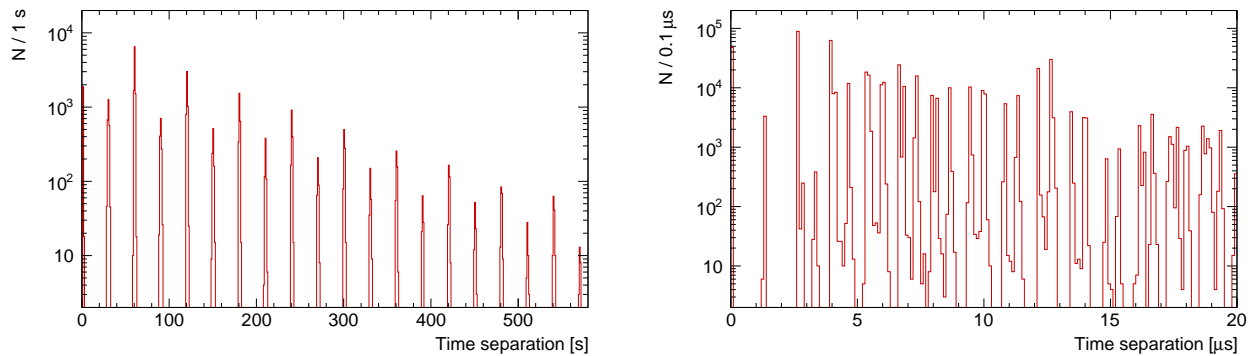

Figure 4.40: Distributions of the time separation between clusters flagged as noise on long (left) and short (right) timescales. Between many muon fills, these noise clusters occur with roughly equal probability every 30 s. Within a fill, they tend to occur multiple times and at distinct time separations that can be as large as 300 μs.

of clusters formed from the characteristic noise signal is around 122.8 MeV, which, as shown in Fig. 4.41, remains constant across the fill as expected. The distribution of their energies is exposed by comparing that observed at early and late times into the fill, as seen in Fig. 4.41. From their deviation, it is concluded that such characteristic noise clusters are reconstructed no higher than 500 MeV—setting a lower bound on the ω_a-measurement's energy threshold.

Nonetheless, these characteristic noise pulses pile up with those from positron hits. To assess
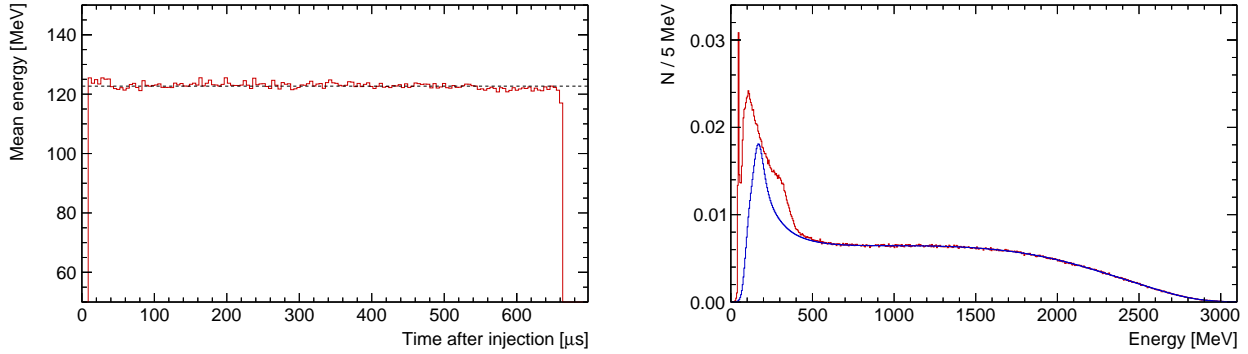
Figure 4.41: Left: The mean energy of clusters flagged as noise throughout the muon fill. A constant fit yields a mean energy of 122.847(81) MeV. Right: The energy distribution of all clusters found in the 1a data set (Sec. 5.1) before (blue) and after (red) 500 μs into the fill, normalized to the integrals between 1500 and 3000 MeV. The noise cluster's energy distribution is exposed by the difference between the early-time (when positrons dominate) and late-time (when noise clusters dominate) distributions. The shoulder around 300 MeV is from crystal–SiPM pairs with a higher energy scale, while the low-energy spike near 45 MeV is instead due to the 160-MHz oscillatory noise in calorimeter 9.

their bias on a positron's best-fit result, a Monte Carlo study was carried out by overlaying the noise's pulse in Fig. 4.39 with a simulated positron's pulse. The summed waveform was then fit to a positron pulse template and a pedestal. Sweeping over the pulses' time separation, a negligible net time bias was found. The positron's energy, on the other hand, was slightly biased by about -1 MeV for the energies above 1 GeV since the noise tends to raise the fitted pedestal. However, as the SiPM's polling is asynchronous to the muon injection, there is no recognized early-to-late bias mechanism to the ω_a-measurement. For that reason, the reconstruction does not attempt to remove any of the flagged clusters from the data, and, in fact, the pileup-correction method used in Sec. 5.4 benefits from keeping such noise clusters.

CHAPTER 5

ANOMALOUS PRECESSION FREQUENCY MEASUREMENT

The first muons were delivered to the Muon $g - 2$ experiment at Fermi National Accelerator Laboratory (FNAL) on May 23, 2017. For the ensuing nine months, the experiment's many systems were commissioned and tuned to maximize the number of muons stored, which is paramount to reach the experiment's target statistical precision on a_μ in the four years budgeted. By the end, approximately 600 high-energy positrons from muon decays were being consistently detected in each fill. From then on, the experiment's state was held as stable as possible (until the accelerator's annual shutdown) for a four-month period called Run 1. This chapter presents a measurement of the muon's anomalous precession frequency, ω_a, on the data taken in Run 1, which have been reconstructed using the previous chapter's algorithm. As detailed in Ch. 2, this chapter's result can be directly used to determine the muon's anomalous magnetic moment, a_μ. The systematic uncertainty on the ω_a-value found will be thoroughly assessed in the next chapter.

5.1 Data Selection

While the experiment's systems were held as stable as possible throughout Run 1, some of them experienced hardware malfunctions that prevented them from operating as intended. In particular, the plates forming the electrostatic quadrupoles and pulsed kicker magnet were unable to reach their design high voltages. Furthermore, these systems' voltage set points changed across Run 1 based on the hardware's condition at the time. This in turn varied the muons' dynamics, as discussed in Ch. 2, making it difficult to analyze the combined data taken in Run 1. Therefore, the full data set is divided into four subsets—defined in Tab. 5.1—with unique combinations of the quadrupole and kicker voltage settings. These smaller data sets, labeled from 1a to 1d in time

Date	Label	Nickname	n	Kicker setting	High-energy e^+
April 22 – April 25	1a	60 hour	0.108	130 kV	1.2×10^9
April 26 – May 2	1b	High kick	0.120	137 kV	1.1×10^9
May 4 – May 12	1c	9 day	0.120	130 kV	2.9×10^9
June 6 – June 29	1d	End game	0.108	125 kV	5.6×10^9

Table 5.1: The data sets taken during Run 1 in 2018, defined by their unique combinations of the quadrupole—quoted in terms of the field index, n, given by Eq. 2.12—and kicker voltage settings. The data sets are labeled by the run number and order in which they were collected but are more commonly referred to by their nicknames within the collaboration. Note that the kicker voltage is the sum over the three plates. The relative size of each data set can be seen by the approximate number of reconstructed positrons with an energy above 1.7 GeV.

order, can then be fit separately for ω_a, accounting for their unique muon dynamics.

Notwithstanding stable set points, the plates' voltages were occasionally disrupted by, for example, dielectric breakdown. Moreover, from time to time, the accelerator complex would deliver muon bunches with atypical profiles as measured by the T0 entrance counter. Such brief instabilities necessitated that data-quality cuts be applied to each data set's muon fills to select only the most stable ones for analysis. This avoids introducing possible systematic ω_a-biases from how the muon dynamics are handled in this chapter. It should also be noted that some data were taken for specific studies in unusual running conditions. For example, about 10 % of fills collected contain laser pulses used to correct the in-fill gain effect (Sec. 4.6). Such fills were also excluded from each data set's ω_a-measurement.

The data-quality cuts, more specifically, define the acceptable range on a number of metrics formed from diagnostic signals of the experiment's systems. For each data set separately, these ranges were chosen to exclude any muon fill—or period of muon fills—appearing atypical. Details of such cuts can be found in [118]. To summarize, beginning with the quadrupoles, voltage and current readouts from the high-voltage power supplies were used to ensure a stable plate voltage over the fill after the scraping procedure (Sec. 2.4). For the kicker, induction coil sensors were used

to ensure a stable profile of the pulsed voltage. For the reconstruction's time alignment (Sec. 4.4), each fill's laser sync pulse in the T0 counter and in every crystal was required to be successfully fit. After being time-aligned, the injected muon-bunch profiles in the T0 counter were then required to be stable. Further, to ensure a consistent muon storage efficiency, the number of lost muons and high-energy positrons reconstructed in the calorimeters were also used as metrics. Finally, since ω_a will be combined with the magnetic field's measurement to obtain a_μ, any time periods with an anomaly in the field strength or a malfunction in the field equipment were excluded from the ω_a-measurement. In total, around 10–20 % of each data set's muon fills were removed to ensure only those suitable for this chapter's analysis were used [118].

5.2   Blinding Procedure

As introduced in Sec. 2.5, ω_a will be determined by fitting a time histogram of (possibly weighted) positrons above an energy threshold and, being a high-precision measurement, may be sensitive to subtle procedural differences. All ω_a (and ω_p) measurements conducted by the Muon $g-2$ collaboration, therefore, are double-blinded to preclude any personal biases for the final a_μ-result to agree or not with the theoretical prediction. That is, this dissertation's ω_a-value for each data set is offset in both hardware and software by a number unknown to the author. Both offsets will not be revealed until all of the ω_a (and ω_p) analyses are finalized and the collaboration unanimously approves of them, which is expected to occur in the coming months.

The hardware-level blinding was implemented by offsetting the primary, 40-MHz clock—which drives the backend electronics—by ε such that

$$-3 \text{ kHz} \leq \varepsilon \leq -1 \text{ kHz} \tag{5.1}$$

[119]. This range of ε can, therefore, shift the fitted value of ω_a by as much as 75 ppm—much larger than the previous ω_a-measurement's discrepancy with theory of about 2.2 ppm (Sec. 1.3).
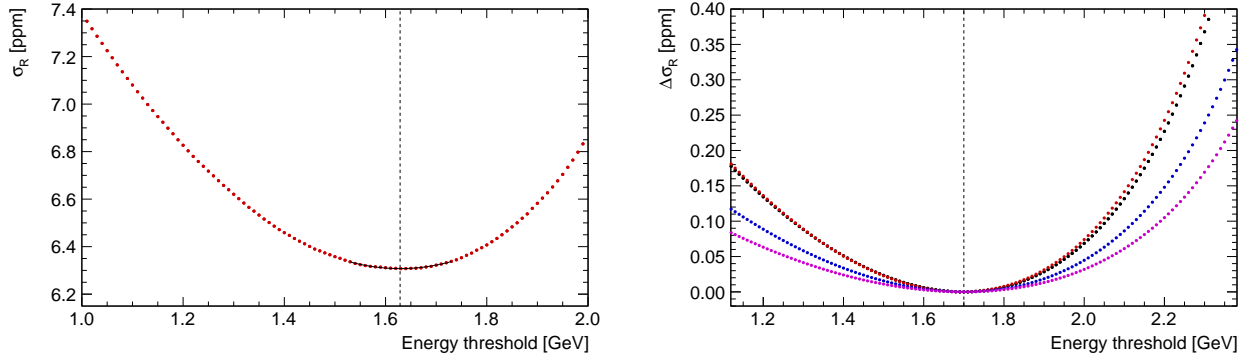
Figure 5.1: Left: The uncertainty of R using increasing energy thresholds before absolute energy calibration of calorimeter 1 in the 1a data set. Right: The calorimeter-combined fit's uncertainty on R, relative to the smallest value sampled, for increasing energy thresholds after applying the absolute energy calibration, using the 1a (black), 1b (red), 1c (blue), and 1d (magenta) data set.

More importantly, this range is within the dual PLL-lock range of the waveform digitizers' clock synthesizers, which can tolerate offsets between -5.2 and 5.8 kHz (Fig. 3.13). As a reminder, every digitizer's lock status was continuously monitored across Run 1 and would have prevented data collection if any one had ever become unlocked. The ε-offset was initially set by two employees of FNAL who are otherwise not involved in the experiment [119]. Every week, they verified (in secret) that the frequency output agreed with the set point to within 3 mHz—for a negligible uncertainty of 0.075 ppb. Between these measurements, the clock's stability was measured according to the procedure described in Sec. 3.5. As ε has yet to be revealed, the time-based data presented throughout this dissertation are tacitly accurate only up to the secret frequency offset.

The software-level blinding was implemented by offsetting the hardware-blinded ω_a value returned by the fit. This offset is derived from the hash of a string chosen by (and only known by) the analyzer to ensure independence among different ω_a-analyses. Each analysis is, therefore, free from potential influence by what another analysis found. More specifically, in every analysis, ω_a is replaced by a blinded parameter, R, according to

$$\omega_a(R) = 2\pi \cdot (0.2291 \text{ MHz}) \cdot \left[1 + (R + \Delta R) \cdot 10^{-6}\right]. \tag{5.2}$$

Here, R is the difference—in units of ppm—from a reference ω_a-value of $2\pi \cdot (0.2291 \text{ MHz})$ and re-places ω_a as one of the fit's free parameters. The analysis-specific, secret offset is introduced by ΔR, which is drawn from a ± 24-ppm flat distribution with 1-ppm Gaussian tails using a random number generator. Accordingly, this chapter will report all ω_a-measurements in terms of R. To allow for relative comparisons among the data sets, however, a common, non-zero ΔR offset has been used—which is known only to the author.

5.3   Absolute Energy Calibration

For each data set, the waveform digitizers' recorded traces were reconstructed using the previous chapter's algorithm in its default configuration. As part of the algorithm, the clusters' energies were calibrated relative to each other by aligning the energy signature of lost muons traversing the calorimeters. However, the photon yield is slightly lower for a muon's ionization than a positron's electromagnetic shower (Sec. 4.6). An extra calibration constant, therefore, must be multiplied to each cluster's energy to convert it to an absolute scale. It should be stressed, though, that an in-accurate calibration constant cannot introduce an early-to-late effect and, thus, cannot bias the ω_a values found in this analysis (Sec. 2.6).

A straightforward way to find this calibration constant is from the energy threshold, E_{th}, that maximizes the best-fit ω_a's statistical precision. As seen in Fig. 2.7, in the baseline method with weights of $w(E) = 1$, the statistical precision is optimized at a well-defined, non-zero E_{th}. When simulating the calorimeter's acceptance and energy resolution, this optimal threshold is about 1.9 GeV. Hence, the absolute energy can be calibrated by defining the empirically found optimal E_{th} to also be 1.9 GeV. This makes, however, the end point of the positron's energy distribution (after correcting for pileup in the next section) to be markedly beyond 3.1 GeV—the muons' energy. At the level of 0.1 GeV, it was found that defining the empirically found optimal E_{th} to be 1.7 GeV

resulted in a reasonable energy end point and, thus, was used instead.

More specifically, to find the optimal E_{th} empirically, time histograms were constructed from 0 to 699.8972 μs with a 0.1492-μs bin width. This range of times after the muons' injection spans the full duration in which the waveform digitizers were recording data for each fill. The time bin width corresponds approximately to the muons' cyclotron period and, as explained in Sec. 5.6, is used to help mitigate the effects of the muons' debunching over the fill. The histograms were then populated with reconstructed clusters above an energy threshold that ranged (in relatively calibrated units) from 1.00 to 2.50 GeV in intervals of 0.01 GeV. Next, they were fit to the minimal, five-parameter model given by

$$N(t) = N_0 \cdot e^{-t/\tau} \cdot [1 + A_0 \cdot \cos\left(\omega_a(R) \cdot t + \phi_0\right)] \tag{5.3}$$

using the procedure in Sec. 5.7. Neglecting muon dynamics and pileup effects, this is the model expected from first principles by combining Eqs. 2.29 and 5.2. It also describes the histogrammed data well enough for the fit's uncertainty on the best-fit R, σ_R, to be meaningful. Afterward, the σ_R-values within 0.1 GeV of the lowest σ_R sampled were fit to a parabola. The best-fit parabola's minimum was then defined to be the optimal E_{th}.

Furthermore, the optimal E_{th} will be slightly different among the calorimeters due to varying amounts of material located around the storage ring. This alters, for instance, the probability that a decay positron hits a particular calorimeter, which in turn alters ω_a's—or, equivalently, R's— expected statistical precision in Eq. 2.30. For this reason, the optimal E_{th} was found separately for each calorimeter using the above procedure with the 1a data set, as illustrated in Fig. 5.1 for calorimeter 1. All of the calorimeters' optimal E_{th}-values are reported in Tab. 5.2, which indeed have a statistically significant spread of about 55 MeV. The mean optimal threshold across the calorimeters, on the other hand, is around 10 % lower than 1.7 GeV.

As the spread in Tab. 5.2 is relatively small, all of the calorimeter's optimal E_{th}-values were

Calorimeter	Threshold	Calorimeter	Threshold	Calorimeter	Threshold
1	1628.9	9	1518.1	17	1520.0
2	1505.9	10	1551.7	18	1588.2
3	1559.4	11	1582.6	19	1554.9
4	1564.9	12	1610.8	20	1525.5
5	1368.8	13	1604.2	21	1455.7
6	1516.9	14	1566.5	22	1474.9
7	1543.8	15	1528.0	23	1522.5
8	1533.0	16	1487.0	24	1548.1

Table 5.2: The optimal energy threshold $(E_{\text{th}})_c$ for each calorimeter c, in uncalibrated MeV, that minimizes ω_a's uncertainty. These values are calibrated to be 1.7 GeV. The uncertainty on each $(E_{\text{th}})_c$-threshold was found to be at most, at most, 0.9 MeV.

chosen to be calibrated as 1.7 GeV. That is, for calorimeter c, each of its reconstructed clusters' energies were multiplied by the calibration constant

$$C_c = 1.7 \text{ GeV}/(E_{\text{th}})_c \tag{5.4}$$

where $(E_{\text{th}})_c$ is the optimal E_{th} found for calorimeter c. This enables R's statistical precision to be optimized in the 1a data set for a *single* energy threshold, $E_{\text{th}} = 1.7$ GeV. As a consistency check, which is shown in Fig. 5.1, sweeping over E_{th} in a calorimeter-combined fit after absolute calibration indeed yields an optimum very close to 1.7 GeV. When performed on the other data sets using the same absolute calibration, the optimum from calorimeter-combined fits remains very close to 1.7 GeV. Therefore, in this dissertation, the same $(E_{\text{th}})_c$ constants in Tab. 5.2 were used for every data set analyzed in Run 1.

5.4 Pileup Correction

In a muon fill, some decay positrons will arrive at a calorimeter too close in time and space to be individually resolved by the reconstruction algorithm. As discussed in Sec. 2.6, these *pileup* occurrences change not only the number of recorded positrons but also the muons' encoded spin-

precession phase. This phase difference poses a direct bias to ω_a because the fraction of pileup is not constant into the fill, causing a phase-pulling effect. To attain the experiment's target systematic uncertainty on ω_a, the presence of pileup clusters in the reconstructed data must then be either incorporated into a more-complete model or corrected for prior to fitting for ω_a.

The leading contamination from pileup in the data will be from the most likely scenario when just two positrons arrive simultaneously in time and space at a calorimeter. As the positrons are independent, the time dependence of such *double pileup* will follow the square of the underlying positrons' hit rate, $\rho(t)$, which will in turn follow the five-parameter model in Eq. 5.3 [78]. The most straightforward approach to account for pileup in the data, therefore, is to augment the fit function in Eq. 5.3 by the leading pileup term

$$N(t) \supset N_{\text{pileup}} \cdot e^{-2t/\tau} \cdot \left[1 + A_0 \cdot \cos\left(\omega_a(R) \cdot t + \phi_{\text{pileup}}\right)\right]^2 \tag{5.5}$$

where N_{pileup} and ϕ_{pileup} are two new free parameters. In this seven-parameter model, however, both ϕ_0 and ϕ_{pileup} couple to ω_a with equal strength, inflating the best-fit ω_a's uncertainty by a factor of $\sqrt{2}$ [120]. This decrease in precision is unacceptable to achieve our target uncertainty on ω_a, leading the author to take the alternative approach of correcting for pileup in each data set's time histogram prior to fitting for ω_a.

The underlying concept in any pileup-correction method is a model for how the reconstruction would have behaved if two (or more) positrons were overlapping in time and/or space. This could, for instance, be estimated from simulation, possibly with some effects being neglected. However, while achievable, it would likely be challenging to implement and then maintain should the reconstruction and/or simulation ever change. Hence, in this section, a more empirical approach is introduced where pileup events are artificially constructed by combining islands together that are then passed through the algorithm in Ch. 4, allowing for a *per-event* modeling of the reconstruction's behavior. Each data set's pileup distribution will thereby be deduced and subtracted from

the time histogram's data—so that no new fit parameters are needed.

5.4.1   Pileup Correction Approach

The approach, in general, is to artificially create known pileup events by combining islands in the data at the waveform level. To that end, for each fill and calorimeter, the first step is to identify pairs of reconstructed clusters, called *pileup candidates*, which share at least one crystal and are separated in time within δt of Δt. The idea is that these candidates should be any pair of clusters that could possibly become unresolved if overlaid. As the reconstruction algorithm has a 100 % pileup identification efficiency for time separations beyond three clock ticks (c.t.) at any energy (Fig. 4.34), δt is generously set to 4 c.t. ≈ 5 ns.

The choice of Δt, on the other hand, is less well-defined. A minimum of $2\delta t$ can be set to ensure that the reconstruction's dead time around the two clusters do not overlap. From this minimum to about 60 c.t., the second cluster's gain will be reduced from the short-term double-pulse effect (Fig. 4.28), and, as a result, its waveforms will be heavily influenced by the first cluster's tail. In addition, the effect of the muons debunching will oscillate the detected-positron rate at the cyclotron period. As an example, in the 1a data set, there are 8.01 ± 0.25 % fewer pileup candidates if Δt is a half-integer multiple rather than an integer multiple of the cyclotron period. The two clusters, therefore, ought to be sampled when the pileup rate is as close as possible. To that end, Δt is set to 119.36 c.t. ≈ 149.2 ns—one cyclotron period.

With the pileup candidates in hand, the next step is to remove any of them that were identified from the same island(s) so as to not double count. An example would be a three-positron event with the first two somewhat close in time and space to each other (but resolved), producing two pileup candidates with the third. In the following procedure, it does not matter which candidate is removed. For each pileup candidate, their raw islands are then prepared and combined in a se-

ries of six steps, which will be expounded next. If a waveform does not exist because of the island chopper's algorithm (Sec. 4.2), it is artificially generated as white Gaussian noise with a standard deviation equal to the crystal's noise level using the `Ranlux64Engine` random number generator in the `CLHEP` software package [121].

First, any underlying tails from earlier clusters are subtracted from each waveform according to the template for the crystal. As the templates last out to 180 c.t., the later cluster, for example, will—by construction—always have a very small component of the first cluster's tail(s). At a low rate proportional to the cube of the positron rate, a third cluster could also overlap with a tail underneath one of the clusters of the pileup candidate.

Second, the short-term double-pulse gain effect—in which a given pulse lowers the amplitude of any subsequent pulse that arrives after it up to roughly 60 ns (Sec. 4.6)—is corrected. For all clusters in both islands, if any pulse's energy was modified—by a factor of $1/g_c^{\mathrm{STDP}}$—due to this effect, an identical pulse but with an amplitude scaled by $1/g_c^{\mathrm{STDP}} - 1$ is added to its waveform's samples. This procedure, in essence, artificially boosts the gain-reduced signal, while not changing the reconstruction's χ^2-minimum if refit.

Third, the time of the first sample in the second cluster's island is shifted by $-\Delta t$, where Δt is rounded down to the closest integer number of clock ticks. Note that the two clusters' peaks are not forced to be aligned, so a representative spread of time separations will remain.

Fourth, any samples at times not shared between these time-shifted waveforms are dropped. They will, otherwise, need to be artificially created using the template model and/or noise around its pedestal; however, if there is a mismatch between the true and the estimated pedestal, a discontinuous step in the waveform will be created, which the reconstruction will mistakenly interpret as a signal and try to fit with a low-energy cluster. This trimming step avoids such a situation. On average, 1.6060(6) and 1.8492(7) samples are removed from the beginning and end of one of the

islands, respectively. The maximum number of samples that can be removed from the beginning or end of either island is the Δt used to shift the second island, which occurs when both clusters are within the same island.

Fifth, the correction to the ADC's interleaved pedestal offset is added to each remaining sample in the second cluster's island. Each island's waveform is then shifted to have half the average pedestal between the islands, which is computed from the cluster's template model or, if missing, from the average of the samples. Sixth, the shared samples between the islands are summed to create a new island, which is run through the reconstruction in Ch. 4—but without chain fitting or the short-term double-pulse gain correction—to form a "double" distribution $\rho_{1+2}(E, t)$.

Separately, after the fourth step, the two islands' (uncombined) waveforms are added to a waveform of white Gaussian noise, as performed above. This causes the individual islands have noise levels similar to the (combined) island that is processed in the sixth step. The two islands are then also run through the reconstruction, as above, to form the "single" distributions $\rho_1(E, t)$ and $\rho_2(E, t)$. The pileup spectrum can then be estimated by

$$\delta\rho_{\mathrm{pileup}}(E, t) = \frac{\rho_{1+2}(E, t) - \rho_1(E, t) - \rho_2(E, t)}{2} \tag{5.6}$$

which can be thought of as the number of clusters—at a given energy and time—that were mistakenly included or excluded. The *pileup correction* is then performed by subtracting $\delta\rho_{\mathrm{pileup}}$ from the measured distribution in each energy–time bin prior to fitting for ω_a.

The factor of $1/2$ in Eq. 5.6 arises from combinatorics. In combining two islands, we are estimating how often a cluster overlaps with another within the reconstruction's dead time such that the individual clusters are unresolved. By allowing the second cluster of a pileup candidate to be δt on either side of Δt from the first, we are thus spanning a time range that is *twice* the dead time. This overcounting is intentionally done to obtain a better estimate of the pileup by statistically sampling more of the data.

In addition, note that the above approach implicitly assumes that the positron rate at times t and $t + \Delta t$ are equal. Neglecting the reconstruction's dead time, the pileup rate actually sampled is the product of these two rates, which we would like to say is

$$\rho(t) \cdot \rho(t + \Delta t) = \rho(t') \cdot \rho(t') \qquad (5.7)$$

with the effective time, t', being simply t. A closer approximation, however, is made by solving Eq. 5.7 for t' given the form of $\rho(t)$. If we consider only the exponential-decay factor, i.e., $\rho(t) \propto e^{-t/\tau}$, it has the exact solution of $t' = t + \Delta t/2$. When the muon-precession factor, $1 + A_0 \cos(\omega_a t + \phi_0)$, is included, though, $t' - t$ becomes dependent on t, bounded between 0 and Δt. For example, the two ρ-factors in Eq. 5.7 can offset each other whenever the muon precession rises more than the exponential falls in Δt, causing $\rho(t) = \rho(t + \Delta t)$.

For a fixed offset from t, the pileup-rate error oscillates at ω_a and lags behind the muon precession by $\Delta t/2$. While this error depends on the asymmetry A_0—which in turn depends on the energy threshold used—it always has a minimum around $t' = t + \Delta t/2$ and maximum around $t' = t$ and $t' = t + \Delta t$. Therefore, in this approach, the reconstructed cluster times are shifted by $\Delta t/2$ before populating the $\rho_1(t)$, $\rho_2(t)$, and $\rho_{1+2}(t)$ distributions in Eq. 5.6. The sub-percent level shape distortion to $\delta\rho_{\text{pileup}}$ from the residual pileup-rate error will be discussed in Sec. 6.4.

This novel approach to correcting for pileup, on average, in the reconstructed data was applied separately to each data set. Examples of the correction's input distributions in Eq. 5.6 are shown in Fig. 5.2. As expected, the single distributions (ρ_1 and ρ_2) closely match each other even though they ought to have slightly different contamination levels from higher-order pileup due to their Δt separation. Also as expected, the double distribution (ρ_{1+2}) has significantly more counts with energies $E \geq 3.1$ GeV than the sum $\rho_1 + \rho_2$ due to the artificially created pileup.
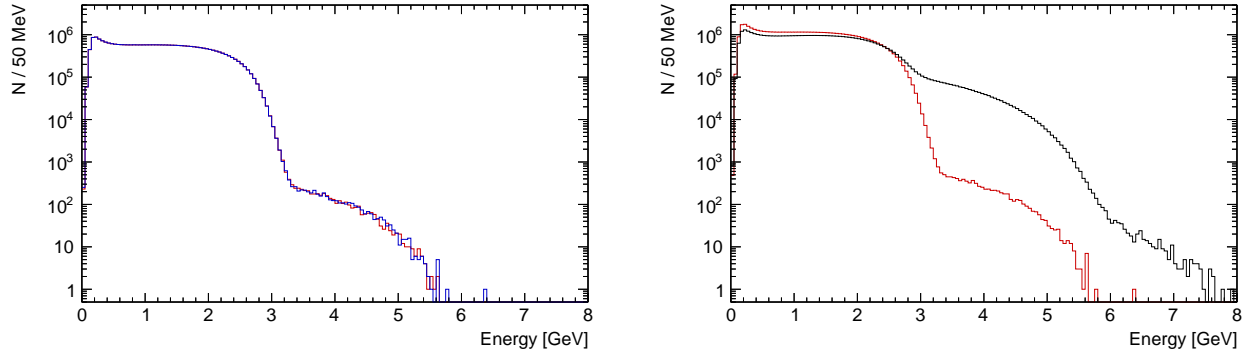
Figure 5.2: Left: The single distribution of $\rho_1(E)$ (red) and $\rho_2(E)$ (blue) used for the pileup correction in Eq. 5.6. Right: The distribution of $\rho_{1+2}(E)$ (black) used in Eq. 5.6 and, as a reference, $\rho_1(E) + \rho_2(E)$ (red). These data are from the 1c data set.

5.4.2 Higher-Order Pileup Correction

More than two positrons, of course, can also overlap in time and space such that they are all unresolved by the reconstruction. Each higher order of such M-positron pileup, however, is increasingly unlikely to occur. In particular, neglecting the reconstruction's resolution, the rate of M-particle pileup should go as the positron rate raised to the power of M [78]. The pileup correction found in the previous section will, thus, need a correction to account for these higher orders in the data. To begin, first consider three-positron pileup—the next highest order.

At a rate roughly proportional to $e^{-3t/\tau}$, one of the pileup candidate's clusters, which was treated as a single positron in Sec. 5.4.1, will have actually been two unresolved positrons. When the islands are then combined, a triple pileup event can occur. Schematically, such a process is able to be formally written as

$$\rho_1(E,t) + \rho_{2+3}(E,t) \rightarrow \rho_{1+2+3}(E,t) \tag{5.8}$$

where third positron could be overlapping with either the first or the second positron. Following Eq. 5.6, this would result in a pileup spectrum contribution of

$$\rho_{1+2+3}(E,t) - \rho_1(E,t) - \rho_{2+3}(E,t) \tag{5.9}$$

with an overall factor of 1/2 that will be tacitly implied. However, applying the logic behind Eq. 5.6 to the three-positron case, the proper contribution would have been

$$\rho_{1+2+3}(E,t) - \rho_1(E,t) - \rho_2(E,t) - \rho_3(E,t) \tag{5.10}$$

which differs from what was actually done in Eq. 5.9 by

$$\rho_{2+3}(E,t) - \rho_2(E,t) - \rho_3(E,t). \tag{5.11}$$

This difference between what was wanted and what was done is the first of two corrections to $\delta\rho_{\text{pileup}}$ from triple pileup, correcting for the *input* to the procedure. As the third positron has a nearly identical chance of overlapping with either positron, the contribution in Eq. 5.9 due to both cases of triple pileup will be twice what was observed. Therefore, the second correction to $\delta\rho_{\text{pileup}}$ is the negative of its contribution in Eq. 5.10, correcting for the *output* to the procedure.

The triple-pileup correction to be added to Eq. 5.6, put together, is then as follows

$$
\begin{aligned}
\delta\rho_{\text{correction}}(E,t) = & - \left[\rho_{1+2+3}(E,t) - \rho_1(E,t) - \rho_2(E,t) - \rho_3(E,t) \right]/2 \\
& + \left[\rho_{1+2}\ \ (E,t) - \rho_1(E,t) - \rho_2(E,t) \right] \\
& + \left[\rho_{2+3}\ \ (E,t) - \rho_2(E,t) - \rho_3(E,t) \right]
\end{aligned}
\tag{5.12}
$$

where the factors from combinatorics, i.e., the number of dead times swept, are included.

Appropriately, Eq. 5.12 must be applied at the rate at which triple-pileup events occur in a given calorimeter. To estimate this rate, the procedure in Sec. 5.4.1 is extended to include a third time window—separated by $2\Delta t \pm \delta t$—in the definition for a pileup candidate. All three islands are then combined together and run through the reconstruction chain in an analogous manner to Sec. 5.4.1 to form a "triple" distribution $\rho_{1+2+3}(E,t)$. In this way, with the underlying double and single distributions, the correction in Eq. 5.12 to the pileup spectrum $\delta\rho_{\text{pileup}}$ can be built—at the necessary rate—out of these three-window matches.

However, it is important to note that such a correction should only be applied to true triple-pileup events and ideally remain blind to what the reconstruction did. In other words, it ought to be zero when triple pileup did not occur, which is not easily achieved. With three positrons close together, many outcomes can occur yielding two-positron pileup, including, schematically,

$$\rho_{1+2+3}(E,t) \to \rho_{1+3}(E,t) + \rho_2(E,t) \tag{5.13}$$

where the second positron was not close enough to be subsumed. If Eq. 5.13 is substituted in Eq. 5.12, the triple-pileup correction is found to be non-zero. Further, if the cluster partitioning procedure is used in the reconstruction, the spatially middle cluster's energy can be split—in a nonlinear way—between the two clusters on either side of it. Assuming uniform temporal and spatial distributions, such nonlinear cluster splitting is estimated to occur for around 13 % of the three-window pileup candidates.

A potential remedy is to deduce what the reconstruction did event-by-event and include only the correction from identified triple-pileup cases. This has been attempted but found infeasible due to the energy resolution and spurious clusters formed. For example, in a pileup event, the energy deposited from one electromagnetic shower's tail can overlap with that of the other, pushing the pulse amplitude high enough for a low-energy cluster to be reconstructed. Additionally, a low-energy cluster can sometimes be formed from large fit residuals since the pileup trace may differ significantly from the single-positron template used. These types of clusters reconstructed in the pileup island—but not in the single-positron islands—make it difficult to reliably match the input and output clusters in this approach.

Instead, a facet of the reconstruction's behavior is exploited. In Sec. 4.5, if the time difference between two clusters that share at least one crystal is less than the artificial dead time, they are merged into one cluster and refit. As a consequence, if the distributions' indices are time ordered, the process in Eq. 5.13 cannot occur. In other words, when the clusters with the largest time sep-
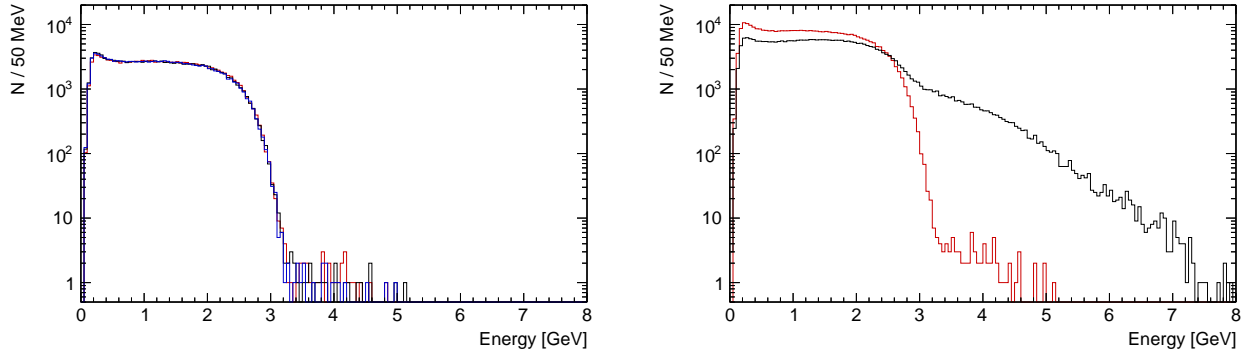
Figure 5.3: Left: The single distribution of $\rho_1(E)$ (red), $\rho_2(E)$ (black), and $\rho_3(E)$ (blue) used for the triple-pileup correction in Eq. 5.14. Right: The distribution of $\rho_{1+2+3}(E)$ (black) used in Eq. 5.14 and, for reference, $\rho_1(E) + \rho_2(E) + \rho_3(E)$ (red). These data are from the 1c data set.

aration are merged together, any clusters in-between them will have been merged too. This allows the triple-pileup correction in Eq. 5.12 to be rewritten as

$$
\begin{aligned}
\delta\rho_{\text{correction}}(E, t) = &- \left[\rho_{1+2+3}(E, t) - \rho_1(E, t) - \rho_2(E, t) - \rho_3(E, t)\right]/2 \\
&+ \left[\rho_{1+2} \quad (E, t) - \rho_1(E, t) - \rho_2(E, t)\right]/2 \\
&+ \left[\rho_{2+3} \quad (E, t) - \rho_2(E, t) - \rho_3(E, t)\right]/2 \\
&+ \left[\rho_{1+3} \quad (E, t) - \rho_1(E, t) - \rho_3(E, t)\right]
\end{aligned}
\tag{5.14}
$$

where the distributions' indices are now time ordered. In essence, Eq. 5.14 overweights the two-positron pileup term between the pileup candidate's first and third cluster—the special case described. As such, it yields a non-zero correction for triple pileup only.

The one exception is if the cluster-partition procedure (Sec. 4.7) is used. This reconstruction step has the advantage of reducing the amount of pileup in the measured data by about a half, leading to a *smaller* pileup correction. However, it breaks the time-order assumption that Eq. 5.13 cannot occur in approximately 10 % of the pileup candidates, leading to a *worse* pileup correction. Together, the cluster-partition procedure has been found to yield no net improvement in ω_a's systematic uncertainty [112]. Therefore, to preclude any (as yet unknown) systematic bias from this
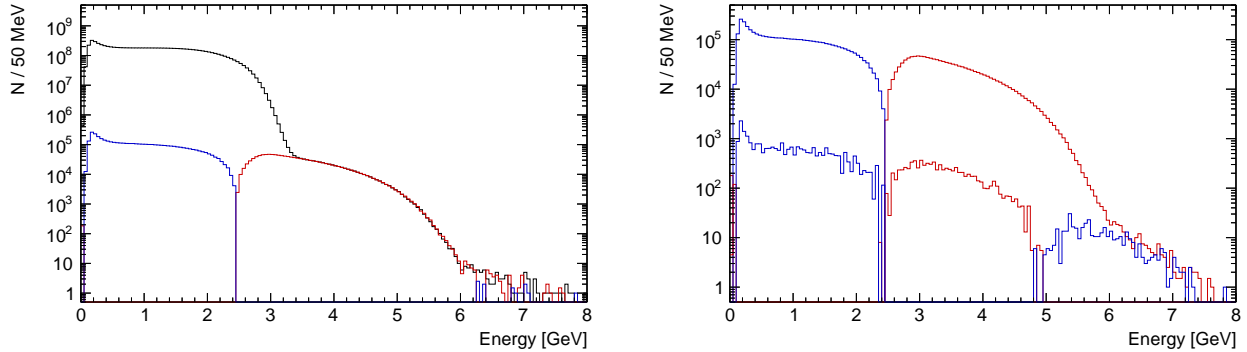
Figure 5.4: Left: The measured energy spectrum (black) and the final pileup spectrum (blue/red). Right: The pileup spectrum in Eq. 5.6 (upper curve) and the triple-pileup correction in Eq. 5.14 (lower curve). Positive is blue; negative is red. These data are from the 1c data set.

step, it is not used in this chapter's ω_a-measurement.

In addition, a conceptual drawback to such a triple-pileup correction is the added assumption that the positron rate is equal at t, $t + \Delta t$, *and* $t + 2\Delta t$ (Sec. 5.4.1). To minimize the resulting pileup-rate error, the reconstructed cluster times are shifted by Δt before populating the distributions in Eq. 5.14. This shift should have a minimal effect on the ω_a-fit, though, because the correction to Eq. 5.6 from triple pileup is estimated to be only 0.5 % [78]. With a rate decrease of another two orders of magnitude, the correction due to quadruple-positron pileup—the next highest order—is then expected to be roughly 0.005 % and, thus, negligible [78]. Notwithstanding, a quadruple-pileup correction (and, in principle, any even higher-order correction) can be found through analogous logic based on the time ordering of the four clusters when combined.

The input distributions to the triple-pileup correction in Eq. 5.14 are seen in Fig. 5.3. As with those for the base pileup correction, this higher-order pileup correction's single distributions (ρ_1, ρ_2, and ρ_3) also closely match each other despite the extra Δt separation in ρ_3. In addition, the triple distribution (ρ_{1+2+3}) has significantly more counts with energies $E \geq 3.1$ GeV than the sum $\rho_1 + \rho_2 + \rho_3$, as expected due to the method's artificially created pileup.

The final pileup spectrum $\delta\rho_{\text{pileup}}(E)$, constructed according to Eqs. 5.6 and 5.14, is depicted
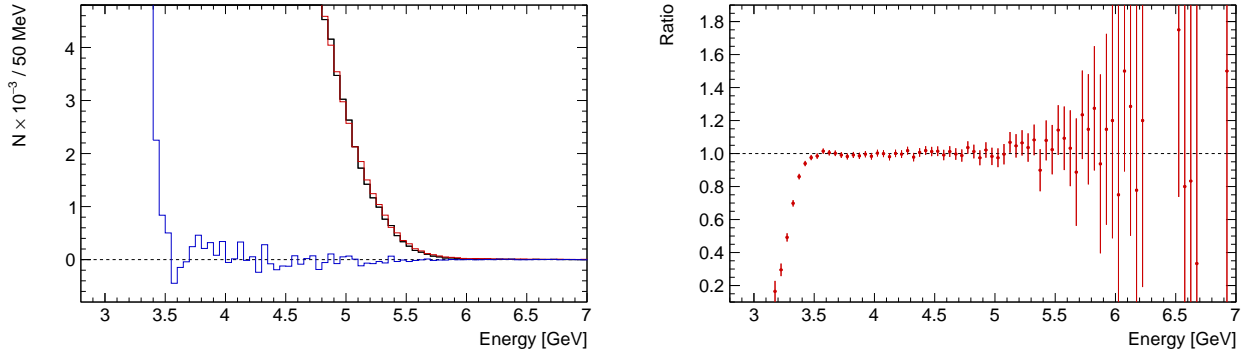
Figure 5.5: Left: The measured (black), pileup (red), and pileup-corrected (blue) spectrum for the 1c data set. Right: The ratio of the pileup to the measured spectrum. A constant fit to the ratio for $E \geq 3.8$ GeV yields 0.05(51)—consistent with a perfect pileup description.

in Fig. 5.4. Two sign transitions can been seen at around 50 and 2450 MeV. The lower-energy transition is from the new clusters identified in certain pileup scenarios discussed in Sec. 5.4.2, while the higher-energy one is from lower-energy counts being effectively moved to higher-energy bins thanks to pileup—resulting in a crossover between removing and adding. Furthermore, in the triple-pileup correction, a third sign transition is seen from an analogous crossover between double and triple pileup events. In total, only about 0.07 % of the measured data are estimated to be pileup. The correction due to triple pileup in Eq. 5.14 is around 0.64 %—close to the 0.5 % back-of-the-envelope prediction mentioned above.

Focusing on the high-energy tail measured in Fig. 5.5—a non-physical regime that must be all pileup counts—we can see that it is fully described by the pileup spectrum. Taking the ratio of $\delta\rho_{\mathrm{pileup}}(E)$ to the measured spectrum for the representative 1c data set, a 0.05 ± 0.51 % under-correction is observed for $E \geq 3.8$ GeV, which is less than 0.1 standard deviations from a perfect correction. There is also no evident deviation from unity at the highest of energies observed, indicating no significant contamination from the uncorrected higher-order pileup. Thus, stopping at triple pileup is reasonable for these data sets' statistics.

Additionally, the pileup correction's time distribution is presented in Fig. 5.6, using the base-
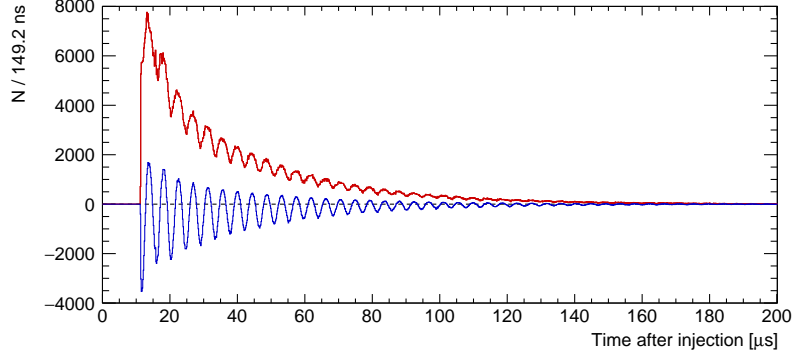
Figure 5.6: The time distribution of the total pileup correction found using the baseline (red) and asymmetry (blue) weighting schemes (Sec. 5.7). These data are from the 1c data set.

line and asymmetry weighting schemes that will be described in Sec. 5.7. The pileup clusters' time spectrum remains visible to about 300 μs after injection. There is also subtle a peak–valley shape asymmetry that arises from the reconstruction's non-zero time resolution; the ω_a-peaks are broadened because the pileup scales approximately quadratically with the underlying positron-hit rate. Moreover, the amplitude of the pileup correction's oscillation at ω_a can be seen to be up to twice as large in the asymmetry-weighted method than in the baseline method.

5.4.3 Covariance Matrix

By construction, the pileup spectrum $\delta\rho_{\mathrm{pileup}}(E, t)$ is determined from and, thus, correlated with the measured data. For an energy-threshold analysis, this correlation must be known in order to apply the correct bin uncertainties in the fit for ω_a. To that end, a bootstrap approach could be used to calculate it empirically from subsets of the data. The low rate of pileup, however, causes large uncertainties, and the positron-rate variations among muon fills limit its accuracy. A more viable approach, therefore, is to calculate the correlations analytically by keeping track of which positrons were used in the pileup construction—as will be described next.

For simplicity, first consider the baseline method in which each positron has unit weight, i.e., $w(E) = 1$. Then let N_m be the number of positrons measured in a given time and energy bin,

which can be partitioned into the number that did not participate in the pileup construction N_b and the number that did participate as follows

$$N_m = N_b + \sum_{n,\Delta n} (n \cdot N_{n,\Delta n}) \tag{5.15}$$

where $N_{n,\Delta n}$ is defined as the number of times n positrons participated in a change of Δn counts. These terms are, by construction, statistically independent as $\Delta n \neq 0$ for a nontrivial pileup correction. In that vein, the bin's pileup counts N_p is simply the net change of counts:

$$N_p = \sum_{n,\Delta n} (\Delta n \cdot N_{n,\Delta n}). \tag{5.16}$$

The bin's corrected number of positrons N_c can, therefore, be written as

$$N_c = N_m - N_p \tag{5.17a}$$

$$= N_b + \sum_{n,\Delta n} (n - \Delta n) \cdot N_{n,\Delta n} \tag{5.17b}$$

and, barring for now the $1/2$ factor in Eq. 5.6, is the number used in the ω_a-fit.

With these equations, the bin uncertainties and their correlations with each other can be exactly found via their covariance. For a given time/energy bin i, its uncertainty is

$$\mathrm{cov}\left[(N_c)_i, (N_c)_i\right] = \mathrm{var}\left[(N_c)_i\right] \tag{5.18a}$$

$$= \mathrm{var}\left[(N_b)_i + \sum (n - \Delta n)_i (N_{n,\Delta n})_i\right] \tag{5.18b}$$

$$= \mathrm{var}\left[(N_b)_i\right] + \mathrm{var}\left[\sum (n - \Delta n)_i (N_{n,\Delta n})_i\right] \tag{5.18c}$$

$$= (N_b)_i + \sum (n - \Delta n)_i^2 (N_{n,\Delta n})_i \tag{5.18d}$$

$$= (N_m)_i + \sum \left[(n - \Delta n)_i^2 - (n)_i\right] (N_{n,\Delta n})_i \tag{5.18e}$$

where the summations are all over $n, \Delta n$. The last three equalities make use of the statistical independence, Poisson distribution, and Eq. 5.17b, respectively. The correlation between time/energy bins i and j can then be expressed by

$$\mathrm{cov}\left[(N_c)_i, (N_c)_j\right] = \mathrm{cov}\left[(N_b)_i + \sum (n - \Delta n)_i (N_{n,\Delta n})_i,\right.$$

$$(N_b)_j + \sum (n - \Delta n)_j (N_{n,\Delta n})_j \Big] \tag{5.19a}$$

$$= \mathrm{cov} \left[\sum (n - \Delta n)_i (N_{n,\Delta n})_i , \sum (n - \Delta n)_j (N_{n,\Delta n})_j \right] \tag{5.19b}$$

$$= \sum \sum (n - \Delta n)_i (n - \Delta n)_j \cdot \mathrm{cov} \left[(N_{n,\Delta n})_i , (N_{n,\Delta n})_j \right] \tag{5.19c}$$

$$= \sum \sum (n - \Delta n)_i (n - \Delta n)_j \cdot (N_{n,\Delta n})_{i,j} \tag{5.19d}$$

where $(N_{n,\Delta n})_{i,j}$ is the number of times a change of Δn_i counts, involving n_i positrons, in bin i caused a change of Δn_j counts, involving n_j positrons, in bin j. Here, the summations are again over $n, \Delta n$. The last equality in Eq. 5.19 is realized by considering two cases. First, if a correction in bin i does not change that in bin j, then the two bins are—by definition—uncorrelated. Second, if a correction in bin i does change that in bin j, then their relative number of occurrences must be proportional to each other. The covariance factor in Eq. 5.19c is then

$$\mathrm{cov} \left[(N_{n,\Delta n})_i , (N_{n,\Delta n})_j \right] = \mathrm{cov} \left[(N_{n,\Delta n})_i , \alpha \cdot (N_{n,\Delta n})_i \right] \tag{5.20a}$$

$$= \alpha \cdot \mathrm{cov} \left[(N_{n,\Delta n})_i , (N_{n,\Delta n})_i \right] \tag{5.20b}$$

$$= \alpha \cdot (N_{n,\Delta n})_i \tag{5.20c}$$

where α is the proportionality constant. More generally, without prior knowledge of α, the covariance factor is $(N_{n,\Delta n})_{i,j}$ between any two time/energy bins i and j.

Moreover, note that the global factor of $1/2$ in Eq. 5.6 effectively scales the number of times n positrons participated in a change of Δn counts, i.e., $N_{n,\Delta n}$. Therefore, any such factor, ξ, can, in general, be accounted for in Eqs. 5.15 and 5.16 via the substitution

$$N_{n,\Delta n} \to \xi \cdot N_{n,\Delta n} \tag{5.21}$$

where nominally $\xi = 1/2$. The generalization is useful in Ch. 6 when artificially scaling the pileup correction to ascertain ω_a's sensitivity to it as part of a systematic assessment.

The covariance matrix built via Eqs. 5.18 and 5.19, however, requires that we know exactly the number of times each pileup situation occurs. In contrast, the pileup-correction procedure in Sec. 5.4.1 is blind to what the reconstruction did to the combined island. Consequently, there is no readily available mapping of clusters. Any attempt at such a mapping would, for instance, have to account for new clusters that are expected, as mentioned in Sec. 5.4.2. A more generic technique, which is equally blind to the reconstruction's behavior, is to build the covariance matrix V per event. That is, for each pileup candidate reconstructed, the quantities n, Δn, and N_c are tallied in each time/energy bin, and a running sum is kept according to

$$V_{i,i} = \sum \left[\xi^2 \left(n - \Delta n \right)_i \left(n - \Delta n \right)_i - \xi \left(n \right)_i \right] + \left(N_m \right)_i, \tag{5.22}$$

$$V_{i,j} = \sum \left[\xi^2 \left(n - \Delta n \right)_i \left(n - \Delta n \right)_j \right] \tag{5.23}$$

where $i \neq j$ and the summations are all over the pileup candidates with $\Delta n \neq 0$. In addition, for the correct correlation between time bins, note that the reconstructed time of the second island—before it was shifted by $-\Delta t$ in the third step in Sec. 5.4.1—is used in the above tally.

When the previous section's triple-pileup correction is applied, the covariance matrix's tallying becomes more intricate. In this case, n is the number of *unique* clusters that participate in a *net* change of Δn, including all corrections for the given time/energy bin. However, in any case, the covariance will be wrong at the next highest order due to pileup candidates, separated by Δt, that share one or more clusters and, thus, are double-counted in the running-sum approach of Eqs. 5.22 and 5.23. For the ω_a-measurement presented in Sec. 5.7, specifically, the triple-pileup correction was ignored in building the covariance matrix. Also, only the diagonal elements, $V_{i,i}$, in Eq. 5.22 were used. These choices are justified in Ch. 6 because their influence on the ω_a-fit were found to be safely negligible, being on the order of 0.001 ppb.

Furthermore, it was demonstrated in Ch. 2 that weighting positrons by their asymmetry, i.e., $w(E) = A(E)$, can yield the highest statistical precision on the best-fit ω_a. Repeating the above
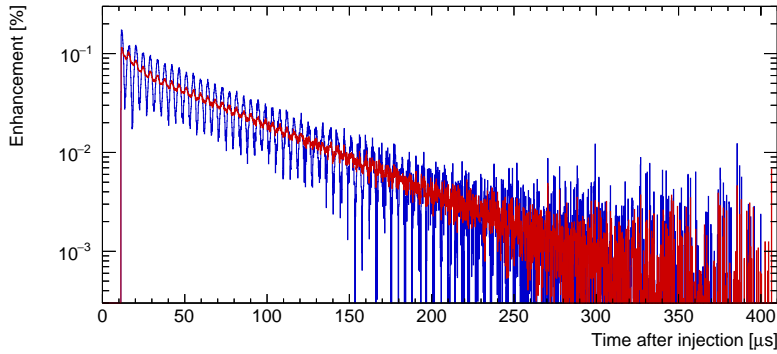
Figure 5.7: The uncertainty enhancement from the pileup subtraction, according to Eq. 5.22, for the baseline (red) and asymmetry (blue) weighting schemes (Sec. 5.7). These data are from the 1c data set, which are representative of the other data sets in Run 1.

logic with each reconstructed positron i weighted, more generally, by w_i, the covariance matrix V in Eqs. 5.22 and 5.23 holds with the replacements

$$N_m \to \sum_i w_i^2, \tag{5.24}$$

$$n \quad \to \sum_i w_i \tag{5.25}$$

where, in a given time/energy bin, the summations are over the positrons measured and over the positrons in the pileup candidate, respectively. Thus, the change in counts in Eqs. 5.22 and 5.23 must similarly be replaced by $\Delta n \to \Delta w$.

It should be noted that, in principle, the weighting scheme used can be a continuous function. In such a case, however, $\Delta w \neq 0$ for nearly all pileup candidates—including those in which the combined island did not result in true pileup event. For example, when two well-separated clusters are combined, the added noise to the waveform alone will slightly perturb its reconstructed energy, leading to a *different* weight. This will, therefore, enhance the correlation between the data and the pileup correction. To mitigate that, the $w(E)$-function can be discretized into a piecewise constant function. As long as the energy-interval width is comparable with the reconstruction's resolution, it should have only a minimal impact on the best-fit ω_a's precision. Such a discretization will be
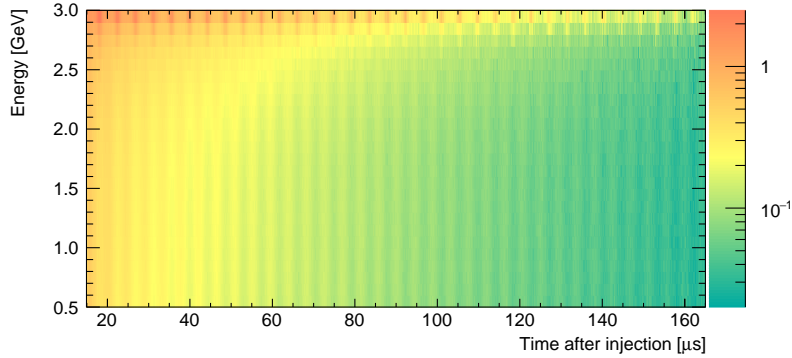
Figure 5.8: The pileup subtraction's uncertainty enhancement, according to Eq. 5.22, as a function of energy and time into the fill when using the baseline (red) and asymmetry (blue) weighting schemes (Sec. 5.7). These data are from the 1c data set.

used in the asymmetry-weighted method in Sec. 5.7.1.

For the baseline and asymmetry-weighted methods, the uncertainty enhancement due to the pileup subtraction, using Eq. 5.22, is shown in Fig. 5.7. This enhancement oscillates at ω_a and is less than 0.1 % after 30 μs into the fill. Furthermore, the uncertainty enhancement is energy-dependent, as illustrated in Fig. 5.8. The relative enhancement is found to grow sharply above about 2.8 GeV to as much as 2.9 %, predominately due to the small number of corrected counts at high energies (Fig. 5.4). All together, the results presented in this section establish the validity of this pileup-correction approach, which will be later used in Sec. 5.7.

5.5 Muon Loss Correction

The minimal, five-parameter model in Eq. 5.3 implicitly assumes that all muons decay inside the storage region. However, as first introduced in Ch. 2, some of the injected muons will inevitably be lost, e.g., due to local field perturbations or betatron motion. While the electrostatic-quadrupole scraping procedure (Sec. 2.4) mitigates these muon losses, a non-negligible number of muons still leave the storage region during the ω_a-measurement period. As a result, there is a deficit in the number of decay positrons in the histogram that is fit.

If the probability for a muon to be lost after scraping were constant in time, the five-parameter model would, nonetheless, describe the histogrammed data well. In this case, the best-fit muon lifetime, τ, would simply be shifted. As will be shown in this section, though, the loss probability varies across the fill, posing a bias to the extracted R. This time dependence, more specifically, will be found by identifying lost muons from coincident clusters in adjacent calorimeters—akin to what was done in Ch. 4. To achieve the experiment's target systematic uncertainty on ω_a, these muon losses will then be incorporated into an extended, six-parameter model—where the new fit parameter accounts for the identification's unknown efficiency.

5.5.1 Lost Muon Identification

A lost muon can be found in the reconstructed data by its characteristic time of flight, $\alpha \approx 6.15$ ns, when passing through adjacent calorimeters. This was the principle that made lost muons an excellent absolute time reference in Sec. 4.4. The simplest approach, therefore, is to identify lost muons as double coincidences—clusters reconstructed in adjacent calorimeters and separated by about α. However, this signal can be easily mimicked, for example, by two independent positrons, which would distort the lost-muon time spectrum. This accidental background is greatly reduced by instead using *triple coincidences*, where a third cluster is found in the consecutive calorimeter and separated from the second by about α. Even higher-order coincidences further mitigate this background, but their signal-to-background ratios are markedly smaller.

Alternatively, the accidental background can be further reduced by requiring that each cluster has an energy below approximately 300 MeV. This is because, as utilized in Sec. 4.6, lost muons deposit a characteristic amount of energy. Some of the time, however, a positron will pile up with a lost muon, raising the reconstructed energy enough for the lost muon to be missed. As such a scenario is more likely to occur early in the fill, it poses a bias to the lost muons' time distribution

and, in turn, to the muon loss correction. Therefore, instead of using an energy cut, it is desirable

to estimate and subtract the residual accidental background from the time distribution of triple

coincidences identified by only using a time cut.

The accidental background, more specifically, can be estimated (on average) by assuming that

the rate of hits, ρ, at times t and $t \pm \Delta t$ are equal—the idea behind the previous section's pileup

correction. As a simpler example, consider the rate of double—instead of triple—coincidences. In

this case, the rate of two clusters in adjacent calorimeters a t and $t + \alpha$ is that of lost muons

and accidentals, while that at times t and $t + \alpha \pm \Delta t$ can only be of accidentals—provided Δt is

more than the time-resolution spread of approximately 1 c.t. (Fig. 4.9). These rates can then be

subtracted to back out the true rate of lost muons.

The same approach can be applied to triple coincidences, where there are many time-window

combinations in which random hits can mimic the signal. Accounting for all of them, the number

of triple coincidences from lost muons, on average, is

$$
\begin{aligned}
L(t) = N_{0,0,0} - & [N_{0,1,1} \quad + N_{0,0,1} \quad + \quad N_{0,1,0} \quad - 2N_{0,-1,1}] \\
- & [N_{0,0,0,0} - N_{0,0,0,1} - \quad N_{0,0,1,0} \quad + 2N_{0,1,-1,0} + 2N_{0,1,-1,1} - \\
& N_{0,0,1,1} - N_{0,1,0,1} - \quad N_{0,1,1,0} \quad + 2N_{0,0,1,-1} + 2N_{0,1,0,-1} - \\
& N_{0,1,0,0} - N_{0,1,1,1} + 2N_{0,1,1,-1} + 2N_{0,-1,1,1} - 6N_{0,1,-1,2}]
\end{aligned}
\tag{5.26}
$$

where the first term in square brackets is the accidental-background correction [120, 98]. The

second term in square brackets will be discussed shortly. Here, each subscript index denotes the

multiplier on Δt with respect to its nominal time in calorimeter order. For example, $N_{0,0,0}$ is the

number of triple coincidences identified without any correction for the accidental background, and

$N_{0,-1,1}$ is the number at t, $t + \alpha - \Delta t$, and $t + 2\alpha + \Delta t$.

To improve the assumption that $\rho(t \pm \Delta t) \approx \rho(t)$, the rate slightly earlier and later in time can

be averaged to give the closer approximation of $[\rho(t + \Delta t) + \rho(t - \Delta t)]/2 \approx \rho(t)$. This is because the change in counts will be linear provided Δt is sufficiently small. In this chapter's analysis, Δt was chosen to be 9 c.t., which is about 7.5 % of a cyclotron period—the relevant time scale based on the discussion in Sec. 5.4.1. The assumptions with this specific Δt-choice have been carefully examined and found to be adequate for $t \geq 30$ μs [98]. At earlier times, they are known to break down due to large oscillations of the positron-hit rate at ω_c from the muons debunching. Accordingly, each term in Eq. 5.26 is implicitly the mean of the term shown and the term under a sign transition, e.g., $N_{0,-1,1}$ stands for $[N_{0,-1,1} + N_{0,1,-1}]/2$.

In addition, to maximize the statistical precision of $L(t)$, every trio of calorimeters should be used. However, lost muons will sometimes pass through more than three consecutive calorimeters, causing them to be over-counted. These cases also pose a bias because the probability for a triple coincidence also being a quadrupole has been found to vary over the course of a fill [98]. To correct for the higher-order coincidences, the above logic can be extended to include coincidence windows in the fourth consecutive calorimeter. The result is the second term in square brackets in Eq. 5.26 after accounting for the effect of accidentals [98].

Furthermore, depending on the timing of each bunch's extraction kicker from FNAL's delivery ring (Sec. 2.2), it is possible for a small amount of particles other than muons to also be injected into the ring and stored. In general, the magnetic storage ring can store any particle with charge e at the magic momentum or with charge $2e$ at twice the magic momentum. Any higher-charged particle should not be stored since it would need a momentum larger than the energy of the initial proton bunches (Sec. 2.2). Consequently, particles aside from muons can be similarly lost during the measurement period, also producing a triple coincidence. However, since they do not change the number of muons still stored, they should not be counted as part of $L(t)$.

As such particle contaminants are also expected to be stable, their relative abundance should
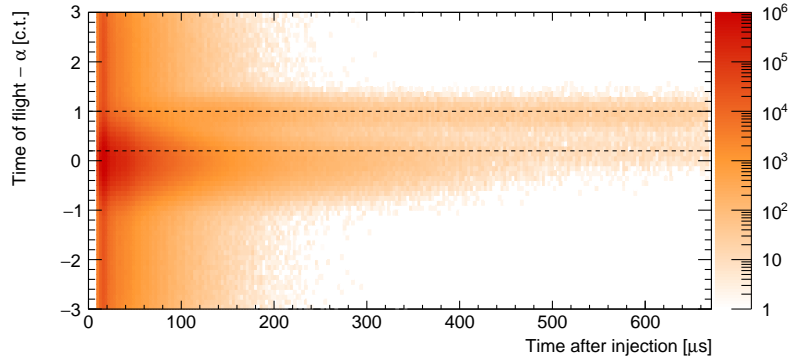
Figure 5.9: The distribution of triple coincidences in the 1d data set as a function of the time separation between the first and second calorimeter, relative to α, and the time into the fill. The populations of lost protons and deuterons are evident at late times with a time-of-flight anomaly around 0.2 and 1 c.t., respectively.

increase over the fill as the muon population decays away. Further, if their mass differs from that of the muon, their characteristic time of flight between adjacent calorimeters will differ as well. They can, therefore, be identified by looking for an unexpectedly large number of coincidences at late times in the fill with a time of flight different than α. As seen in Fig. 5.9, two populations other than muons are indeed evident in the data. They exhibit times of flights of around 0.2 and 1 c.t. beyond α—consistent with them being protons and deuterons. Their relative abundances, on the other hand, are found to vary between the data sets. The 1d data set by far has the largest contamination with roughly 0.3 and 1.6 % of the injected particles estimated to be protons and deuterons, respectively [98].

A straightforward way to correct for these non-muon coincidences in $L(t)$ is to model the time distribution after 500 μs into the fill—when nearly all of the muons have decayed. The model can then be extrapolated to early times and subtracted from the measured $L(t)$. More sophisticated techniques, however, have revealed that the time distribution is complex and cannot be described well by a simple function [98]. For that reason, in this chapter's analysis, the only coincidences counted in Eq. 5.26 were those with a time separation of $\delta t \in (\alpha - 1.2 \text{ c.t.}, \alpha - 0.2 \text{ c.t.})$. As seen in Fig. 5.9, this time-of-flight cut should exclude nearly all of the lost deuterons/protons—as well
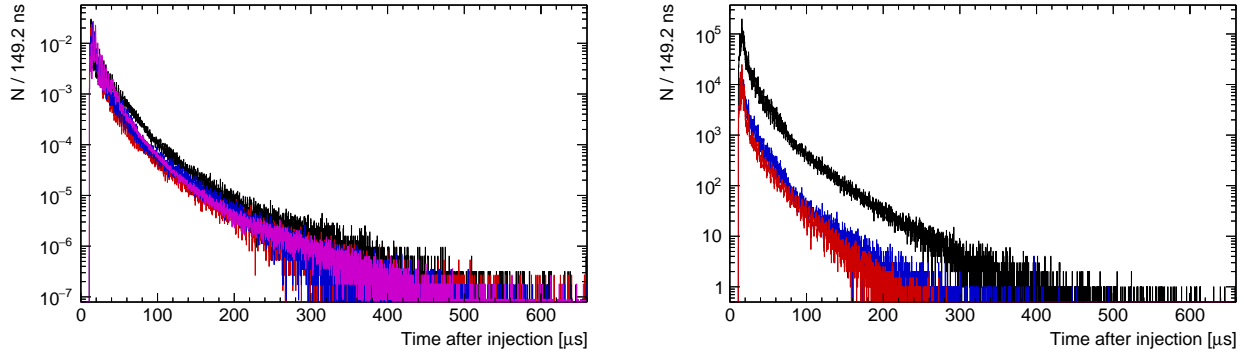
Figure 5.10: Left: The lost muons' time distribution, $L(t)$, found with the 1a (black), 1b (red), 1c (blue), and 1d (magenta) data set. Each distribution's integral is normalized to one. Right: The time distribution for the 1c data set of triple coincidences (black), accidentals (red), and higher-order coincidences (blue). The accidentals and higher-order coincidences compose about a 20 % correction to the number of triple coincidences identified.

as, more generally, any lost particle heavier than the muon. Note that the lower bound of δt was chosen to reduce the accidental background, which is also evident in Fig. 5.9, thereby reducing the size of the accidental correction needed.

The obvious downside to the δt-cut is the loss of statistics used to obtain $L(t)$. It should also be noted that it implicitly assumes that the distribution of the lost muons' times of flight remains constant throughout the fill. Otherwise, an early-to-late effect would be introduced that could bias the best-fit ω_a. Additionally, it is possible for a positron to pile up with a lost muon, pulling the cluster's time enough for the lost muon to be missed. However, the rate of such pileup, as suggested by the last section's results, is very small. All together, the cost of non-muon triple coincidences in $L(t)$ is believed to outweigh these downsides, and, in any case, the influence of these choices on R will be carefully assessed in Ch. 6.

Using the δt-cut, $L(t)$ was constructed for each data set in Run 1 and is shown in Fig. 5.10. It should be noted that the cluster-partition procedure (Sec. 4.7) cannot alter $L(t)$. If an energy cut had been used, however, the procedure's muon–positron pileup reduction would have helped to mitigate the aforementioned bias from the lost muon's energy being raised. As can be seen in

Fig. 5.10, the shape of $L(t)$ varies slightly among the data sets. This is expected since the chance of a muon being lost partly depends on its betatron oscillations' amplitude—and in turn on the quadrupole/kicker settings. It is also evident that each data set's $L(t)$ has a time structure that is not described by the five-parameter model in Eq. 5.3. Therefore, as motivated at the beginning of this section, these lost muons must be accounted for—either by correcting the histogrammed data prior to fitting for ω_a or by extending the model function used to fit.

5.5.2 Muon Loss Correction Approach

As not every lost muon produces a triple coincidence, $L(t)$ is only *proportional* to the rate of lost muons. The proportionality constant, ϵ, characterizes the above method's efficiency of detecting the lost muons. Assuming ϵ remains constant over the fill, the true loss rate is then $L(t)/\epsilon$, and, together with the muons' decay rate, the ordinary differential equation governing the number of muons in the storage region, N, is given by

$$\frac{\mathrm{d}N}{\mathrm{d}t} = -\frac{N}{\tau} - \frac{L(t)}{\epsilon} \tag{5.27}$$

where τ is the muon's time-dilated lifetime. This can be solved via separation of variables to give the number of muons stored,

$$N(t) = N(t_0) \cdot e^{-(t-t_0)/\tau} \cdot \left[1 - \frac{1}{N(t_0)\epsilon} \int_{t_0}^{t} L(t') \cdot e^{(t'-t_0)/\tau}\, \mathrm{d}t' \right] \tag{5.28}$$

where $N(t_0)$ is the number stored at an initial time t_0 [120]. Here, the weighted integral of $L(t)$ accounts probabilistically for when the lost muons would have decayed after $t = t_0$ had they not been lost instead. Therefore, letting $t_0 = 0$ be when the muons were injected, the histogrammed data can be corrected for the muon loss by subtracting the term dependent on $L(t)$ in Eq. 5.28 given a value of $N(0) = N_0$.

However, the efficiency ϵ is unknown *a priori* and, thus, must be found by letting it be a free
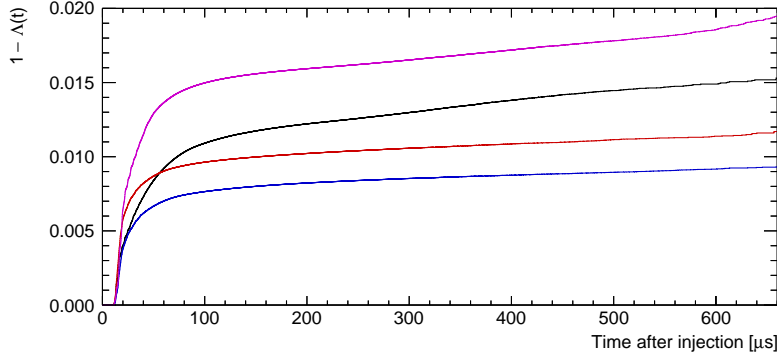
Figure 5.11: The muon-loss correction, expressed as $1 - \Lambda(t)$, for the 1a (black), 1b (red), 1c (blue), and 1d (magenta) data set. Each data set's best-fit κ_{loss} has been used so that each distribution can be interpreted as the loss fraction (since the reconstruction began) over the fill.

parameter in the fit. Adhering to Eq. 5.28, the updated, six-parameter model extends Eq. 5.3 by replacing N_0 with $N_0 \cdot \Lambda(t)$, where the muon-loss correction, $\Lambda(t)$, is

$$\Lambda(t) = 1 - \kappa_{\mathrm{loss}} \int_0^t L(t') \cdot e^{t'/\tau} \, \mathrm{d}t' \tag{5.29}$$

with $\kappa_{\mathrm{loss}} = 1/(N_0\epsilon)$. In practice, κ_{loss} is the new parameter floated, which depends only on the initial number of stored muons and the detection efficiency of lost muons.

To implement Eq. 5.29, $L(t)$ is constructed as a histogram using the same bins as the decay positrons' histogram. For numerical stability, $L(t)$ is also normalized to unity so that the best-fit κ_{loss} is closer to $\mathcal{O}(1)$. The weighted integral is then approximated as a Riemann sum over these time bins with t' being the bins' centers. Moreover, as detailed in Sec. 4.5.9, the reconstruction's fit algorithm only processes islands after about $t_r = 11.25$ μs, which explains why $L(t)$ in Fig. 5.10 is zero until t_r. The only consequence is that N_0 is simply be scaled by $\Lambda(t_r)$, which should be remembered when interpreting the best-fit N_0 results in Sec. 5.7.

The $\Lambda(t)$ distribution for each data set is presented in Figs. 5.11 and 5.12, where the best-fit κ_{loss} values from Sec. 5.7 have been used. If the deuteron/proton coincidences are not removed by the δt-cut, $\Lambda(t)$ can be seen to sharply increase beyond roughly 500 μs. The fact that, using the chosen δt range, the distributions remain linear at such late times supports that the effect from
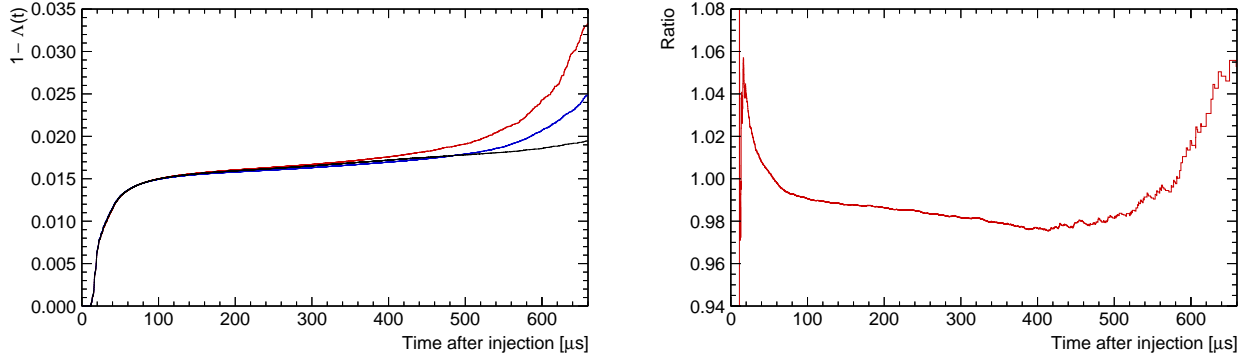
Figure 5.12: The muon-loss correction, expressed as $1 - \Lambda(t)$, for the 1d data set when using an upper bound of δt of $\alpha - 0.2$ c.t. (black), $\alpha + 0.2$ c.t. (red), and $\alpha + 1.2$ c.t. (blue). As the upper bound of δt is increased, more lost protons and deuterons will be included in $L(t)$. Right: Using the 1c data set, the ratio of $\Lambda(t)$ with $L(t)$ in Eq. 5.26 to $\Lambda(t)$ with $L(t) = N_{0,0,0}$.

non-muon populations has been completely removed. Further, the ratio of $\Lambda(t)$ when using the entire expression in Eq. 5.26 and only $N_{0,0,0}$ reveals the non-negligible distortion in $\Lambda(t)$ at early times mostly due to accidental coincidences. This is important because there are over 200 times more positrons detected just between 30 and 40 μs than there are after 500 μs.

5.6 Muon Dynamics

For the baseline method, a histogram is populated with the times of positrons reconstructed from each calorimeter that have energies above 1.7 GeV. After correcting for pileup in Sec. 5.4, ω_a can then be ideally extracted by fitting it to

$$N(t) = N_0 \cdot \Lambda(t) \cdot e^{-t/\tau} \cdot [1 + A_0 \cdot \cos(\omega_a(R) \cdot t + \phi_0)] \tag{5.30}$$

where $\Lambda(t)$ is the muon-loss correction in Eq. 5.29. To reveal if this six-parameter model explains the data adequately, the fast Fourier transform (FFT) of its fit residuals can be used to look for frequencies of any important dynamics that are absent from it. In this section, such an empirical approach will be taken to identify and then account for such phenomena via a mathematical framework to construct the final fit function used in the next section's ω_a-measurement.

(a) Without time randomization.
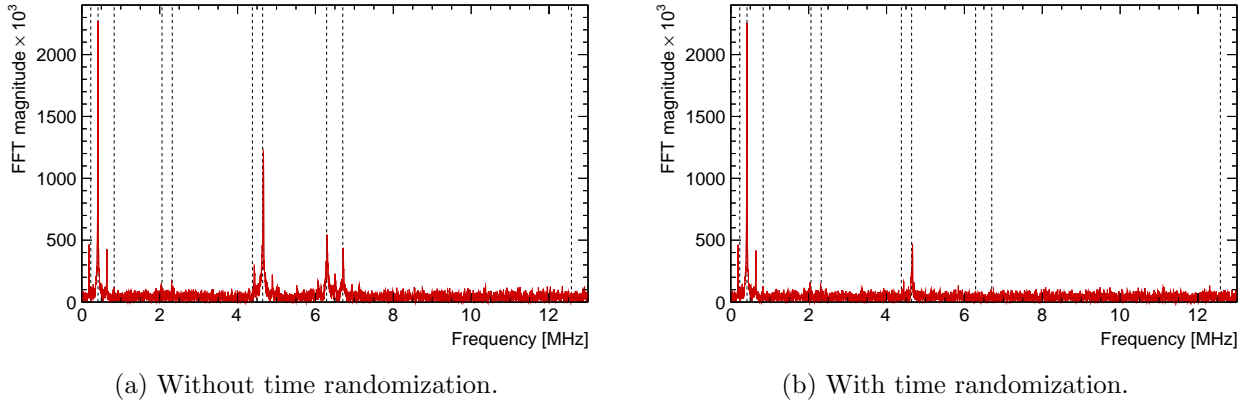
(b) With time randomization.

Figure 5.13: The FFT magnitude of the fit residuals (a) without and (b) with time randomization applied, using the six-parameter model in Eq. 5.30. These data are from the 1c data set. From low to high frequency, the dashed lines are at f_a, $f_c - f_x$, $2(f_c - f_x)$, $f_c - 2f_y$, f_y, $f_c - f_y$, $2f_y$, f_x, f_c, and $2f_x$. The remaining visible frequencies arise from beating with f_a.

As it is—in principle—arbitrary, the histogram is initially binned in 37.3 ns—about a quarter of a cyclotron period. Applying a fit to Eq. 5.30, its fit residuals' FFT can then be observed out to the Nyquist frequency of about 13.4048 MHz. As seen in Fig. 5.13a and as will be described next, there are four fundamental frequencies—f_a, f_c, f_x, and f_y—along with mixed frequencies among them, which are given in Tab. 5.3 for the 1b and 1c data sets' field index of 0.120.

Aside from $f_a = \omega_a/(2\pi)$ and $f_c = \omega_c/(2\pi)$, these frequencies arise from coherent muon oscillations inside the storage region. In particular, the muons are known to undergo simple harmonic motion vertically and radially due to a non-ideal injection and subsequent kick onto the magic radius, respectively (Sec. 2.4). The observed frequencies f_y and f_x—given by Eqs. 2.16 and 2.15, respectively—are their corresponding, characteristic frequencies, which result in modulations in a calorimeter's number of reconstructed positrons.

Moreover, the muons' vertical position is found in the tracker data to be symmetric about the storage region's center, as shown in Fig. 5.14. The y–$\dot{y}$ phase space will, thus, be the same per a 180 ° betatron phase advance, leading to a more dominant frequency of $2f_y$—as observed in the FFT. The symmetry is not perfect, however, causing a FFT peak at f_y to still be present. The

Physical or aliased frequency	Expression	Frequency [MHz]	Period [μs]
Anomalous precession	f_a	0.2291	4.3649
Cyclotron	f_c	6.7024	0.1492
Horizontal betatron	f_x	6.2874	0.1590
Vertical betatron	f_y	2.3218	0.4307
Coherent betatron oscillation	$f_{\mathrm{CBO}} = f_c - 1 \cdot f_x$	0.4150	2.4097
Vertical oscillation	$f_{\mathrm{VO}} = f_c - 1 \cdot f_y$	4.3806	0.2283
Vertical waist	$f_{\mathrm{VW}} = f_c - 2 \cdot f_y$	2.0589	0.4857

Table 5.3: The physical and aliased frequencies that are observable in the calorimeter data. The values quoted are found from the ideal field equations for the 1b and 1c data sets' field index of 0.120, assuming a cyclotron period of 149.2 ns.

radial position about the magic radius, on the other hand, is quite asymmetric, as illustrated in Fig. 5.14. In contrast, this leads to a dominant frequency at f_x—and not $2f_x$.

In addition, each calorimeter effectively samples these frequencies at fixed azimuthal positions as the muons orbit the storage ring at f_c. As a result, some of the beam frequencies will appear lower in the data collected via the stroboscopic effect—regardless of the time-bin width used. In particular, the f_x, f_y, and $2f_y$ frequencies will be aliased since they are higher than the Nyquist frequency $f_c/2$ and are called the coherent betatron oscillation (CBO), vertical oscillation (VO), and vertical waist (VW) frequencies, respectively. As the muons debunch over the fill, however, such sampling at f_c will diminish, raising the relative component of the fundamental frequencies found in the data. In sum, the four physical frequencies, together with their harmonics and aliased frequencies, can account for all of the peaks observed in the fit residuals' FFT.

5.6.1   Time Randomization

For an unbiased ω_a-measurement, the muon dynamics must all be corrected for before the fit or modeled in the fit function. To that end, first consider the dynamics at the cyclotron frequency f_c. As seen in Fig. 4.8, the muons are injected into the ring with a temporal profile lasting less than a
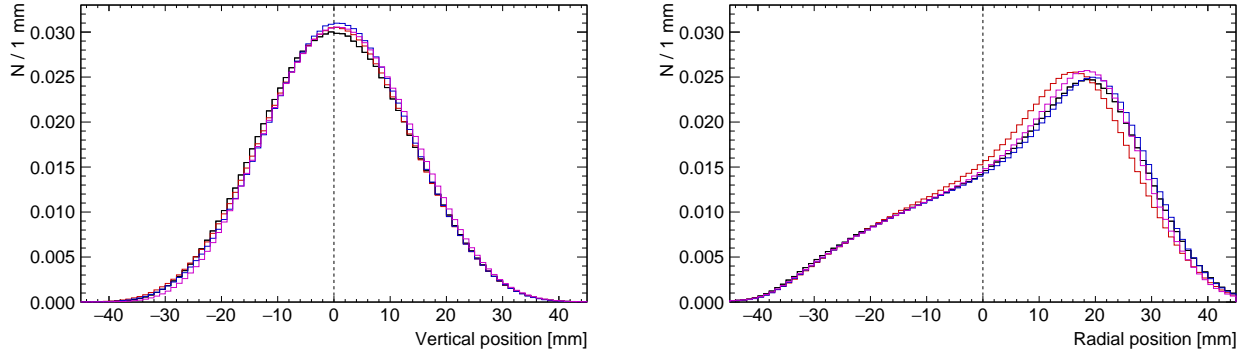
Figure 5.14: Preliminary results of the vertical (left) and radial (right) distribution of muons, reconstructed from tracker data after 30 μs into the fill [122].

cyclotron period, $T_c = 1/f_c \approx 149.2$ ns. The muons, therefore, start out traveling around the ring spatially localized. From the calorimeter's point of view, the positron-hit rate will oscillate at f_c with a peak when more muons are on one side and a trough when more muons are on the other. As time elapses, the muons' momentum spread will eventually cause the stored muons to spatially smear out uniformly in azimuth. However, as illustrated in Fig. 5.15, at 30 μs after injection when the histogram's fit typically starts, this *fast rotation* signal at f_c is still evident.

In addition to the muons' debunching, their spread in momenta leads to that in the cyclotron periods realized, which is presented in Fig. 5.16. Together, the data's fast-rotation signal would be difficult to model sufficiently well. It is, therefore, instead averaged out, as seen in Fig. 5.15, by binning in 149.2 ns—the typical value of T_c. This specific bin width was chosen because it is the most probable value in Fig. 5.16 to the appropriate number of significant digits. As a result, any frequency above the Nyquist frequency, $f_c/2 \approx 3.4$ MHz, will also be aliased. In particular, the frequencies f_{VO}, $2f_y$, and f_x become aliased to f_y, f_{VW}, and f_{CBO}, respectively—frequencies already present in the fit residuals' FFT.

The spread in cyclotron periods, however, means that such binning cannot *exactly* average out the fast-rotation signal. As each calorimeter effectively samples a different phase of the cyclotron motion, any residual fast-rotation signal should then manifest in a calorimeter-dependent bias to
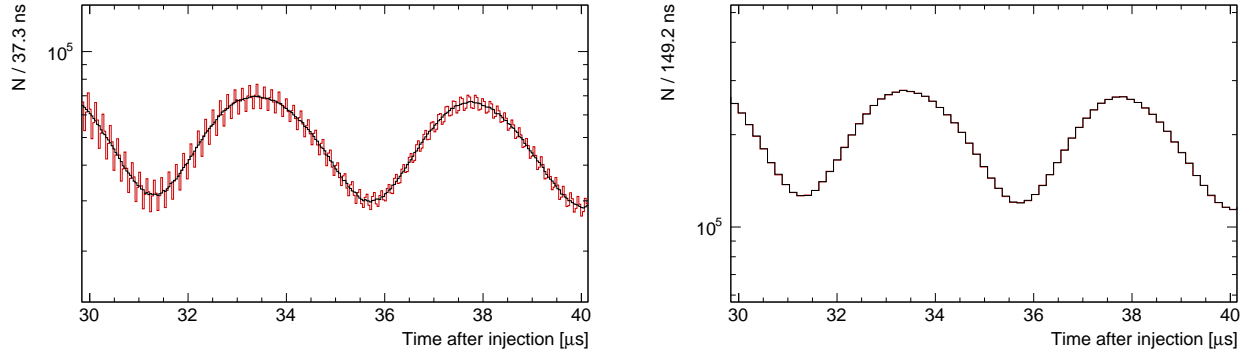
Figure 5.15: The time histogram of positrons in the 1c data set, using $E_{\text{th}} \geq 1.7$ GeV. When the time-bin width is a quarter of a cyclotron period (left), the fast-rotation signal at f_c can be seen without (red) but not with (black) the time-randomization procedure applied. In contrast, when the time-bin width is equal to a cyclotron period (right), the fast-rotation signal is not directly evident even without the time-randomization procedure applied.

ω_a. When the calorimeter-combined data are fit, these biases are not expected to average to zero because of the calorimeters' different acceptances.

To remove this potential ω_a-bias, the fast-rotation signal is further smeared out by artificially randomizing its phase among the muon fills. Specifically, each fill's injection time is offset by a random number drawn from a uniform distribution between $-149.2/2$ and $149.2/2$ ns, using the `Ranlux64Engine` generator with a luxury level of two. In contrast to randomizing each cluster's time, this procedure preserves the cancellation of any muon-dynamics signal between calorimeters 180 $°$ apart. Nonetheless, such alternative methods to time randomization have been attempted but will be shown in Sec. 6.7 to be no better than the one described in this section.

As seen in Fig. 5.17, this time-randomization procedure successfully removes the fit residuals' FFT peak at f_c—even without binning in T_c. The removed calorimeter-dependence of R can be seen too. As a secondary consequence, the other frequencies in the data are also partially smeared out—and to a larger extent the closer they are to f_c. For example, as shown in Fig. 5.13, the $2f_y$ peak has been greatly reduced, and the f_x peak has been effectively removed. This should make the muon-dynamics effects somewhat less important in the fit but, at the same time, harder to fit
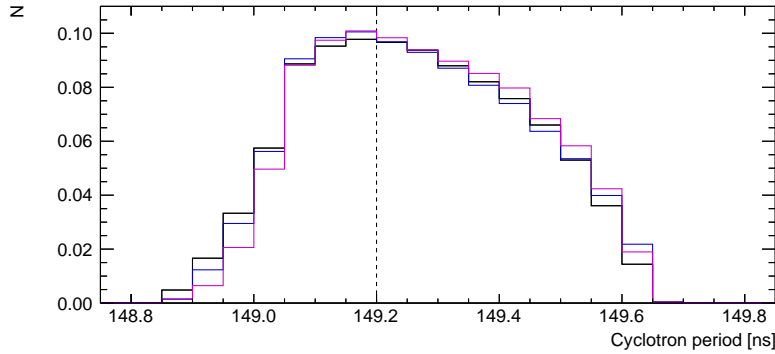
Figure 5.16: Preliminary results of the muons' distribution of cyclotron periods for the 1a (black), 1c (red), and 1d (magenta) data sets [123]. The result for the 1b data set is not available as of the time of this writing (Dec. 2019) but is expected to be similar in shape to the other data sets.

since their signals are now diminished.

5.6.2 Muon Dynamics Model

This time-randomization approach applied to f_c breaks down for the lower frequencies closer to f_a—the primary frequency of interest. The period of the largest remaining FFT peak at f_{CBO}, for example, is about 55 % of that of f_a. A time randomization to smear out f_{CBO} would, thus, also smear out an appreciable fraction of the anomalous precession's signal. The residual frequencies due to muon dynamics should, therefore, be instead modeled in the fit function. In the rest of this section, the form of the model to be used will be motivated by accounting for the observed frequencies within a mathematical framework [124].

To begin, the muons' motion is mainly expected to modulate the normalization, N_0, by varying in time if a given decay positron, on average, will be detected. Following the notation defined in Ch. 2, this acceptance can be parameterized by the muon's position in the storage ring when it decayed, (x, y, s), and momentum, $(\dot{x}, \dot{y}, \dot{s})$. Here, s is the muon's azimuthal position. This can be formalized into a joint probability distribution, $\mathcal{N}(x, \dot{x}, y, \dot{y})$, where its s- and $\dot{s}$-dependencies are averaged over a given calorimeter's phase space. Further, to leading order, x and $\dot{x}$ (and y and
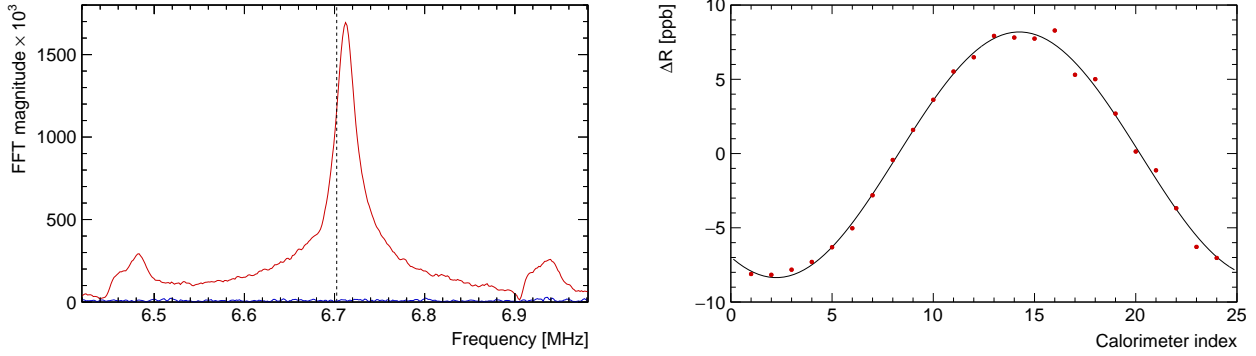
Figure 5.17: Left: The fit residuals' FFT around the cyclotron frequency, using the model in Eq. 5.30, for calorimeter 1 in the 1c data set without (red) and with (blue) the time randomization applied. Right: The difference in the best-fit R across the calorimeters in the 1c data set without and with the time randomization applied, using the fit procedure described in Sec. 5.7. A fit to sine wave with a fixed period of 24 is overlaid (black). The deviation from zero shows the ω_a-bias introduced by the (unaccounted for) fast-rotation signal.

$\dot{y}$) are coupled by simple harmonic motion as their phase space rotates with the muons traveling around the ring (Sec. 2.4). That is to say that the probability that a positron hits a calorimeter, on average, simply depends on where the muon was when it decayed. Also to leading order, x and y are expected to be independent of each other. Together, the acceptance's distribution can then be redefined as the product of two probability distributions, $\mathcal{N}(x) \cdot \mathcal{N}(y)$, allowing the x- and y-dependencies to be considered separately.

The normalization's time dependence arises from the muons' time-varying spatial distribution in the storage region. Moreover, the shape of the muons' distribution, $\mathcal{B}$, can be described by its moments. Considering only the x-dependency, the i-th raw moment is given by

$$\mu_i(t) = \int_{-\infty}^{\infty} x^i \cdot \mathcal{B}(x \mid t)\, \mathrm{d}x \tag{5.31}$$

where $\mathcal{B}(x \mid t)$ is the marginal probability distribution at a given time t. By definition, the zeroth moment is unity. The first raw moment is the mean, $\mu(t)$, while the second can be shown to be $\mu^2 + \sigma^2$, where $\sigma(t)$ is the standard deviation.

At a time t into the fill, the probability distribution in x of decay positrons that go on to hit a

calorimeter can then be formalized as the product, $\mathcal{N}(x) \cdot \mathcal{B}(x \mid t)$. An analogous expression holds for y since x and y have been taken as independent of each other. By integrating over all of the possible positions, the time-dependent modulation on N can be written as

$$N(t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [\mathcal{N}(x) \cdot \mathcal{B}(x \mid t)] \cdot [\mathcal{N}(y) \cdot \mathcal{B}(y \mid t)] \, \mathrm{d}x \, \mathrm{d}y \tag{5.32a}$$

$$= N_x(t) \cdot N_y(t) \tag{5.32b}$$

where $N_x(t)$—as with $N_y(t)$—is given by

$$N_x(t) = \int_{-\infty}^{\infty} \mathcal{N}(x) \cdot \mathcal{B}(x \mid t) \, \mathrm{d}x \tag{5.33a}$$

$$= \int_{-\infty}^{\infty} \left[\sum_{i=0}^{\infty} a_i \cdot x^i \right] \cdot \mathcal{B}(x \mid t) \, \mathrm{d}x \tag{5.33b}$$

$$= \sum_{i=0}^{\infty} a_i \cdot \left[\int_{-\infty}^{\infty} x^i \cdot \mathcal{B}(x \mid t) \, \mathrm{d}x \right] \tag{5.33c}$$

$$= \sum_{i=0}^{\infty} a_i \cdot \mu_i(t) \tag{5.33d}$$

$$= a_0 + \sum_{i=1}^{\infty} a_i \cdot \mu_i(t) \tag{5.33e}$$

and $\mathcal{N}(x)$ is expanded as a Maclaurin series in Eq. 5.33b—which was verified to be reasonable in a Monte Carlo simulation [124]. As a result, the normalization's modulation from muon dynamics can be expressed as a series of the raw moments. This provides a framework upon which a model for $N_x(t)$ and $N_y(t)$ can be built—by postulating the form of each moment, $\mu_i(t)$.

From the physics of the muons' dynamics inside the ring, the non-ideal injection/kick onto the magic orbit is expected to cause the muons to oscillate radially and vertically with characteristic frequencies, f_x and f_y, respectively. As the muons repeatedly circle the ring, though, they will spread out to cover the ring's admittance due to their momentum spread and quadrupole-field non-linearities. This causes the phases of the individual muons' oscillations to increasingly interfere destructively over the fill. After enough time has passed, the muon ensemble will eventually reach an equilibrium state—collectively with no time dependence.

Based on that intuition, each $i \geq 1$ moment is postulated to have the form of an equilibrium

constant plus a periodic term that diminishes over time, i.e.,

$$\mu_i(t) = E_i + D_i(t) \left[\sum_{j=1}^{\infty} A_{i,j} \cdot \cos(j\omega \cdot t + \phi_{i,j}) \right] \tag{5.34}$$

where $D_i(t)$ is known as the "decoherence" envelope. Without any loss of generality, the periodic

modulation is treated as a Fourier series. The $j = 0$ constant term, however, is set to zero as it

corresponds to a displacement that is not expected.

The exact envelope function and the number of Fourier terms can be empirically determined

from the data, and, with Eq. 5.33e, the functional form of $N_x(t)$ and $N_y(t)$ can be constructed.

Together, the modulation of the normalization, N_0, from muon dynamics can then be modeled by

$N_x(t) \cdot N_y(t)$. Further, the modulation on the asymmetry, A_0, and phase, ϕ_0—which result from

oscillations of the positron energy and drift time, respectively—can be analogously derived using

the substitutions $N \to A$ and $N \to \phi$, respectively, in the equations.

5.6.3 Fit Function

The previous section's framework can now be employed to build the final fit function for the ω_a-

measurement. In general, the guiding principle is to include just enough moments and modulation

terms to account for each evident frequency in the fit residuals' FFT. The terms absent are consid-

ered to be negligible for the size of the data sets in Run 1 and, thus, not large enough to bias the

best-fit ω_a-value. The functional form of the decoherence envelopes will then be estimated from

a combination of muon-dynamics simulations and tracker measurements.

First, consider the f_x-related frequencies in Fig. 5.13 after binning in 149.2 ns—i.e., the two

aliased frequencies, f_{CBO} and $2f_{\text{CBO}}$. From the trackers' measurement of the muons' spatial dis-

tribution over the muon fill, the f_{CBO} peak is understood to be predominantly from the muons'

centroid oscillating due to the non-ideal kick and, therefore, is associated with the first moment
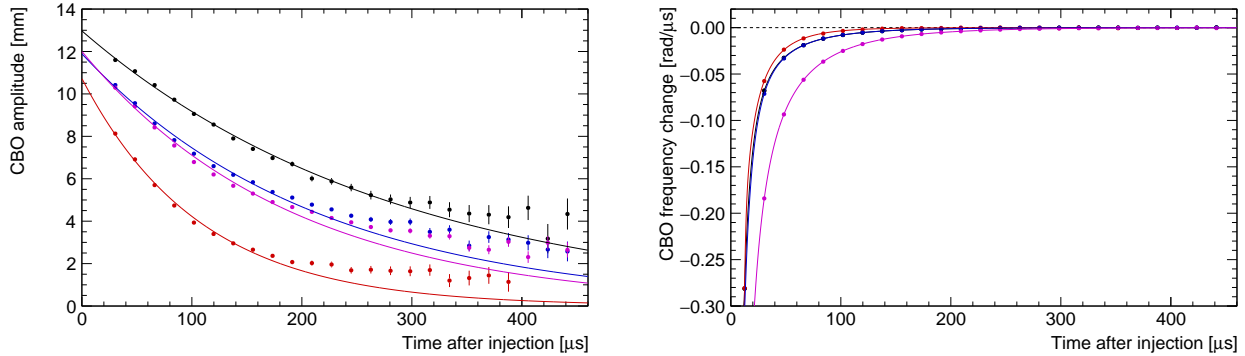
Figure 5.18: The amplitude (left) and frequency change (right) of the f_{CBO}-oscillation for the 1a (black), 1b (red), 1c (blue), and 1d (magenta) data set. These data are based on the positrons' mean horizontal position measured by the tracker in front of calorimeter 18 [125]. Each data set's measured amplitudes have been fit to an exponential model, while their frequency changes have been fit according to the *ad hoc* model in Eq. 5.41 using the values in Tab. 5.5.

[125]. Accordingly, it can be modeled by a $j = 1$ term in $\mu_1(t)$, where ω_x is aliased to ω_{CBO}. Its decoherence envelope measured by the trackers, as seen in Fig. 5.18, roughly follows a decaying exponential, which is consistent with muon-dynamics simulations.

At this point, it is important to distinguish between the frequency components aliased by the muons' cyclotron motion or by the time-bin width. In general, the former can have a different, characteristic decoherence time as its decoherence incorporates that of the fundamental frequency *and* f_c while the muons debunch. When faster muons begin to lap slower ones, for instance, the phases of the individual muons' oscillations will once again partially interfere constructively. A different decay, therefore, of the f_{CBO} and f_{VW} modulations from the muon's cyclotron motion is expected—as compared to those from the time bin's width. This must be taken into account when estimating the form of each moment's decoherence envelope, $D_i(t)$.

To study their possible differences, the decoherence time of each frequency in Fig. 5.13a was separately estimated. To do so, the six-parameter model in Eq. 5.30 was extended by the minimal form for a moment in Eq. 5.34. Assuming an exponential-decay model, the finely binned,

Physical or aliased frequency	Expression	Decoherence time constant, τ_i [μs]
Cyclotron	f_c	25.4 ± 3.1
Horizontal betatron	f_x	12.2 ± 2.7
Double vertical betatron	$2 \cdot f_y$	37.6 ± 2.1
Coherent betatron oscillation	$f_{\mathrm{CBO}} = f_c - 1 \cdot f_x$	215 ± 16
Vertical waist	$f_{\mathrm{VW}} = f_c - 2 \cdot f_y$	16 ± 15

Table 5.4: The time constant, τ_i, of the exponential-decay decoherence envelope in Eq. 5.35 for the dominant frequencies in the 1c data set. The best-fit values for f_y, f_{VO}, and $2f_{\mathrm{CBO}}$ are less than one standard deviation away from zero and, therefore, have been omitted.

baseline-weighted histogram (without time randomization) was then fit to

$$N(t) = N_0 \cdot \Lambda(t) \cdot e^{-t/\tau} \cdot [1 + A_0 \cdot \cos(\omega_a(R) \cdot t + \phi_0)]$$

$$\cdot \{1 + e^{-t/\tau_i} \cdot [A_i \cdot \cos(\omega_i \cdot t + \phi_i)]\} \tag{5.35}$$

where ω_i is the specific frequency peak of interest. This allows one frequency in addition to ω_a to be probed. For each frequency fit, its residuals' FFT exhibited little to no peak at the frequency of interest, verifying the fits were successful. As shown in Tab. 5.4, the decoherence time constant, τ_i, found at f_{CBO} is indeed about an order of magnitude larger than that at f_x or f_c. The f_{CBO} component from the muons' cyclotron motion and from the time-bin width will, therefore, have different decoherence envelopes. However, after time randomization as seen in Fig. 5.13b, the f_x component aliased (by binning) into f_{CBO} is expected to be negligible compared to the size of the peak at f_{CBO} from the muon's cyclotron motion. Consequently, each observed moment can be modeled with one exponential describing the f_{CBO} (or $2f_{\mathrm{CBO}}$) frequency.

Moreover, the $2f_{\mathrm{CBO}}$ peak observed could reasonably be from a $j = 2$ term in either $\mu_1(t)$ or $\mu_2(t)$. As measured by the trackers, the radial position's mean and width both exhibit a $2f_{\mathrm{CBO}}$-frequency component [126, 127]. However, that corresponding to the width, i.e., the second central moment, is significantly larger. While the second raw moment differs from the central moment by μ^2, to leading order, it can be modeled by a $j = 2$ term only in $\mu_2(t)$. This suggests, though,

that there should be a higher-order, $j = 1$ term in $\mu_2(t)$—at ω_{CBO}—which is indeed evident in the trackers' measured frequency spectrum of the radial beam width [127]. Further, as the second moment is formed from the square of x, it may be expected for $D_2(t)$ to likewise be the square of that for $D_1(t)$. As it is consistent with tracker data and the $2f_{\mathrm{CBO}}$ frequency is predominately muon-aliased, the decoherence envelope will accordingly be fixed to the same exponential decay function but with half the time constant, τ_{CBO} [127].

Next, consider the f_y-related frequencies in Fig. 5.13 with 149.2-ns binning—i.e., the physical frequency, f_y, and the aliased frequency, f_{VW}. In a similar way to that in x, the f_y frequency is understood to largely come from the mean oscillating, while the $2f_y$ (aliased to f_{VW}) frequency is predominately from the width oscillating. They can, thus, be modeled with the first and second moments as in $N_x(t)$. Further, tracker measurements of the decoherence envelope suggest that a decaying exponential will similarly be an adequate model [128]. As seen in Tab. 5.4, the decoherence's time constant, τ_i, found at f_{VW} is also significantly different from that at $2f_y$ or f_c. From Fig. 5.13b, though, the $2f_y$ component aliased (by binning) into f_{VW} will dominate in the fit over the f_{VW} component due to the muon's cyclotron motion. Also, the f_{VO} component due to the muons' cyclotron motion will be aliased by binning into f_y, but its peak in the FFT is almost undetectable above the noise floor. Each observed beam moment, therefore, can be modeled with one exponential describing the f_y and f_{VW} frequencies not aliased by the muons. For the same reason as in x, $D_2(t)$ is then expected to have half the time constant in $D_1(t)$. In fact, if the two moments' time constants are allowed to float in the 1c data set's fit, the ratio of their best-fit values is found to be around 0.51—consistent with such intuition.

When fitting with the $N_x(t) \cdot N_y(t)$ factor, a few small—but evident—frequency peaks remain in the fit residuals' FFT. They vanish, however, when accounting for a modulation in the asymmetry, A_0, and phase, ϕ_0. To do that, a minimal model was used, i.e., only the $i = 1$ and $j = 1$ term,
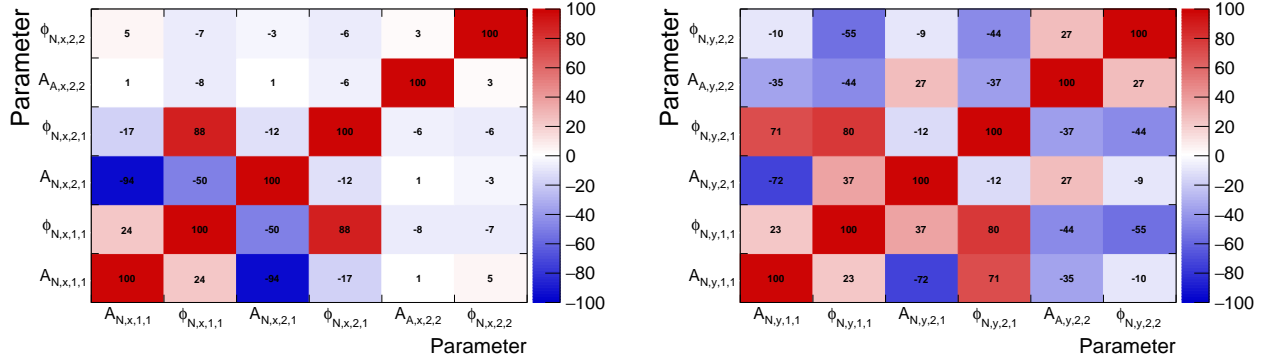
Figure 5.19: The correlation matrix in percent between the amplitudes and phases in $N_x(t)$ (left) and $N_y(t)$ (right) when the $i = 2$ and $j = 1$ terms (denoted by the "2, 1") are included in a fit to the 1c data set, using the fit procedure described in Sec. 5.7.

for these two fit parameters. All together, the final fit function becomes

$$N(t) = N_0 \cdot N_x(t) \cdot N_y(t) \cdot \Lambda(t) \cdot e^{-t/\tau} \left[1 + A_0 \cdot A_x(t) \cdot \cos(\omega_a(R) \cdot t + \phi_0 \cdot \phi_x(t)) \right] \quad (5.36)$$

where the modulations from muon dynamics are modeled via Sec. 5.6.2 by

$$N_x(t) = 1 + e^{-1t/\tau_{\mathrm{CBO}}} \cdot A_{N,x,1,1} \cdot \cos(1\omega_{\mathrm{CBO}}(t) \cdot t + \phi_{N,x,1,1})$$

$$+ e^{-2t/\tau_{\mathrm{CBO}}} \cdot A_{N,x,2,2} \cdot \cos(2\omega_{\mathrm{CBO}}(t) \cdot t + \phi_{N,x,2,2}), \quad (5.37)$$

$$N_y(t) = 1 + e^{-1t/\tau_y} \quad \cdot A_{N,y,1,1} \cdot \cos(1\omega_y \quad (t) \cdot t + \phi_{N,y,1,1})$$

$$+ e^{-2t/\tau_y} \quad \cdot A_{N,y,2,2} \cdot \cos(1\omega_{\mathrm{VW}}(t) \cdot t + \phi_{N,y,2,2}), \quad (5.38)$$

$$A_x(t) = 1 + e^{-1t/\tau_{\mathrm{CBO}}} \cdot A_{A,x,1,1} \cdot \cos(1\omega_{\mathrm{CBO}}(t) \cdot t + \phi_{A,x,1,1}), \quad (5.39)$$

$$\phi_x(t) = 1 + e^{-1t/\tau_{\mathrm{CBO}}} \cdot A_{\phi,x,1,1} \cdot \cos(1\omega_{\mathrm{CBO}}(t) \cdot t + \phi_{\phi,x,1,1}) \quad (5.40)$$

and where the equilibrium constants, E_i, for $N(t)$ and $A(t)$ have been factored out and combined into overall factors of N_0 and A_0, respectively. Note that, in both $N_x(t)$ and $N_y(t)$, the expected $i = 2$ and $j = 1$ term has been omitted. This is because, if included, they are found to be highly correlated with other, more dominant fit parameters, as seen in Fig. 5.19. The differences among the frequencies and envelopes in $N_x(t)$ and $N_y(t)$ are not great enough to constrain all of the fit parameters with the data sets' sizes in Run 1. The primary consequence is that the decoherence

envelope is altered for the remaining terms at ω_{CBO} and ω_y. With the exponential modeling for $D_i(t)$, this means that the best-fit τ_{CBO} and τ_y values will be lower than what would have been expected including the $i = 2$ and $j = 1$ terms in $N_x(t)$ and $N_y(t)$. How such lower values influence the best-fit R will be assessed as a systematic uncertainty in Sec. 6.7.

In Eqs. 5.37–5.40, the muon-dynamics frequencies are expressed as functions of time. This is because the recovery of the quadrupole plates from their scraping to nominal voltages was much longer than expected in Run 1—which was caused by damaged/defective resistors in the system. This in turn caused ω_{CBO} and ω_y to vary throughout the muon fill (Sec. 2.4). The exact time dependence of ω_{CBO} was extracted from the tracker data using the *ad hoc* model

$$\omega_{\mathrm{CBO}}(t) = \omega_0 \left[1 + A_1 e^{-t/\tau_1}/t + A_2 e^{-t/\tau_2}/t \right] \tag{5.41}$$

which can be seen in Fig. 5.18 to be a reasonable functional form. Here, ω_0 is a free parameter in the fit for ω_a, while the other parameters—given in Tab. 5.5—are fixed from the tracker data's fit for each data set. The ω_a-uncertainty from fixing these values is assessed in Sec. 6.4.

The time dependence of ω_y and ω_{VW}, on the other hand, cannot be precisely measured by the trackers because, in comparison, these oscillations decohere too quickly. However, for the same reason, these frequencies are also less important in the ω_a-fit. Nonetheless, the time dependence expected can be estimated because ω_y is related to ω_{CBO} by

$$\omega_y(t) = \kappa_y \cdot \omega_{\mathrm{CBO}}(t) \sqrt{\frac{4\pi/(149.2 \text{ ns})}{\kappa_y \cdot \omega_{\mathrm{CBO}}(t)} - 1} \tag{5.42}$$

where $\kappa_y = 1$, using Eqs. 2.15 and 2.16. This is based on the field-index equations assuming the quadrupole plates were continuous around the storage ring which are known to be inaccurate up to about 1 % [122]. A fudge factor, κ_y, is, therefore, introduced in Eq. 5.42 to account for such inaccuracy. For the 1a data set, this model for $\omega_y(t)$ using $\kappa_y = 1.013$ has been found to agree at the level of the tracker measurement's precision [128]. Thus, based on $\omega_y(t)$ in Eq. 5.42, $\omega_{\mathrm{VW}}(t)$ can

Data set	A_1 [μs]	τ_1 [μs]	A_2 [μs]	τ_2 [μs]
1a	1.223 ± 0.017	81.45 ± 0.35	2.1806 ± 0.0086	7.95 ± 0.25
1b	1.2433 ± 0.0077	49.5 ± 3.3	2.596 ± 0.067	6.70 ± 0.10
1c	1.1012 ± 0.0057	76.0 ± 3.2	2.095 ± 0.011	8.85 ± 0.35
1d	3.1803 ± 0.0021	95.150 ± 0.050	2.055 ± 0.041	9.10 ± 0.10

Table 5.5: The best-fit parameters of the $\omega_{\mathrm{CBO}}(t)$-model in Eq. 5.41 for each data set, found by averaging the values from fits to the two straw trackers' data.

then be estimated by $2\pi/(149.2 \text{ ns}) - 2\omega_y(t)$.

5.7 Fit Procedure and Results

The aforementioned pileup and muon-loss corrections together with the muon-dynamics model enable the anomalous precession frequency measurement, which was outlined in Sec. 2.5. Beginning with the baseline method—i.e., with a weighting scheme of $w(E) = 1$—a time histogram was constructed for each data set in Run 1 from 0 to 699.8972 μs, using a bin width of 149.2 ns. After randomizing the muon fills' injection times, each histogram was populated with decay positrons above a calibrated energy of 1.7 GeV. The pileup-corrected histograms—one of which is given in Fig. 5.20a—were then fit to the 22-parameter model in Eq. 5.36.

In addition to the fundamental five parameters (N_0, τ, A_0, R, and ϕ_0), the free parameters in the fit include the lost-muon detection efficiency (κ_{loss}), the muon-dynamics frequencies (ω_0 and κ_y), the decoherence time constants (τ_{CBO} and τ_y), the amplitude and phase of the normalization's modulation from muon dynamics ($A_{N,x,i,i}$, $\phi_{N,x,i,i}$, $A_{N,y,i,i}$, and $\phi_{N,y,i,i}$ for $i \in \{1, 2\}$), the amplitude and phase of the asymmetry's modulation from muon dynamics ($A_{A,x,1,1}$ and $\phi_{A,x,1,1}$), and, lastly, the amplitude and phase of the phase's modulation from muon dynamics ($A_{\phi,x,1,1}$ and $\phi_{\phi,x,1,1}$). Furthermore, the same software blinding offset, ΔR, was used in each data set's fit so their best-fit values of R can be directly compared.

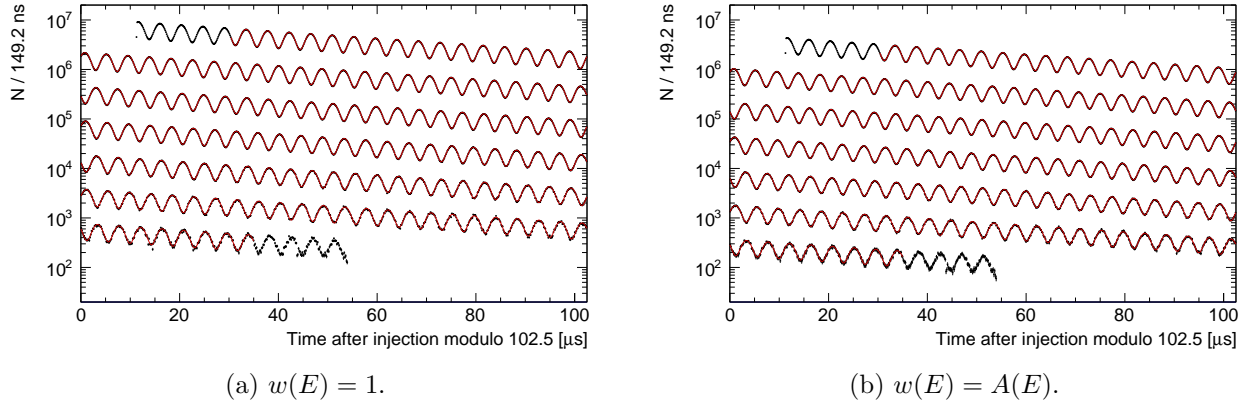(a) $w(E) = 1$.   (b) $w(E) = A(E)$.

Figure 5.20: The time histogram of reconstructed positrons in the 1c data set after correcting for pileup and randomizing the muon fills' injection times using the (a) baseline and (b) asymmetry weighting schemes. The fit to the 22-parameter model (Eq. 5.36) is overlaid (red).

The best fit parameters, $\boldsymbol{p}$, are found by minimizing the Neyman's χ^2 formulation—where the variance is taken as the observed number of counts—as formalized by

$$\chi^2 = \sum_b \frac{1}{N_b} \left[N_b - N(t_b \mid \boldsymbol{p}) \right]^2 \tag{5.43}$$

where the N_b is the number of counts in the time bin b. Note that the model is evaluated at each time bin's center, t_b. To obtain the fit parameters' initial guesses, an iterative procedure was used. In particular, the parameters were introduced in eight successive fits, with the best-fit values fed forward to the next, in the order of $\{N_0, \ \tau, \ A_0, \ R, \ \phi_0\}$, $\{\tau_{\text{CBO}}, \ \omega_0, \ A_{N,x,1,1}, \ \phi_{N,x,1,1}\}$, $\{\kappa_{\text{loss}}\}$, $\{\tau_y, \ \kappa_y, \ A_{N,y,2,2}, \ \phi_{N,y,2,2}\}$, $\{A_{N,x,2,2}, \ \phi_{N,x,2,2}\}$, $\{A_{A,x,1,1}, \ \phi_{A,x,1,1}\}$, $\{A_{\phi,x,1,1}, \ \phi_{\phi,x,1,1}\}$, and $\{A_{N,y,1,1}, \ \phi_{N,y,1,1}\}$. After the initial five-parameter fit, the order of the ensuing parameter groups was determined by the size of the fit residuals' FFT peaks. Also, when fitting for κ_{loss} in these successive fits, the τ parameter inside $\Lambda(t)$ in Eq. 5.29 is fixed to be 64.44 μs in order to speed up the χ^2-minimization. After the last iteration—when the parameter seeds are very good—this τ is coupled to the fitted τ, resulting in a minute change in R of $\mathcal{O}(0.001 \text{ ppb})$.

The χ^2-minimization itself is performed with the `MnMigrad` routine in the `Minuit2` software package, using a strategy level of two [100]. This is the same routine used in the reconstruction's

simultaneous template fit, which was thoroughly characterized and described in Sec. 4.5.5. The routine's tolerance value is set to 0.1 for the iterative fits, and, at the end, the fit is repeated with a lower tolerance of 0.001. Repeated minimizer calls change the best-fit R on the order of 0.01 ppb, which is negligable compared to the statistical precision that will be found.

With the strategy level of two, the parameters' uncertainties returned by the `MnMigrad` routine are, in fact, determined by the `MnHesse` routine in the same software package [100]. More specifically, the uncertainties are estimated by calculating the Hessian matrix at the minimum by finite differences and inverting it—implicitly assuming a parabolic shape in each parameter's χ^2 space. Even though this assumption is likely sufficient, the uncertainty on R—the primary parameter of interest—quoted in this dissertation is instead found using the `MnMinos` routine (with a tolerance value of 0.001) also in the `Minuit2` software package [100]. This routines follows the χ^2-function away from the minimum to locate where it increases by one—with no assumptions as to the χ^2-function's curvature. For each data set, the difference in R's uncertainty reported by the `MnHesse` and `MnMinos` routines, however, is only of $\mathcal{O}(0.001$ ppb$)$.

Moreover, the time bins used in Eq. 5.43 are those between 30.2876 and 650.0644 μs. A lower bound on the start time of around 30 μs was chosen based on how quickly the quadrupole plates can reach their nominal voltage following the scraping process that ends at about 7 μs after injection. This limit corresponds to 4.6 times the quadrupole system's typical (resistor-capacitor) time constant (Sec. 2.4). However, for those plates with damaged/defective resistors mentioned in Sec. 5.6, this limit corresponds to only about one time constant. If the fit is started earlier, the added muon dynamics due to the voltage transition would require further modeling in the fit function, which would be challenging. From 30 μs, the exact fit start time used was selected to correspond to the first node of the ω_a's oscillation. As will be seen in Ch. 6, doing so reduces the best-fit R's sensitivity to the gain correction applied in Ch. 4. The stop time, on the other hand, was chosen so that
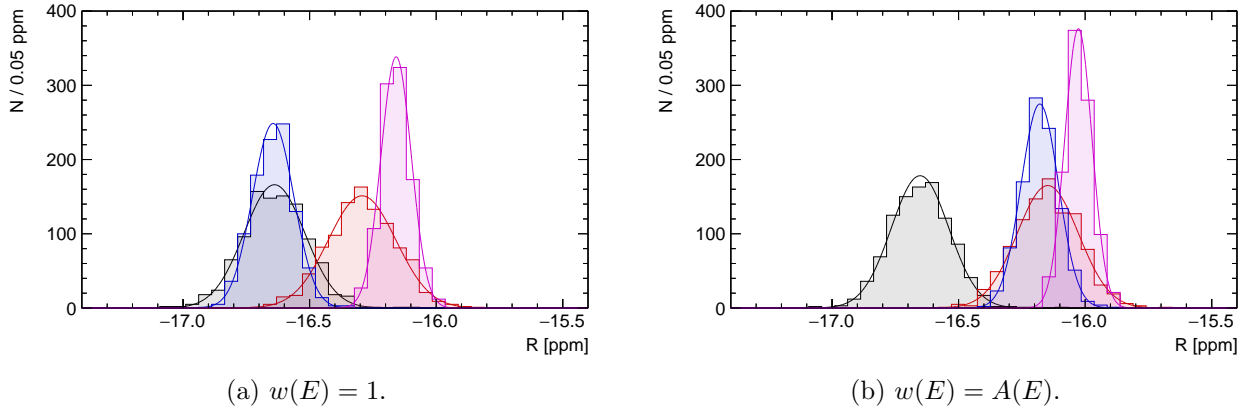
(a) $w(E) = 1$.
(b) $w(E) = A(E)$.

Figure 5.21: The R distribution for 1000 unique seeds of the random number generator used to randomize each muon fill's injection time (Sec. 5.6.1), using the (a) baseline and (b) asymmetry-weighted methods on the 1a (black), 1b (red), 1c (blue), and 1d (magenta) data sets.

(a) $w(E) = 1$.
(b) $w(E) = A(E)$.

Figure 5.22: The reduced-χ^2 distribution for 1000 unique seeds of the random number generator used to randomize each fill's injection time (Sec. 5.6.1), using the (a) baseline and (b) asymmetry-weighted method on the 1a (black), 1b (red), 1c (blue), and 1d (magenta) data sets.
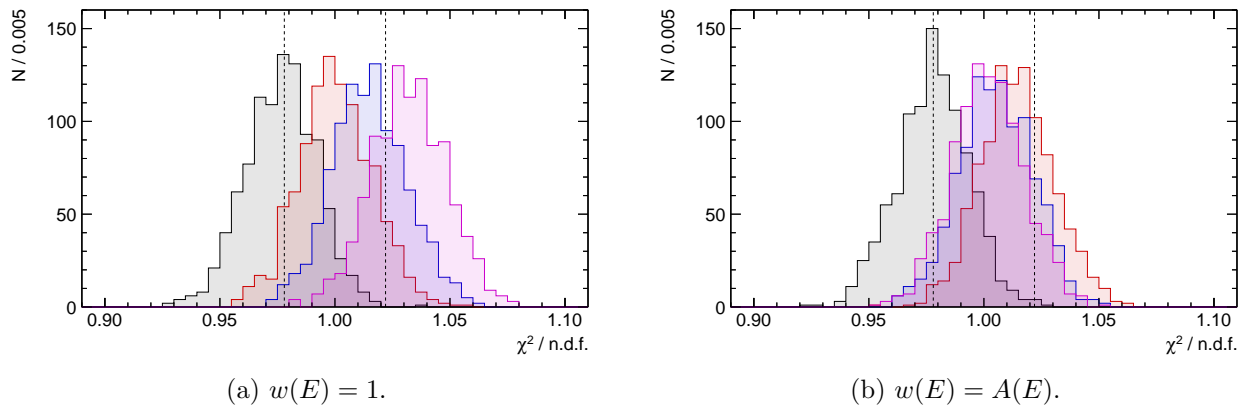
at least 30 counts are present in each time bin used, allowing the variance to be well modeled by a normal distribution as needed for a χ^2-fit approach. The stability of the fit parameters on these start/stop time choices will be presented in Sec. 5.7.2.

The best-fit R is presented in Fig. 5.21 for 1000 unique seeds of the random number generator used in the time-randomization procedure (Sec. 5.6.1). This randomization of the fills' injection times reshuffles the statistics in the fitted histogram, leading to a spread in the best-fit R and χ^2 found when using different seeds—observed in Figs. 5.21 and 5.22. As expected, the spread in R

Data set	R [ppm]	
	$w(E) = 1$	$w(E) = A(E)$
1a	$-16.6304 \pm 1.3377 \pm 0.0350$	$-16.6591 \pm 1.2077 \pm 0.0270$
1b	$-16.3156 \pm 1.3845 \pm 0.0565$	$-16.1410 \pm 1.2508 \pm 0.0574$
1c	$-16.6637 \pm 0.8829 \pm 0.0265$	$-16.1781 \pm 0.7971 \pm 0.0271$
1d	$-16.1485 \pm 0.6311 \pm 0.0260$	$-16.0182 \pm 0.5706 \pm 0.0186$

Table 5.6: The best-fit R found for each data set in Run 1, using the chosen seed of the random number generator used to randomize the fills' injection times (Sec. 5.6.1). From left to right, the uncertainties are statistical and systematic. The combined R-uncertainty is 450 and 410 ppb for the baseline and asymmetry-weighted method, respectively. Note that the same software-blinding offset is used across the data sets (Sec. 5.1) and that several corrections need to be applied once their analyses are completed (Secs. 5.8, 6.7.5, and 6.8.3).

closely follow a normal distribution, with a standard deviation ranging between 60 and 134 ppb among the data sets. Monte Carlo simulations, however, indicate that the time randomization procedure does not alter the time bins' statistical fluctuations. Therefore, the observed changes in R when using different seeds should already be accounted for in the statistical uncertainty. As a result, each best-fit R found is equally valid. The spread in R due to the choice of seed, though, should be considered if comparing results among independent analyses.

Nonetheless, in this dissertation's analysis, the random-number generator's seed used for each data set was chosen to yield representative R and χ^2 values (for both the baseline and asymmetry-weighted methods). Note that, among the results using different seeds, the correlation between R and χ^2 is negligible with a population correlation coefficient less than 0.02 % in magnitude. The best-fit R values using the chosen seeds are presented in Tab. 5.6, including their systematic uncertainties to be detailed in the next chapter. The statistical uncertainty of R is consistent with that predicted by Eq. 2.30, ranging from 0.63 to 1.38 ppm among the data sets. The difference in R for each pair of data sets is below 0.52 ppm and is consistent within a standard deviation. This agreement suggests that the magnetic field's strength—which is proportional to ω_a up to the

Parameter		$w(E) = 1$		$w(E) = A(E)$	
		Value	$\rho_{i,R}$ [%]	Value	$\rho_{i,R}$ [%]
R	[ppm]	$-16.6637(8829)$	100.0	$-16.1781(7971)$	100.0
$N_0 \times 10^{-6}$		$7.9205(25)$	-0.1	$3.8040(10)$	-0.2
τ	[μs]	$64.4412(26)$	-0.2	$64.4407(22)$	-0.2
A_0		$0.370090(29)$	0.5	$0.354023(25)$	0.5
ϕ_0	[rad]	$2.08025(15)$	-82.8	$2.07965(13)$	-82.8
ω_0	[rad/μs]	$2.61162(34)$	2.4	$2.61126(27)$	2.3
τ_{CBO}	[μs]	$211(15)$	-0.5	$213(12)$	-0.6
$A_{N,x,1,1}$		$0.003277(92)$	-0.6	$0.003620(80)$	-0.7
$\phi_{N,x,1,1}$	[rad]	$3.490(28)$	-3.1	$3.530(22)$	-3.1
κ_{loss}		$0.00518(26)$	-0.2	$0.00517(22)$	-0.3
κ_y		$1.01455(78)$	0.4	$1.01463(69)$	0.5
τ_y	[μs]	$47(13)$	-0.2	$42(11)$	-0.3
$A_{N,y,2,2}$		$0.0027(13)$	0.2	$0.0034(17)$	0.4
$\phi_{N,y,2,2}$	[rad]	$3.54(45)$	1.0	$3.58(39)$	1.0
$A_{A,x,2,2}$		$0.000166(62)$	-0.5	$0.000162(53)$	-0.4
$\phi_{N,x,2,2}$	[rad]	$3.76(37)$	-1.1	$3.75(33)$	-1.1
$A_{A,x,1,1}$		$0.00037(16)$	-2.6	$0.00059(15)$	-3.7
$\phi_{A,x,1,1}$	[rad]	$2.71(44)$	-4.1	$3.08(24)$	-3.0
$A_{\phi,x,1,1}$		$0.000234(82)$	-0.1	$0.000241(74)$	-0.1
$\phi_{\phi,x,1,1}$	[rad]	$3.95(35)$	-3.1	$3.98(30)$	-3.0
$A_{N,y,1,1}$		$0.00031(14)$	0.3	$0.00033(15)$	0.4
$\phi_{N,y,1,1}$	[rad]	$4.15(46)$	-0.8	$4.21(41)$	-0.9

Table 5.7: The best-fit values of the 22 free parameters in Eq. 5.36, found using the 1c data set with the baseline and asymmetry weighting schemes. The correlation of each parameter i with R, $\rho_{i,R}$, is also given. The correlation among other parameters is shown in Fig. 5.23.

electric-field correction (Eq. 2.33)—was consistent to $\mathcal{O}(1\ \mathrm{ppm})$ across Run 1.

All of the best-fit parameters for the 1c data set, as an example, are shown in Tab. 5.7. The decoherence time constant τ_{CBO}, in particular, is significantly smaller than that measured more directly by the straw tracker [126]. This is likely because the $j = 1$ term in $\mu_2(t)$ of $N_x(t)$ was neglected to improve the fit's stability (Sec. 5.6.3). This in turn effectively changes the decoherence envelope model, pulling the best-fit time constant to a lower value. When such a term is included

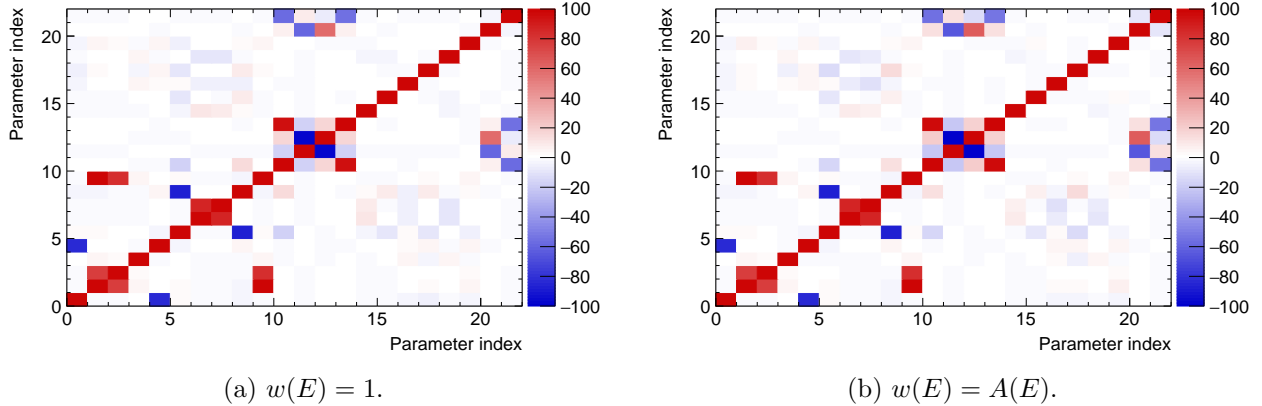(a) $w(E) = 1$.

(b) $w(E) = A(E)$.

Figure 5.23: The correlation matrix, in percent, among the 22 fit parameters in Eq. 5.36, found using the 1c data set with the (a) baseline and (b) asymmetry weighting schemes. The parameters are indexed in the order given in Tab. 5.7 from top to bottom. Accordingly, the lowest and highest parameters indices correspond to R and $\phi_{N,y,1,1}$, respectively.

in the fit function, τ_{CBO} increases to 315 ± 74 μs—statistically consistent with the tracker data. For the same reason, τ_y increases to 74 ± 18 μs when adding the $j = 1$ term in $\mu_2(t)$ of $N_y(t)$. In any case, for each data set, the muon dynamics cause less than a 0.4 % modulation to the acceptance at injection, and their related parameters are only weakly correlated with R. As illustrated in Fig. 5.23, R is only strongly correlated with ϕ_0 (by about -82.8 %) for every data set. The next largest R-correlation is with $A_{N,x,1,1}$, $A_{A,x,1,1}$, or their related phases, depending on the data set, and varies between 3.5 and -5.1 %. Therefore, even though most of the fit parameters come from the muon dynamics, they do not significantly affect R's uncertainty.

As a goodness of fit, the reduced-χ^2 values for the fit results in Tab. 5.6 are within 1.19, 0.08, 0.73, and 1.53 standard deviations of one for the 1a, 1b, 1c, and 1d data sets, respectively. This good agreement indicates that the 22-parameter model in Eq. 5.36 adequately describes the observed data. Further, as shown in Fig. 5.24, the FFT of the fits' residuals in the time bins used in the χ^2 does not reveal any clear peaks above the noise floor. This supports that the muon dynamics are sufficiently modeled for the size of the data sets. Also, the lack of an FFT peak near zero suggests that the slowly varying effects—such as gain changes, pileup, and muon losses—have

been sufficiently corrected or modeled. Moreover, as illustrated in Fig. 5.25, the pulls—defined as the residuals divided by their standard deviations—for the time bins used in the χ^2 are normally distributed about zero with unit width—as expected. The time distribution of pulls near the start time of the fit are well behaved with no clear structures. As seen in Fig. 5.26, the pulls are only large before around 26 μs into the muon fill due to the added muon dynamics from the scraping procedure that are not modeled in the fit function. From these metrics, the fit results using the baseline method of $w(E) = 1$ appear to be satisfactory. Even so, they will be further evaluated later in this section to increase confidence in the best-fit R for each data set.

5.7.1   Asymmetry-Weighted Method

As shown in Sec. 2.5, the best-fit ω_a's statistical precision can be improved over that from the baseline method by as much as 15 % by weighting each positron according to its asymmetry, i.e., $w(E) = A(E)$. This stems from preferentially weighting the highest-energy positrons, which have the strongest correlation with the muons' spin direction. Consequently, the asymmetry-weighted result will also be more sensitive to systematic effects that dominate at high energies such as gain changes and pileup. However, as will be presented in Ch. 6, the change in systematic precision is small relative to the statistical precision gained by using $w(E) = A(E)$.

The asymmetry function, $A(E)$, was empirically found by applying the baseline method to histograms populated with positrons in 40 disjoint energy ranges from 1 to 3 GeV. The lower bound of 1 GeV was primarily chosen to eliminate any noise contamination (Sec. 4.9), while the upper bound of 3 GeV was chosen to be approximately at the end-point of the decay positrons' energy spectrum. The energy bin width of 50 GeV corresponds to reconstruction's estimated energy resolution at 1 GeV (Sec. 4.8). Each histogram was then fit according to the procedure detailed above using the five-parameter model in Eq. 5.3. Note, however, that the stop time used was reduced to

(a) 1a data set.

(b) 1b data set.
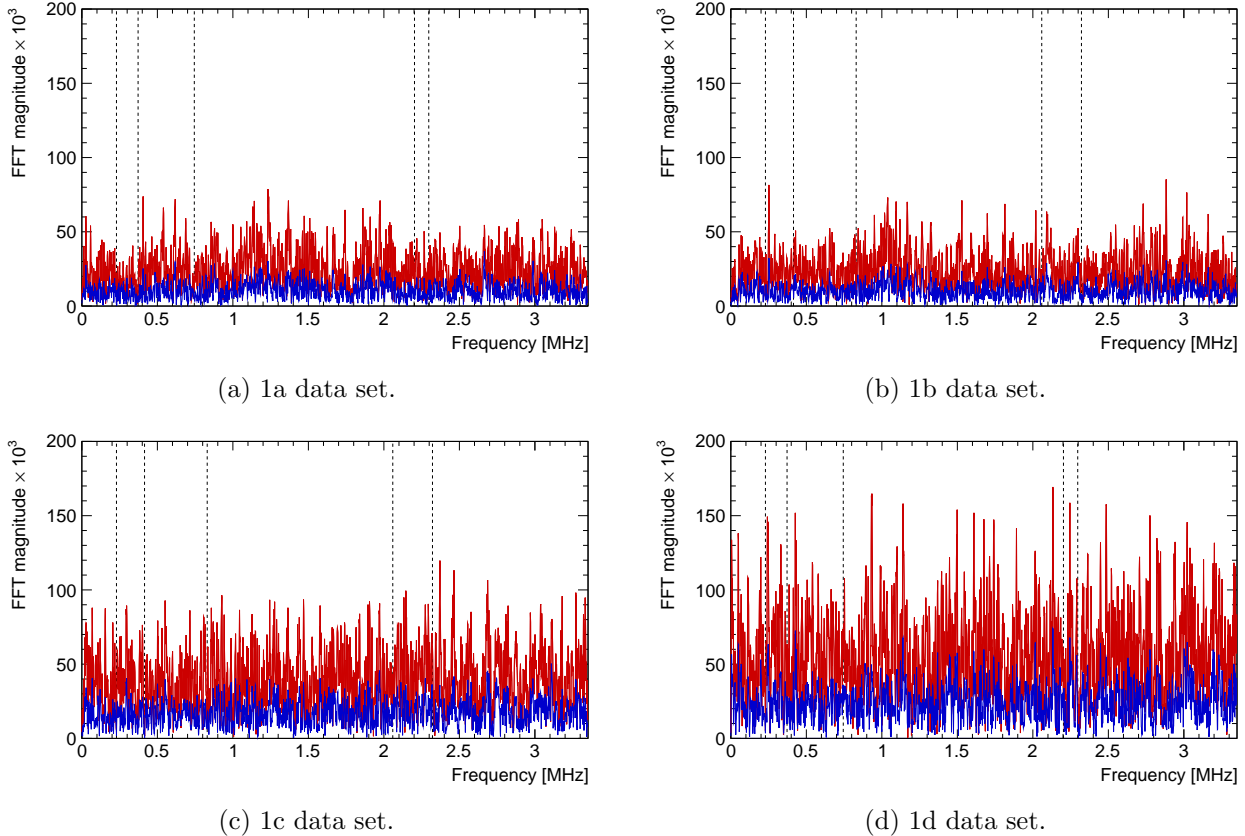
(c) 1c data set.

(d) 1d data set.

Figure 5.24: The FFT of the fit residuals using the 22-parameter model in Eq. 5.36 for the (a) 1a, (b) 1b, (c) 1c, and (d) 1d data sets when applying the baseline (red) and asymmetry (blue) weighting schemes. No clear frequency peak is visible above the noise floor, suggesting that this model adequately describes the histogrammed data.

ensure at least 30 counts in each time bin used in the χ^2. This minimal model is adequate since, as seen in Fig. 5.23, A_0 is only weakly correlated with the parameters describing the muon loss or dynamics. In any case, because the weights applied are not time-dependent, any $w(E)$-function does not pose a bias to the best-fit R. Any deviation in $A(E)$ used for weighting will only degrade the potential gain in precision. The $A(E)$-weighting function used for each data set is shown in Fig 5.27, where the best-fit A_0 extracted from each energy bin's histogram is held constant over its 50-MeV range. As detailed in Sec. 5.4, this piecewise form is advantageous because it reduces the pileup correction's correlation with the reconstructed data.

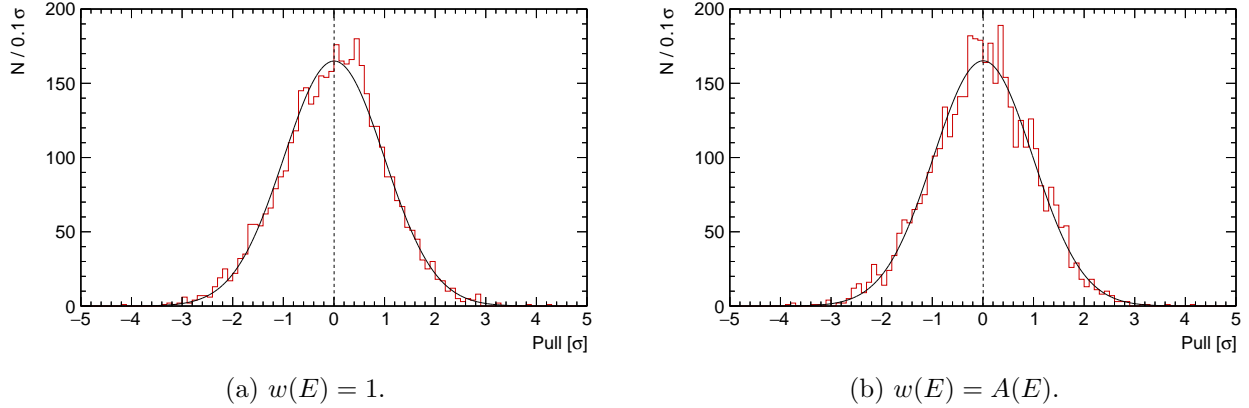For each data set, a time histogram was then populated with positrons between 1 and 3 GeV,

(a) $w(E) = 1$.

(b) $w(E) = A(E)$.

Figure 5.25: The distribution of pulls, i.e., standardized residuals, within the fit's range between 30.2876 and 650.0644 μs for the 1c data set, using the (a) baseline and (b) asymmetry weighting schemes. A normal fit to these data yields a mean of (a) 0.003(16) and (b) 0.006(0.16) and a width of (a) 0.993(12) and (b) 0.984(12).
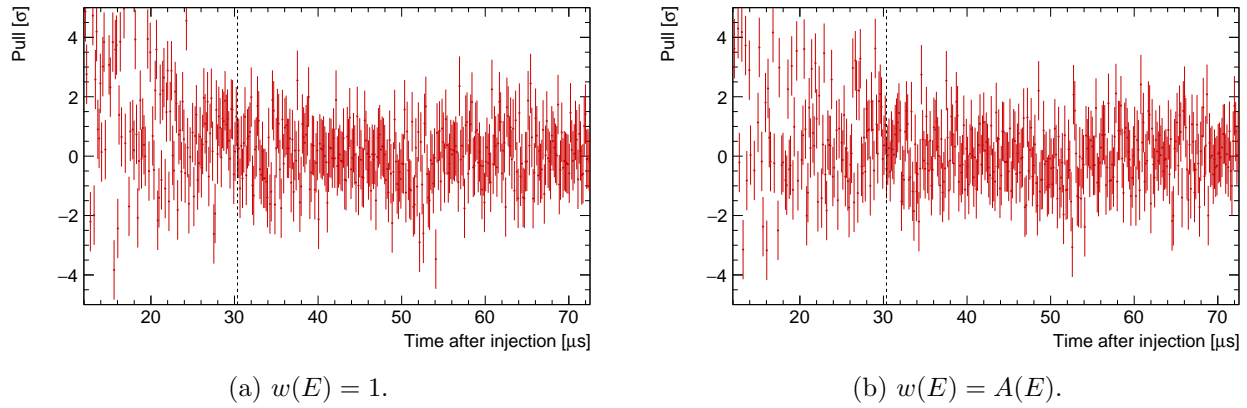

(a) $w(E) = 1$.

(b) $w(E) = A(E)$.

Figure 5.26: The time distribution of pulls, i.e., standardized residuals, around the fit's start time at 30.2876 μs, denoted by the dashed line, for the 1c data set when using the (a) baseline and (b) asymmetry weighting schemes.

weighted according to Fig 5.27. Each histogram was afterward fit following the above procedure using the 22-parameter model in Eq. 5.36. Using the same 1000 random-number generator seeds previously mentioned, the asymmetry-weighted results are presented in Figs. 5.21 and 5.22. The result using the seeds selected to yield representative R and χ^2 values for each data set is given in Tab. 5.6. Note that, for each data set, the same seed is used for both the baseline and asymmetry-weighted methods. As can be seen, R's statistical precision improved over that from the baseline method by about 10.7 % for each data set—very close to the 10.5-% prediction in Fig. 2.11. When
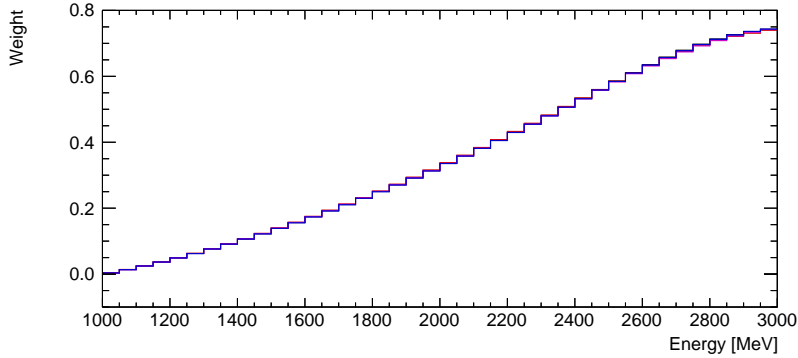
Figure 5.27: The asymmetry function, $A(E)$, found for the 1a (black), 1b (red), 1c (blue), and 1d (magenta) data sets, which is used in the asymmetry-weighted method. The weights below 1 GeV and above 3 GeV are assigned to be zero.

the four data sets are combined, the statistical precision of ω_a will become 0.4093 ppm—roughly four times the experiment's target precision after four years of data collection.

All of the best-fit parameters are in Tab. 5.7 for the 1c data set, as an example, alongside the baseline method's result. As expected, the values of some energy-dependent parameters significantly differ between the two methods. This includes the anomalous precession's phase at injection, ϕ_0, partially because of the energy dependence in the positrons' drift time (Fig. 2.10). However, as seen in Fig. 5.23, these parameters are similarly correlated with each other and with the other parameters. All of the frequencies and time constants, on the other hand, agree between the two weighting methods to about a standard deviation.

The two methods' results, though, are partially correlated because their histograms were constructed from the same underlying positrons. To estimate the allowed variation in R between the two methods, a Monte Carlo simulation of 100 pseudo-experiments was conducted for each data set. In each pseudo-experiment, time histograms—using both weighting methods—were built in 10 MeV ranges according to the five-parameter model in Eq. 2.29, where N_0 and A_0 were empirically estimated from fits to each data set. The time histograms in the energy ranges used in this chapter's analyses were then combined, incorporating the expected energy resolution in Eq. 4.38,
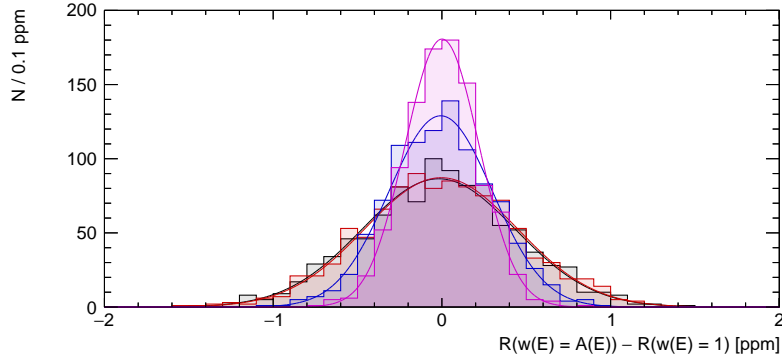
Figure 5.28: The difference between the best-fit R found with the baseline method, $R(w(E) = 1)$, and the asymmetry-weighted method, $R(w(E) = A(E))$, for 100 pseudo-experiments mimicking the 1a (black), 1b (red), 1c (blue), and 1d (magenta) data set. Note that, for each data set, the statistics used agrees with those measured to less than 0.005 %.

and fit to the five-parameter model. The distribution of the R-differences between the baseline and asymmetry-weighted methods is given in Fig. 5.28. As can be seen, the two methods' best-fit R values typically differ on the order of 0.1 ppm. This difference in R measured in each data set—reported in Tab. 5.8—is less than 1.6 standard deviations of that expected from this Monte Carlo simulation. Therefore, the R values extracted using the two methods are concluded to be consistent with each other. In fact, if the R values from each data set in Tab. 5.6 are combined, the baseline and asymmetry-weighted results would agree to within 1.4 standard deviations of that expected from the above simulation.

Moreover, as with the baseline method, the asymmetry-weighted method's fit result appears to sufficiently describe the data. The reduced-χ^2 for this method's fit results in Tab. 5.6 are within 0.90, 0.59, 0.09, and 0.13 standard deviations of one for the 1a, 1b, 1c, and 1d data sets, respectively. As seen in Fig. 5.24, the FFT of the fit residuals in the time bins used in the χ^2 also does not contain any clear peaks above the noise floor. Further, as illustrated in Figs. 5.25 and 5.26, the pulls are normally distributed as expected and are well behaved around the fit start time. Therefore, the fit results when weighting by $A(E)$ appear to be satisfactory but, along with the baseline method's results, will, nonetheless, be more extensively evaluated next.

Data set	$R(w(E) = A(E)) - R(w(E) = 1)$ [ppm]	Expected 1σ deviation [ppm]
1a	$-0.0174(18)$	$0.454(11)$
1b	$0.1818(20)$	$0.452(11)$
1c	$0.4869(14)$	$0.3059(71)$
1d	$0.13958(80)$	$0.2186(51)$

Table 5.8: The R-difference found in data between the baseline and asymmetry-weighted method, $R(w(E) = A(E)) - R(w(E) = 1)$. Each data set's R-difference is less than 1.6 standard deviations of that expected from the Monte Carlo simulation's results shown in Fig. 5.28. The quoted uncertainties are statistical only.

5.7.2 Start/Stop Time Scans

While necessary, no clear structure in the fit residuals is insufficient to verify that the fit result is reliable. For example, if the fit starts at a much earlier time near 20 μs, the pulls' distribution can still appear adequate despite the muon dynamics from the quadrupole system's recovery from scraping not being explicitly modeled. This is because there are enough free parameters in the fit to compensate for the different dynamics. The best-fit values, however, will not be stable with increasing fit start times, as such time-dependent effects diminish. More generally, this is true for any uncorrected time-dependent effect—as demonstrated in Fig. 5.29—which poses a bias to ω_a (Sec. 2.6). It is, therefore, important to scan over the fit's start time to verify that there is no significant time dependence of the fit parameters.

For a later start time, the statistics will decrease and yield a less precise value for some of the fit parameters, including R. As a result, some variation in the best-fit values is expected due to the difference in statistics in the fit. The standard deviation of this allowed statistical drift can be shown to be approximated by

$$\Delta\sigma = \pm\sqrt{\sigma_1^2 + \sigma_2^2 - 2f_{21}\sigma_1^2} \tag{5.44}$$

where σ_1 and σ_2 are the given parameter's statistical uncertainty when the fit starts at a time t_1

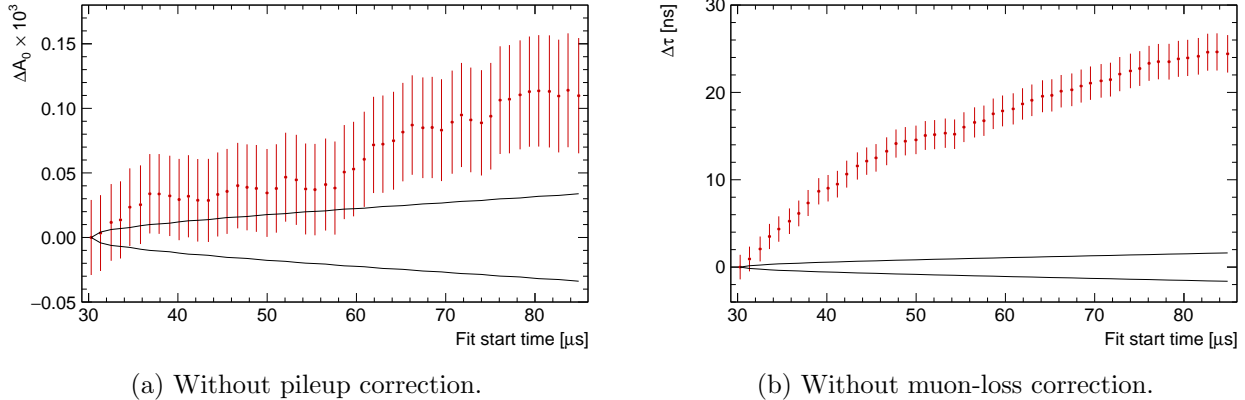(a) Without pileup correction. (b) Without muon-loss correction.

Figure 5.29: A scan over the fit start time for the 1c data set when not correcting for (a) pileup or (b) muon losses. The bands (black) show the standard deviation of the expected drift from statistical differences (Eq. 5.44). As shown, the best-fit asymmetry, A_0, and boosted muon lifetime, τ, found using the baseline method significantly depart from their allowed bands when a time-dependent effect is uncorrected. Similar trends are found when using the asymmetry-weighted method.

and a later time t_2 [33]. Here, f_{21} is the fraction of counts in the fitted data set using t_2 that are also in the fitted data set using t_1—which for a fit start (or stop) time scan is one. Moreover, as the data fit using two different start times share a subset, they are partially correlated. The trend in a fit parameter's statistical drift, therefore, will be moderately smooth and not scattered about zero uniformly from one start time to the next. Any significant trend deviating beyond the $\pm\Delta\sigma$-expectation band is evidence of an uncorrected early-to-late effect.

For each weighting method, the fit start time was scanned over more than 12 ω_a-periods from the nominal start time of 30.2876 μs—with about four start times sampled per ω_a-period. When starting the fit at later times, however, the vertical N_0-modulations are increasingly difficult to observe. Therefore, as they are not vital parameters of interest, κ_y and τ_y are fixed to their values found using the nominal start time, while the related asymmetries and phases are still allowed to float. This, in essence, helps tell the fitter where to look for the vertical oscillations.

In the 1a, 1b, and 1c data sets, all of the best-fit values are consistent with purely statistical variations, remaining within one standard deviation of the allowed $\pm\Delta\sigma$-drift. There is also
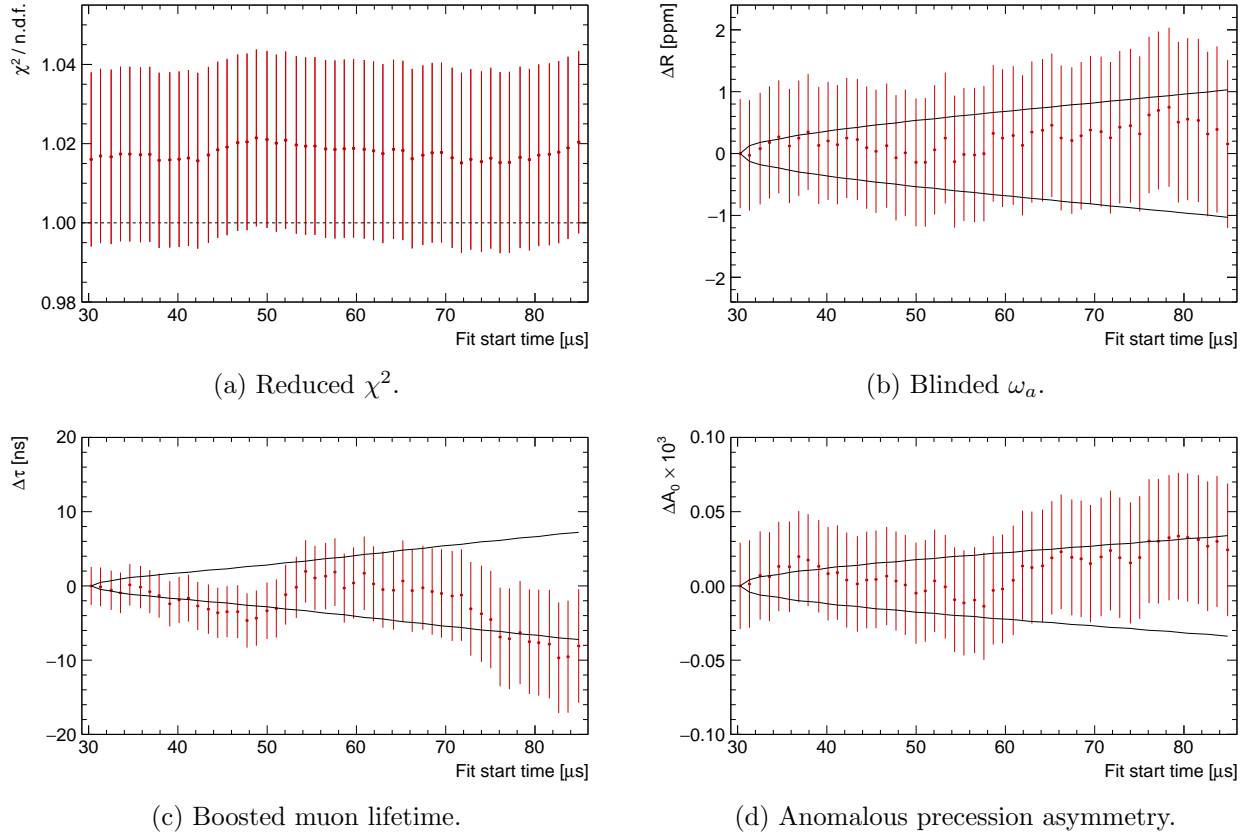
(a) Reduced χ^2.

(b) Blinded ω_a.

(c) Boosted muon lifetime.

(d) Anomalous precession asymmetry.

Figure 5.30: A scan of (a) the reduced-χ^2 and the change in (b) R, (c) τ, and (d) A_0 over the fit start time, using the baseline method with the 1c data set. The bands (black) show the allowed drift due to statistical differences (Eq. 5.44).

no evident peak close to the aliased frequency of ω_a in the FFT of each fit start time scan for R.

Moreover, note that many of the parameters' fit start time scans are very similar due to their cor-

relations. For example, the trends in R and ϕ_0 are nearly perfect reflections of each other. A few

of the representative scans for the 1c data set are shown in Fig. 5.30. In the 1d data set, how-

ever, the (highly correlated) N_0, τ, and κ_{loss} parameters significantly deviate from their expected

$\pm\Delta\sigma$-bounds. This is caused by a known inaccuracy in the gain correction applied to the 1d data

set's clusters, which will be resolved in the coming months. For the purpose of this dissertation,

this inaccuracy's influence on R will be assessed in Sec. 6.2 as a systematic uncertainty. In any

case, this inaccuracy serves as an excellent demonstration of how a fit start time scan can reveal
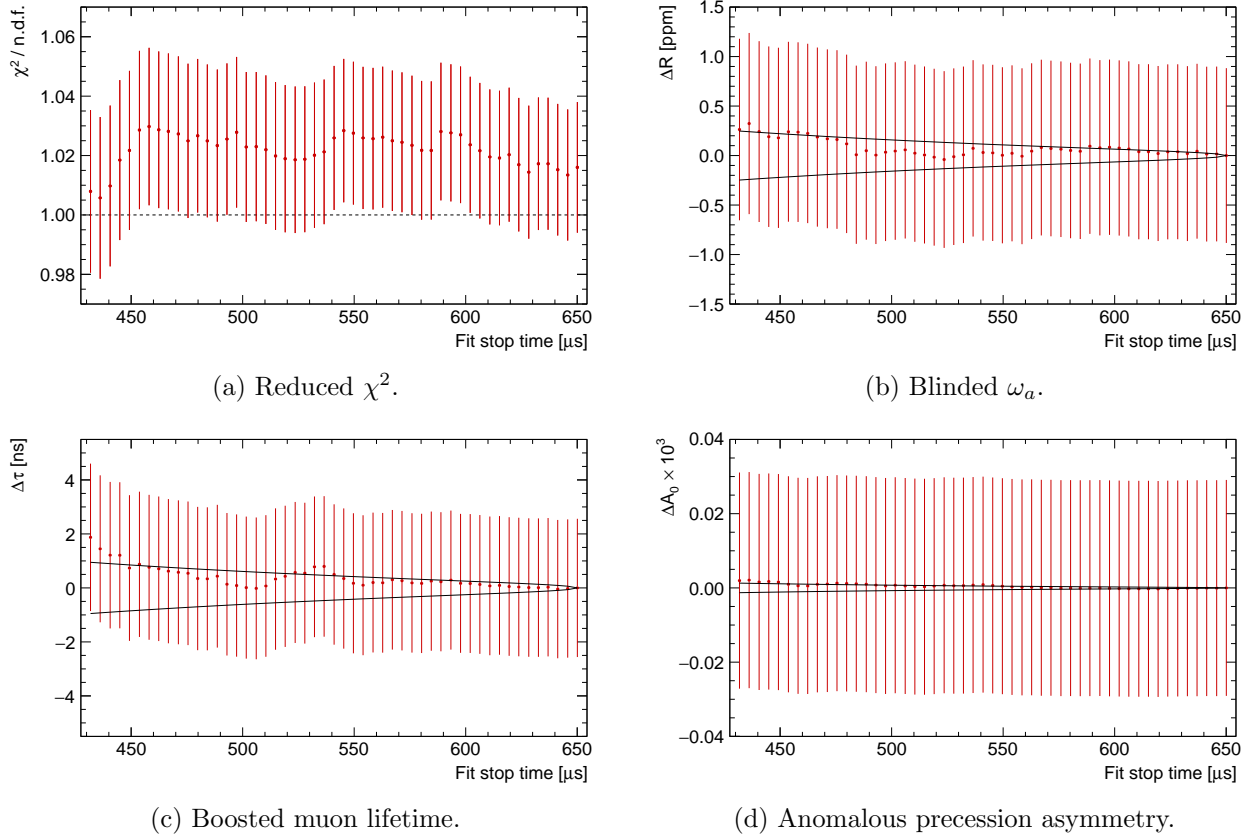
(a) Reduced χ^2.

(b) Blinded ω_a.

(c) Boosted muon lifetime.

(d) Anomalous precession asymmetry.

Figure 5.31: A scan of (a) the reduced-χ^2 and the change in (b) R, (c) τ, and (d) A_0 over the fit stop time, using the baseline method with the 1c data set. The bands (black) show the allowed drift due to statistical differences (Eq. 5.44).

a time-dependent effect that is hidden from the fit residuals' metrics.

Furthermore, for the same reason, the best-fit values should also not depend on the fit's stop time. In addition, later stop times cause the time bins to have fewer counts, worsening the implicit assumption in Eq. 5.43 that their variances are well modeled by normal distributions. Therefore, for each weighting method, the fit stop time was scanned over more than 50 ω_a-periods before the nominal stop time of 650.0644 μs—with about one stop time sampled per ω_a-period. Note that, in these fits, none of the 22 fit parameters were fixed. For every data set, all of the best-fit values remain within a standard deviation of the allowed $\pm\Delta\sigma$-drift. Some of the representative scans for the 1c data set are shown in Fig. 5.31 as examples. These results help build confidence

that all relevant effects at late times have been captured and that the data are not contaminated with noise clusters (Sec. 4.9). It also indicates that the χ^2-minimization method—as opposed to a maximum likelihood method—is sufficient for the nominal stop time used.

5.7.3 Statistical Consistency

To further look for any inadequacies in the analysis, each data set was divided by calorimeter, energy range, muon bunch in the accelerator's cycle, crystal column, and crystal row, which compose internal consistency checks. Many of the muon-dynamics modulations, for instance, are enhanced when applying the above fit procedure to data from only one calorimeter. This enables the limits of their modeling in the fit function to be explored. Furthermore, some of the fit parameters are expected to vary when sorting the data in these different ways, while some are not. An excellent example discussed in Sec. 5.6.1 is that R ought to remain constant among individual calorimeter fits. As seen in Fig. 5.17, the observation of a calorimeter-dependent bias can indicate a residual fast-rotation signal in the fitted data. These predictions can be tested to build more confidence in the final result of R when the data are all combined.

However, when any data set is divided up even by a half, the signal of the vertical modulations on N_0 becomes too small for a reasonable fit to all of the parameters in $N_y(t)$. Therefore, as done in the fit start time scans, κ_y and τ_y are fixed to their nominal value found above for each subset's fit. As a consequence, these fit parameters' correlations with R will no longer be included in the statistical uncertainty found by the `MnMinos` routine, as discussed in Ch. 6. In addition, some of the systematic effects are correlated between subsets, such as between energy bins (Sec. 5.4). With that in mind, the best-fit R values among the subsets can be approximately combined by computing the weighted sum given by

$$R = \left[\sum_i R_i/\sigma_{R_i}^2 \right] \left[\sum_i 1/\sigma_{R_i}^2 \right]^{-1} \tag{5.45}$$

with an approximate statistical uncertainty of

$$\sigma_R = \left[\sum_i 1/\sigma_{R_i}^2 \right]^{-1/2} \tag{5.46}$$

where R_i and σ_{R_i} are the R value and uncertainty in subset i. This combined-R value should be statistically consistent with the nominal result in Tab. 5.6 and, therefore, serves as an important check. The fit stop time used for the subsets, however, is earlier than the nominal value to ensure more than 30 counts in each time bin. The allowed deviation from a difference in statistics can, therefore, be calculated using Eq. 5.44.

As an initial consistency check, the data from each of the 24 calorimeters around the magnetic storage ring were fit separately. The fit stop time used ranged between roughly 450 and 550 μs, depending on the data set's size. As presented in Fig. 5.32 for the 1c data set, the fits to each calorimeter's data yielded a reduced-χ^2 reasonably consistent with one, indicating that the 22-parameter model is still adequate. Also, for every data set, the best-fit values of R, τ, κ_{loss}, ω_0, and τ_{CBO} are all statistically consistent, as expected, across the calorimeters—as well as with their nominal values in Tab. 5.6. A constant fit to each data set's R values, for example, yields a reduced-χ^2 of one to within 1.3 standard deviations, while the combined-R value using Eq. 5.45 is less than 1.5 standard deviations of the expected statistical drift away from the nominal value. Moreover, if R versus calorimeter index is fit to a sine wave, the best-fit amplitude is consistent with zero, suggesting that the time-randomization procedure in Sec. 5.6.1 is sufficient.

Furthermore, the calorimeters' absolute energy calibration (Sec. 5.3) was also only intended to be approximate, so their true energy thresholds used may differ slightly. There is also a different amount of material shadowing each calorimeter, leading to unique acceptances. Consequently, the energy-dependent fit parameters, such as A_0 and $A_{N,x,1,1}$, are not expected to be the same across the calorimeters—as indeed observed in Figs. 5.32 and 5.33. It can also be seen in Fig. 5.33 that $A_{N,x,1,1}$ is between 5.6 and 7.2 times larger than in the data set's calorimeter-combined fit. This
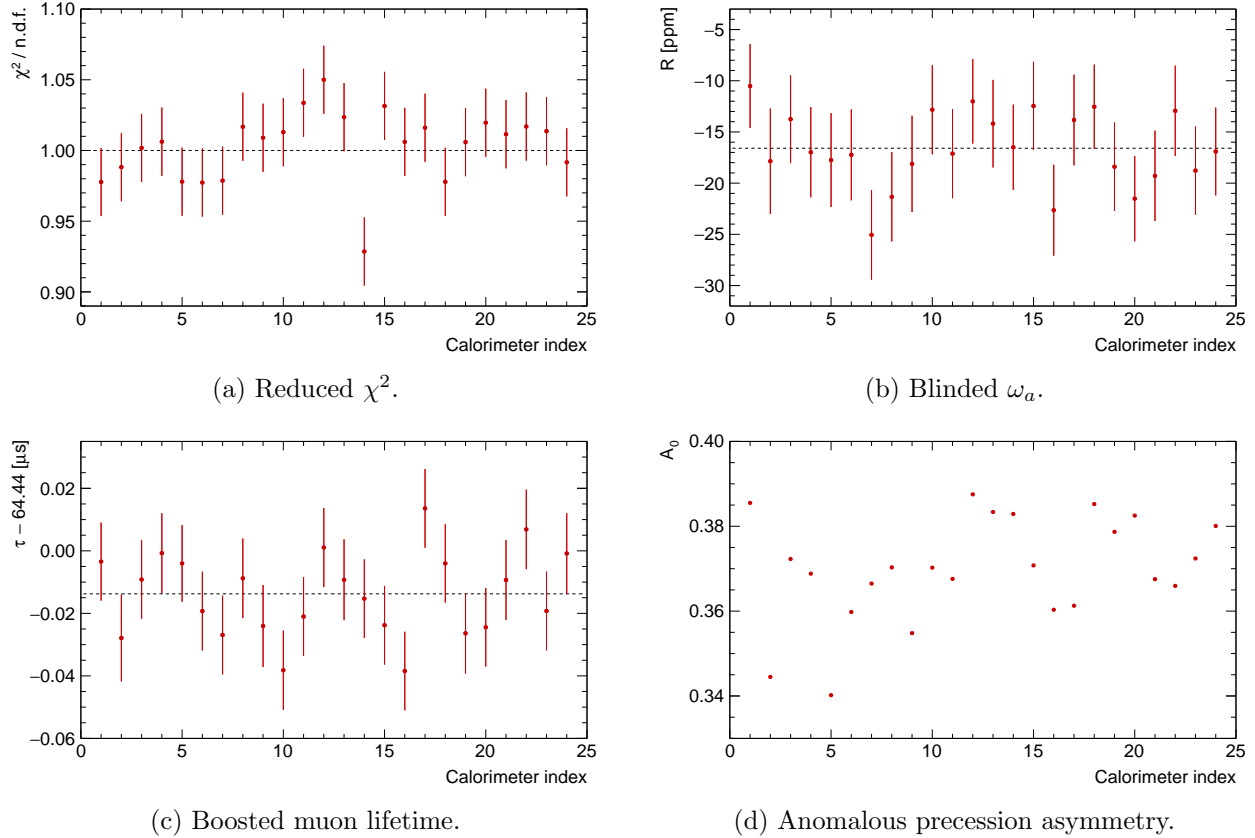
(a) Reduced χ^2.

(b) Blinded ω_a.

(c) Boosted muon lifetime.

(d) Anomalous precession asymmetry.

Figure 5.32: The (a) reduced-χ^2, (b) R, (c) τ, and (d) A_0 for each calorimeter in the 1c data set using the baseline method. The dashed lines in (b) and (c) are the result of constant fits to their data. For reference, the muons are injected into the storage ring near calorimeter 2.

is because the phase, $\phi_{N,x,1,1}$, also varies around the ring—which depends on the data set's field index (Sec. 2.4)—and because each calorimeter samples a different azimuthal range. What is seen in the nominal, calorimeter-combined fit is the oscillation's imperfect deconstructive interference around the ring. Additionally, each data set exhibits two phase jumps in $\phi_{N,x,1,1}$ at calorimeters 12–14 and 18–20. These phase variations—as well as the observed variation in $A_{N,x,2,2}$—agree very well with those predicted from tracker measurements combined with estimates of the calorimeter acceptances from simulation [129]. Only $A_{N,y,2,2}$'s variation across the calorimeters is found to be unusual. However, as illustrated by the overlay in Fig. 5.33d, this is explained in simulation when accounting for the quadrupole system's damaged/defective resistors in Run 1 [130]. All together,

(a) Amplitude of μ_1-moment in $N_x(t)$.

(b) Phase of μ_1-moment in $N_x(t)$.

(c) Amplitude of μ_2-moment in $N_x(t)$.

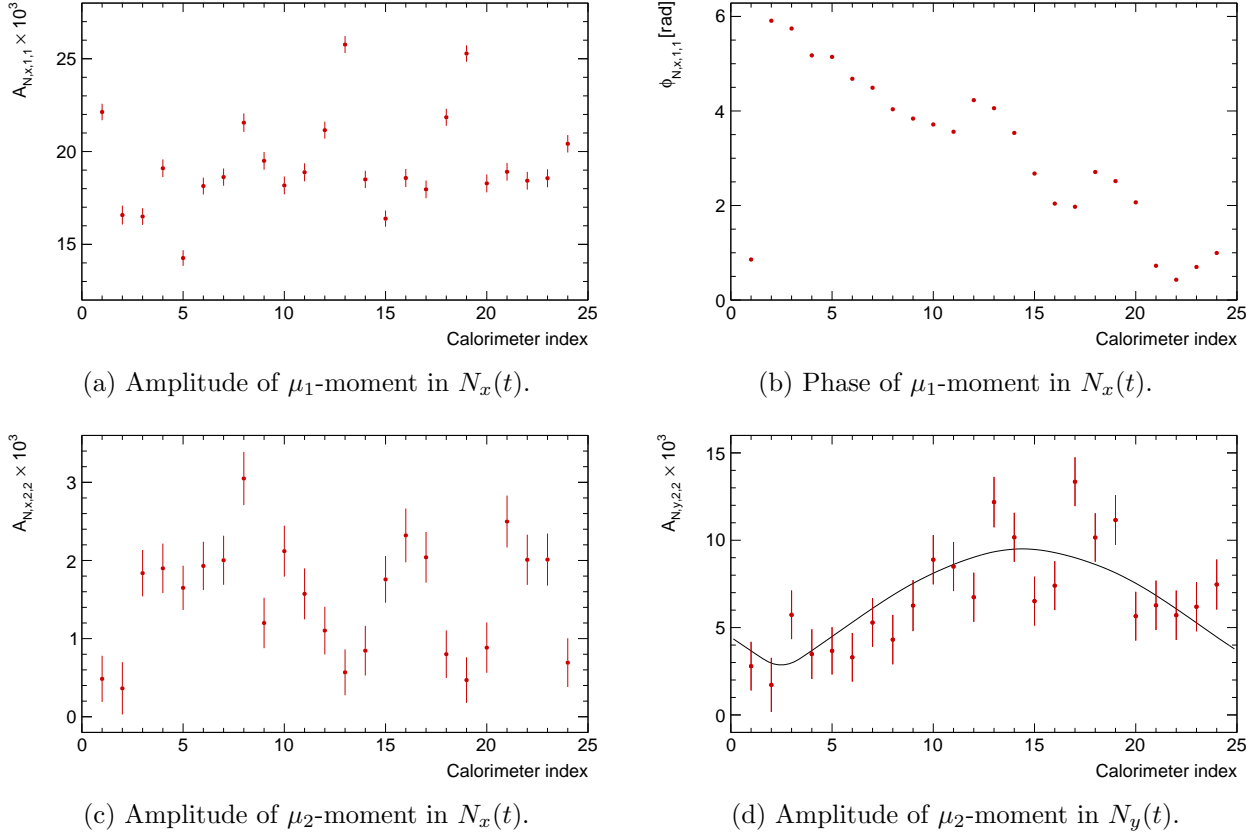(d) Amplitude of μ_2-moment in $N_y(t)$.

Figure 5.33: The value of (a) $A_{N,x,1,1}$, (b) $\phi_{N,x,1,1}$, (c) $A_{N,x,2,2}$, and (d) $A_{N,y,2,2}$ for each calorimeter in the 1c data set using the baseline method. The shape of $A_{N,y,2,2}$'s variation around the ring, predicted from simulation with the quadrupole system's damaged/defective resistors, is overlaid (black) [130]. For reference, the muons are injected into the storage ring near calorimeter 2.

the fit results per calorimeter are as expected with the 22-parameter model sufficiently describing the enhanced muon-dynamics effects.

As another consistency check, the calorimeter-combined data were divided in disjoint, 100-MeV energy ranges from 0.5 to 3 GeV. For each energy bin, the time histogram was fit only using the baseline method with $w(E) = 1$. With this small energy-bin width, the asymmetry weights in Fig. 5.27 that would be applied simply amount to scaling each histogram's normalization. In any case, as will soon be seen, the number of counts in each energy range varies over six orders of magnitude from 0.5 to 3 GeV. The fit stop time used was, therefore, individually adjusted for each energy bin's histogram, being as early as about 270 μs. Also due to low statistics in some of
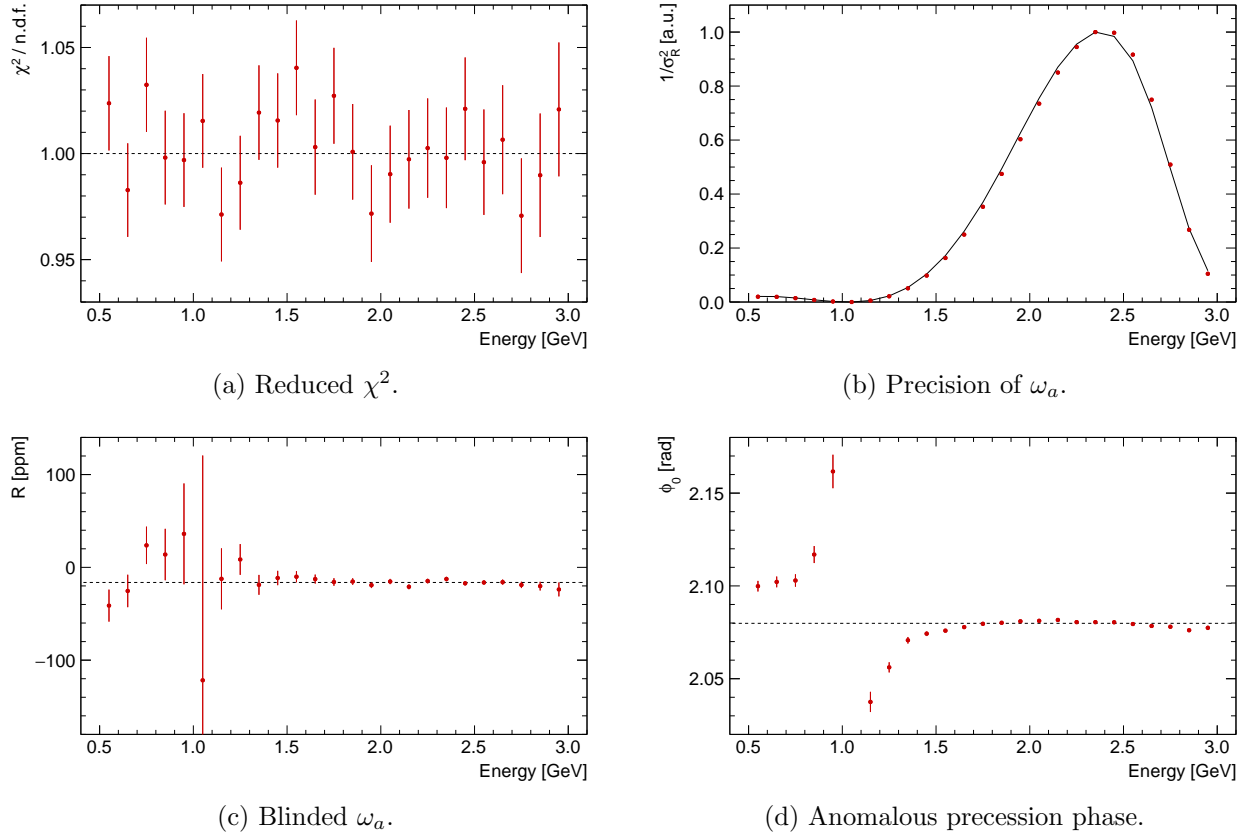
(a) Reduced χ^2.

(b) Precision of ω_a.

(c) Blinded ω_a.

(d) Anomalous precession phase.

Figure 5.34: The (a) reduced-χ^2, (b) $1/\sigma_R^2$, (c) R, and (d) ϕ_0 for each energy bin in the 1c data set using the baseline method. For each energy bin, the product $N_0 A_0^2$ from the best-fit values in Fig. 5.35, normalized to the largest value, is shown (black) on top of ω_a's precision. The dashed lines in (c) and (d) are the result of constant fits to their data. Note that the downward trend of R in the highest energy bins is believed to be a statistical fluctuation as it is not found in other three data sets in Run 1.

the histograms, ω_0 and τ_{CBO} were fixed to their nominal values for all of the fits to improve the χ^2-minimization's stability. The fit results are shown in Figs. 5.34 and 5.35 for the 1c data set, which is representative of the other data sets.

For every data set, the reduced-χ^2 values from the energy bins' fits are well distributed about one, and the R values are consistent with one another. The most imprecise R values occur around 1.05 GeV, where the anomalous-precession asymmetry, A_0, changes sign. A constant fit to the R values yields a reduced-χ^2 of one to within 1.3 standard deviations for all of the data sets. Further, when the energy bins' R values are combined via Eqs. 5.45 and 5.46, they are consistent with
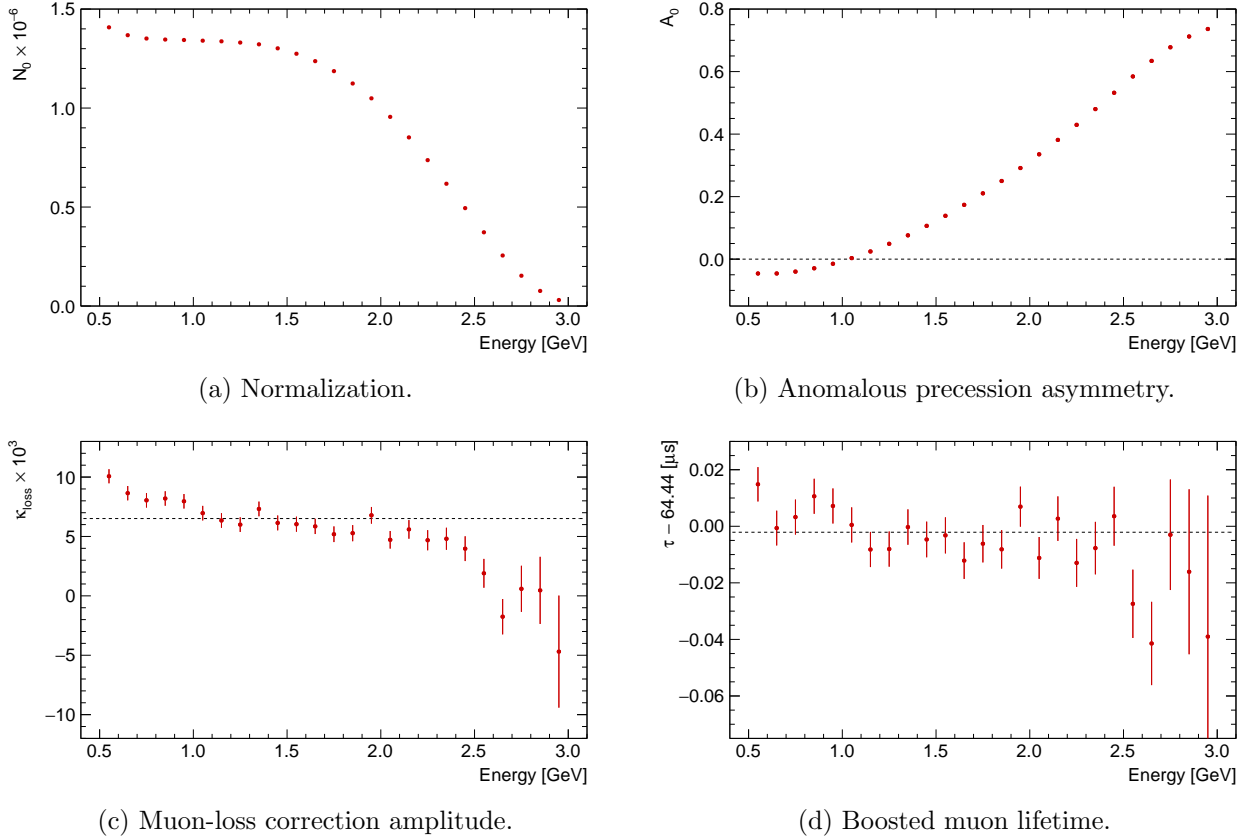
(a) Normalization.

(b) Anomalous precession asymmetry.

(c) Muon-loss correction amplitude.

(d) Boosted muon lifetime.

Figure 5.35: The value of (a) N_0, (b) A_0, (c) κ_{loss}, and (d) τ for each energy bin in the 1c data set using the baseline method. The dashed lines in (c) and (d) are the result of constant fits to their data. Note that the unexpected trend in κ_{loss} and τ with energy is very likely due to an uncorrected gain effect, as demonstrated in Sec. 6.2.

the nominal values in Tab. 5.6 to less than $0.9 \cdot \Delta\sigma$. The combined-R values are also about 8.6 % *more* precise than those from the baseline method—but about 1.2 % *less* precise than those from the asymmetry-weighted method. In general, these combined measurements present an alternative approach to weighting the positrons by $A(E)$. However, there are (small) correlations among the energy bins' data—such as from the pileup correction applied—that complicate the proper R-combination and systematic uncertainty assessment. Consequently, in this chapter's analysis, the combined-R values are only used as a cross-check.

In addition, the anomalous-precession phase, ϕ_0, is observed to vary between 1.5 and 3.0 GeV by about 10 mrad—about half of that expected from the energy dependence of the positron drift

time. The shape of ϕ_0 versus energy also does not match that of the drift time versus energy in Fig. 2.10. This is, at least in part, due to the energy dependence of the calorimeter's acceptance, which is described by $\phi_x(t)$ in Eq. 5.40. The fit results in Fig. 5.35, though, do reproduce the $N_0(E)$ and $A_0(E)$ shapes expected from first principles (Sec. 2.5). Further, the observed energy dependence of R's statistical precision (Fig. 5.34b) closely follows the figure of merit $N_0 A_0^2$, which is anticipated via Eq. 2.30. Also note that the statistical precision of R below 1 GeV is minimal. Very little precision, therefore, would have been gained by lowering the energy threshold used in the asymmetry-weighted method in Sec. 5.7.1.

Moreover, systematic effects perturbing the positrons' energies should be most pronounced in the energy ranges where the spectrum has a steep gradient. Any deficiencies in the gain or pileup corrections are, therefore, expected to bias the fit result most severely in energy bins near the spectrum's endpoint. As shown in Fig. 5.35c, such a bias is evident in the best-fit κ_{loss}. While κ_{loss} should be energy independent, it is observed in every data set to (almost linearly) decrease with increasing energy by a statistically significant amount. Note that, since κ_{loss} is correlated with τ to around 70 %, a similar trend can also be identified in $\tau(E)$. As will be demonstrated in the next chapter, these observations can be explained by an uncorrected gain perturbation that is less than 10^{-3} at injection. This size of a gain change over the muon fill is also plausible, being near the limit of what the laser-calibration system can detect. The influence of a possible residual gain variation on R will be assessed and treated as a systematic uncertainty in Sec. 6.2.

Additionally, later bunches in each cycle of eight—illustrated in Fig. 2.2—rotate more times in the recycler synchrotron; thus, the phase-space distribution should vary bunch-to-bunch from nonlinearities in the radio-frequency manipulation. It is, therefore, important to look for any unexpected bunch dependence of the fit parameters. To that end, the data were divided by the bunch index in the pattern of eight and fit separately. The fit stop time used ranged from around 490 to
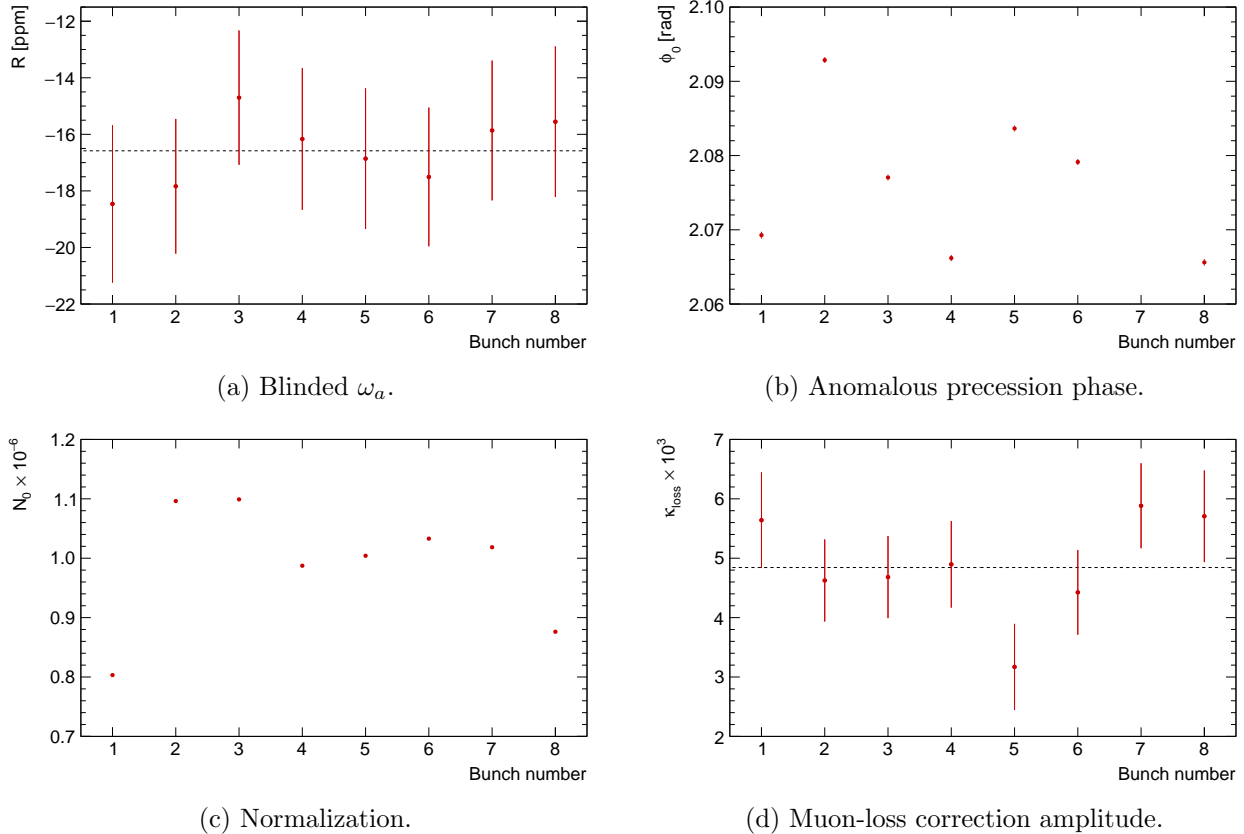
(a) Blinded ω_a.

(b) Anomalous precession phase.

(c) Normalization.

(d) Muon-loss correction amplitude.

Figure 5.36: The value of (a) R, (b) ϕ_0, (c) N_0, and (d) κ_{loss} for each muon bunch in the accelerator's cycle of eight, using the 1c data set with the baseline method. The dashed lines in (a) and (d) are the result of constant fits to their data.

640 μs, and each fit yielded a reduced-χ^2 consistent with one. For each data set, the best-fit R values are consistent among the bunches to less than 1.4 standard deviations. Each combined-R value, on the other hand, agrees with their nominal value in Tab. 5.6 to less than 1.2 standard deviations of the expected statistical drift.

However, as presented in Fig. 5.36, the best-fit ϕ_0 value varies significantly among the bunches. This is due to the different T0 pulse shapes used to define the time when each bunch was injected (Sec. 4.4). It is also observed that the first bunch in the cycle of eight has the lowest number of positron hits. This is plausible since, aside from variations in the accelerator, the kicker plates' peak voltages are known to be lower when the system has not fired recently, in turn reducing
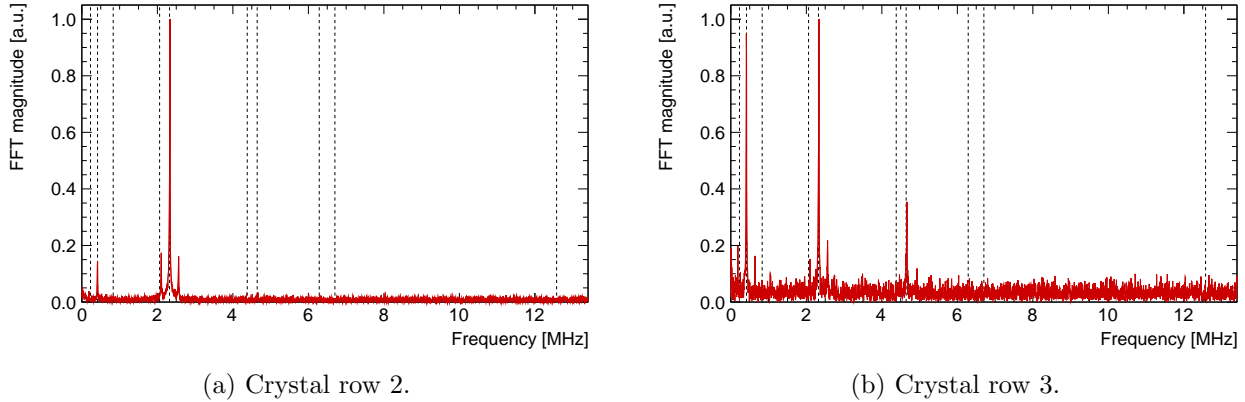
(a) Crystal row 2.

(b) Crystal row 3.

Figure 5.37: The FFT magnitude of the fit residuals for the crystal row (a) second and (b) third from the calorimeter bottom, using the five-parameter model in Eq. 5.3. These data are from the 1c data set with a 37.3-ns time-bin width. From low to high frequency, the dashed lines are at f_a, f_{CBO}, $2f_{CBO}$, f_{VW}, f_y, f_{VO}, $2f_y$, f_x, f_c, and $2f_x$.

the muon storage efficiency. Further, because the bunches vary slightly, one might also expect that their muon-loss spectra will differ. The shape of each bunch's $L(t)$, though, are remarkably similar—with only slight differences in the normalization, explaining the small variation in κ_{loss} versus the bunch number. All together, the bunch dependence observed for every fit parameter is as expected, boosting confidence in bunch-combined fit result.

For a final consistency check, the data were divided by the calorimeter's crystal row and by the calorimeter's crystal column. Each subset was then fit individually, revealing several muon-dynamics effects. In the crystal-row data, for instance, the vertical oscillation at f_y is significantly enhanced relative to the horizontal oscillation at f_{CBO}, as shown in Fig. 5.37. This is expected because the vertical movement is more comparable to the crystal height than to the calorimeter height. The fit results per crystal row are given in Figs. 5.38 and 5.39 for the 1c data set and are representative of the other data sets. As can be seen, the magnitude of the vertical oscillation at f_y—$A_{N,y,1,1}$—linearly increases the farther away the row is from the calorimeter's center. In the rows' fits where the f_y-oscillation is dominant, there tends to be a small, residual peak in their fit residuals' FFT near f_y, resulting in a reduced-χ^2 away from one by as many as four standard de-
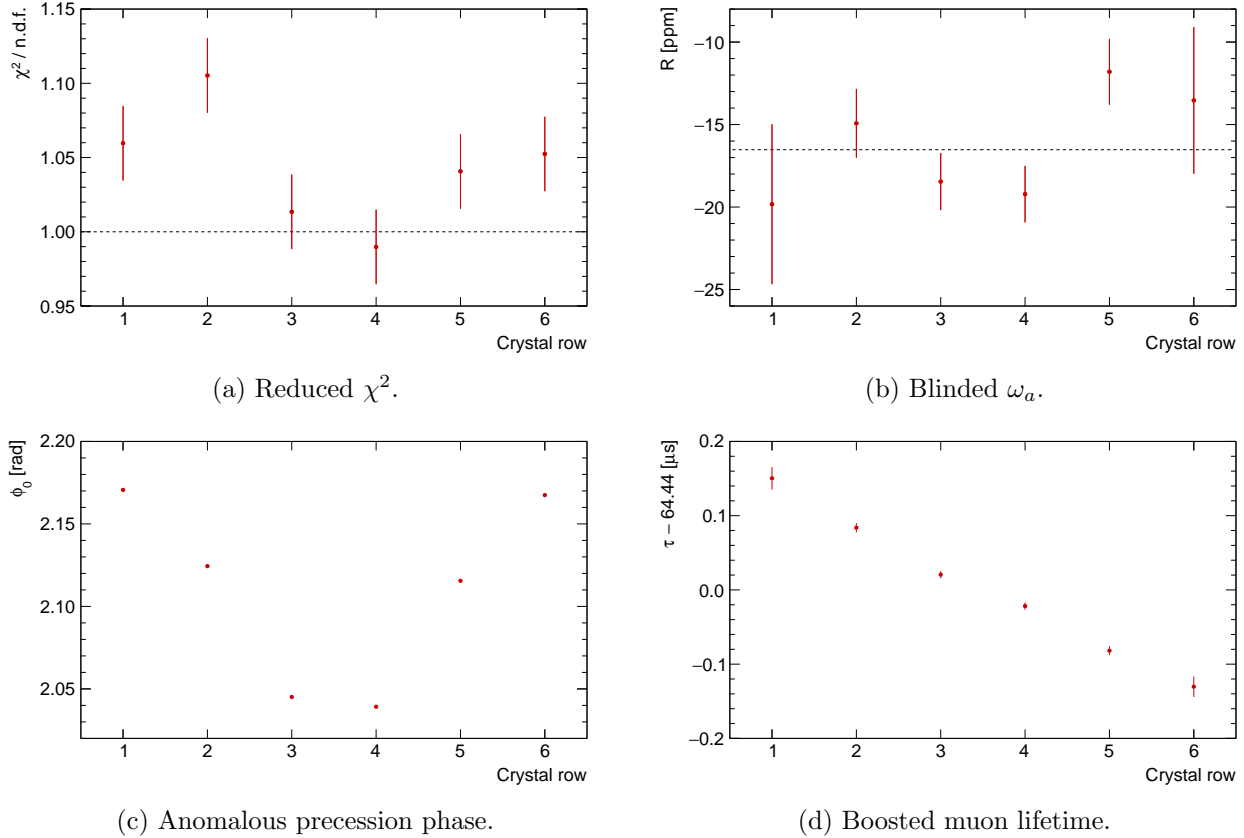
(a) Reduced χ^2.

(b) Blinded ω_a.

(c) Anomalous precession phase.

(d) Boosted muon lifetime.

Figure 5.38: The (a) reduced-χ^2, (b) R, (c) ϕ_0, and (d) τ for each crystal row, using the 1c data set with the baseline method. The dashed line in (b) is the result of a constant fit. For reference, the crystal rows are indexed from the bottom to the top of the calorimeter.

viations. These fits are most likely uncovering the limit of the ω_y-approximation in Eq. 5.42 and of the decoherence envelope used. However, since the related fit parameters are only weakly correlated with R, the combined-R value across the rows still agrees with the best-fit R value from the row-combined data to less than $2 \cdot \Delta\sigma$.

Nonetheless, it should be possible to use the individual rows' results to better estimate or constrain the vertical-betatron models in the row-combined fit. However, as seen in Fig. 5.39b, the best-fit τ_y appears to vary systematically across the crystal rows. This suggests that how the vertical dynamics should be modeled may be row-dependent, and, because of $A_{N,y,1,1}$'s sign, the row-combined fit may be particularly sensitive to such differences. This could be why τ_y in each

(a) Amplitude of μ_1-moment in $N_y(t)$.

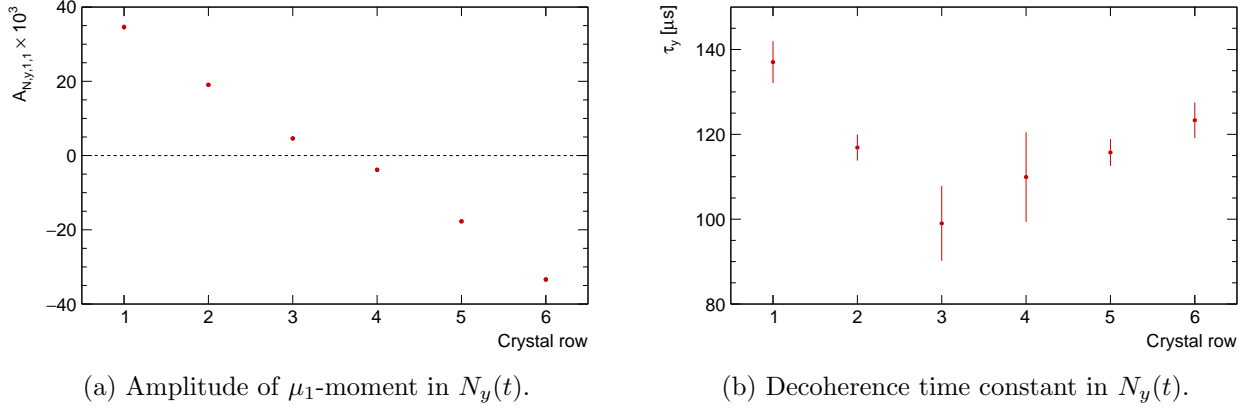(b) Decoherence time constant in $N_y(t)$.

Figure 5.39: The value of (a) $A_{N,y,1,1}$ and (b) τ_y for each crystal row, using the 1c data set with the baseline method. The rows are indexed from the bottom to the top of the calorimeter.

data set's row-combined fit is more than two standard deviations smaller than the typical τ_y in the crystal rows' fits—even when the $j = 1$ term in $\mu_2(t)$ of $N_y(t)$ is added. Therefore, without additional input from the tracker data, it is difficult to discern an improved model of the vertical dynamics. More importantly, though, since the vertical oscillations at f_y diminish by an order of magnitude when the crystal rows' data are combined, the change to the final results in Tab. 5.6 from a small inaccuracy in their frequency/decoherence model is likely negligible.

In addition, for each data set, the boosted muon lifetime, τ, is found to increase from the top to the bottom row of the calorimeter (Fig. 5.38d). This is caused by the coupling of two phenomena. First, lower-momentum muons in the storage ring will decay faster, increasing the mean momentum—and, thus, lifetime—over the fill. Second, as shown in Fig. 5.40, the mean vertical position of the positron drifts downward during the fit range between 30.2876 and 650.0644 μs. As a result, muons that survive longer will preferentially populate the lower row bins. Together, a roughly linear relationship between τ and the crystal row is expected. It must also be noted that ϕ_0 has a strong row dependency. Therefore, as the mean vertical position drifts over the fill, the best-fit ϕ_0 from the row-combined data will also drift, leading to a non-negligible systematic bias on the best-fit R due to their strong correlation (Sec. 6.7).
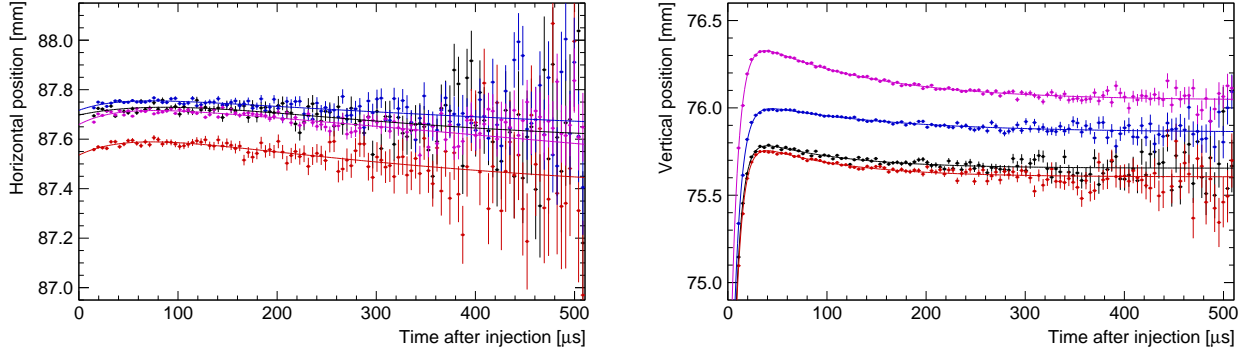
Figure 5.40: The mean horizontal (left) and vertical (right) position of positrons with reconstructed energies above 1.7 GeV over the fill for the 1a (black), 1b (red), 1c (blue), and 1d (magenta) data set. Lower horizontal and vertical positions are closer to the storage region and the calorimeter bottom, respectively. For each data set, a fit to the sum of two exponentials is overlaid. The positions' initial increase results from the quadrupole system's voltage transition from the scraping to nominal configuration (Sec. 2.4), while the later decrease results primarily from the quadrupole system's damaged/defective resistors.

The fit to each crystal column's data, on the other hand, was reasonable with a reduced-χ^2 less than 1.5 standard deviations away from one. Further, the best-fit R values are consistent among the columns to less than 1.3 standard deviations, and, for each data set, the combined-R value is within $1.2 \cdot \Delta\sigma$ of its nominal value. Moreover, the column dependence of τ and ϕ_0 reveals further muon dynamics. As the mean muon momentum increases over a fill, the decay positrons have a preferentially higher energy and, thus, a larger decay radius, resulting in more calorimeter impacts at columns closer to the storage region. However, with a stored momentum width of about 0.15 %, only a 25 ppm mean momentum shift is expected over the fill. This should lead to a τ-change less than 2 ns, which—as seen in Fig. 5.41—is undetectable given the data-set sizes.

Furthermore, the larger decay radii for higher-energy positrons leads to a longer drift time between when the muon decays and when the positron hits the calorimeter. The encoded spin-precession phase of the muon is, therefore, expected to be larger for columns closer to the storage region, which is indeed observed. The ϕ_0-change seen in Fig. 5.41, however, is about an order of magnitude larger than would be expected based on Fig. 2.10. This is because the quadrupole
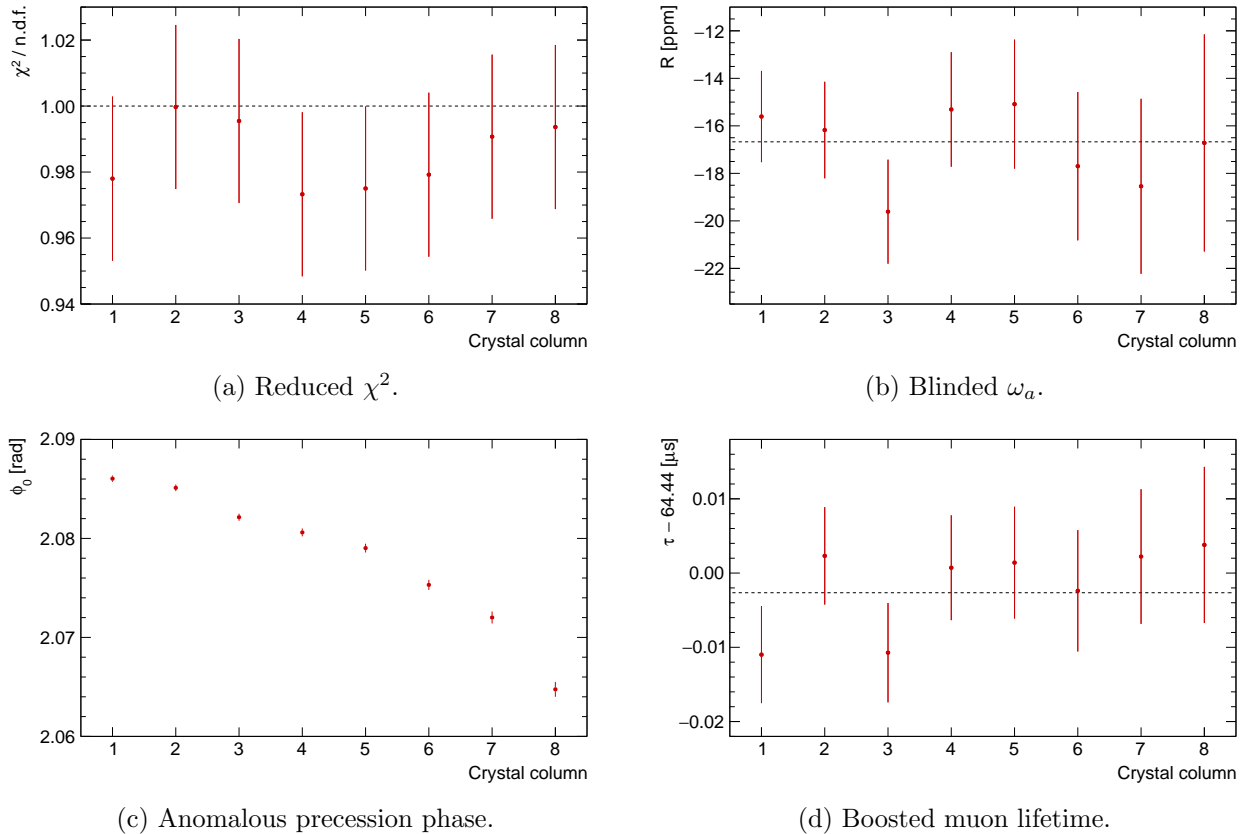
(a) Reduced χ^2.

(b) Blinded ω_a.

(c) Anomalous precession phase.

(d) Boosted muon lifetime.

Figure 5.41: The (a) reduced-χ^2, (b) R, (c) ϕ_0, and (d) τ for each crystal column for the 1c data set with the baseline method. The dashed lines in (b) and (c) are the result of constant fits. For reference, the columns are indexed from towards to away from the storage region.

system's damaged/defective resistors in Run 1 caused a steeper drift of the mean horizontal position over the fill—illustrated in Fig. 5.40—accentuating the change in ϕ_0. All together, except for the aforementioned possible residual gain effect, these numerous consistency checks presented did not identify any significant inadequacies in this chapter's measurement technique. The data collected in Run 1 are, thus, concluded to be well understood.

5.8 Outlook

As discussed in Sec. 2.7, the best-fit ω_a found by the above analysis must be slightly increased by the electric field and pitch corrections to obtain an unbiased measurement, from which a_μ can be

ultimately determined. The dedicated analyses to obtain these corrections, however, are still on-going for the Run 1 data sets as of this writing (Dec. 2019). The electric-field correction's analysis entails extracting the muons' equilibrium radius from calorimeter data, while the pitch correction's analysis involves extracting the muons' vertical distribution from tracker data. Preliminary results for some of the Run 1 data sets indicate that the electric field and pitch corrections will be about 480 and 160 ppb, respectively, with the uncertainties given in Sec. 6.9 [131, 132].

After these corrections are finalized—and the ω_a- and ω_p-measurements are unblinded—the value of a_μ can be determined via Eq. 2.4. Because the magnetic field's strength likely varied over Run 1, a_μ is expected to be found separately for each data set. Their weighted average will then be calculated to obtain the experiment's first measurement of the muon's anomalous magnetic moment. Assuming an ω_p-uncertainty of 140 ppb, the projected a_μ-uncertainty for Run 1 is 460 ppb—based on this dissertation's results using the asymmetry-weighted method in Tab. 5.6, including the next chapter's systematic uncertainty assessment and the systematic uncertainty estimates of separate corrections that need to be applied (Secs. 6.7.5, 6.8.3, and 6.9). Such an uncertainty would make this a_μ-measurement the world's most precise—about 80 ppb more precise than the previous experiment's value. Further, along with theoretical improvements, this a_μ-measurement alone may come close to resolving the current tension in its value (Ch. 1). In any case, as more data are collected and analyzed, the Muon $g - 2$ experiment will be setting an increasingly stringent constraint on new physics.

CHAPTER 6

SYSTEMATIC UNCERTAINTY ASSESSMENT

In the previous chapter, the muon's anomalous precession frequency, ω_a, was measured for each of the four data sets taken during the experiment's first running year, Run 1. In particular, the ω_a values were found by fitting time histograms of each data set's reconstructed positrons, which was performed using a χ^2-minimization approach. From the curvature of the χ^2, the statistical uncertainty on each best-fit ω_a was determined, leaving the *systematic* uncertainty—that not the result of stochastic fluctuations—to be evaluated separately. In this chapter, the data sets' ω_a-measurements are completed by estimating their systematic uncertainties. This amounts to assessing the techniques used, such as the reconstruction algorithm; the corrections made directly to the data; and the time-dependent phenomena modeled in the fit function. A summary of this chapter's assessment is shown in Tab. 6.1. Although the ω_a-measurements in Run 1 are statistics-limited, the techniques established here to characterize the many systematic uncertainties identified will become increasingly valuable when larger data sets become available. As more data are taken and more simulations are developed, some of these systematic uncertainties should also diminish as a better understanding of, for instance, the muon dynamics is enabled. After taking data for the four years budgeted, this reduction is expected to be enough to realize the experiment's systematic ω_a-precision goal of only 70 ppb.

6.1 Backend Electronics

As detailed in Sec. 5.2, this dissertation's ω_a-measurement is intentionally blinded in hardware to preclude any personal bias of the final result in developing the above analysis procedures. As a reminder, this hardware-level blinding was implemented by shifting the frequency of the backend

Uncertainty source	$w(E) = 1$				$w(E) = A(E)$			
	1a	1b	1c	1d	1a	1b	1c	1d
Digitization clock	2.1	2.1	2.1	2.1	2.1	2.1	2.1	2.1
Time correction	1.5	3.7	3.6	2.2	1.8	1.5	5.1	4.8
Energy calibration	11.2	22.4	8.6	9.9	7.6	18.1	20.5	9.9
Pileup	21.2	23.0	19.9	17.8	1.5	2.9	0.8	1.3
Muon loss	2.6	3.1	6.8	2.8	2.1	3.5	6.3	2.7
Muon dynamics	25.3	46.2	12.9	15.7	25.6	54.2	15.6	14.5

Table 6.1: A summary of the estimated ω_a-systematic uncertainties, in ppb, from the dominant sources for each data set in Run 1. The underlying uncertainties from energy calibration, pileup, muon loss, and muon dynamics are reported in Tabs. 6.4, 6.7, 6.9, and 6.10, respectively. The quadrature sum for each data set and weighting method is given in Tab. 5.6. Note that, as described in Secs. 6.7.5, 6.8.3, and 6.9, several corrections to the best-fit ω_a will be applied once their dedicated analyses are completed. In total, these corrections are expected to contribute an additional systematic uncertainty on ω_a of around 150 ppb.

electronics' input clock by ε. In the previous chapter, though, ε was neglected as it is small and has yet to be revealed to the collaboration. Further, as discussed in Ch. 3, this input clock at $f_{\text{input}} = (40 + \varepsilon)$ MHz is upconverted by every waveform digitizer's clock synthesizer to the sampling rate used to record the calorimeter signals. Therefore, when ε is eventually revealed, each data set's $\omega_a(R)$-measurement in Tab. 5.6—which assumed $\varepsilon = 0$—will need to be unblinded by converting from rad/c.t. to physical units of rad/s according to

$$\omega_a = \omega_a(R) \cdot (f_{\text{input}} \cdot \alpha) / (800 \text{ MHz}) \tag{6.1}$$

where the upconversion factor is $\alpha = 20$. As a result, the unblinded-ω_a's uncertainty will be enhanced due to that in the values of f_{input} and α.

After ε was set for Run 1, the central value of f_{input} was directly checked every week to be within 0.075 ppb of its frequency set point [119]. Further, its long-term stability was continuously monitored throughout Run 1 (Sec. 3.5), and, as seen in Fig. 3.20, the central value of f_{input} also remained within 0.075 ppb between the direct measurements. However, the ω_a-measurement

could still be biased if f_{input} fluctuated coherently across the fills. This was primary motivation for having the backend electronics' clocks be asynchronous to the accelerator's operation. As detailed in Sec. 3.5, such short-term stability was directly measured before the start of Run 1, and no dependence on the clock-distribution system's operations was found—as expected.

The upconversion factor, α, was also directly measured before the start of Run 1, as described in Sec. 3.3.3. For each of the 12 waveform digitizers tested, α is found to be equal to the nominal value of 20 to less than 2.1 ppb. Moreover, these α-measurements are reasonably described by a normal distribution about the nominal value. An estimate of α's uncertainty, therefore, could be taken as the measurements' standard deviation, 0.31 ppb. However, until more of the remaining 95 % of waveform digitizers are similarly tested, a conservative estimate of 2.1 ppb will be used. It should also be noted that such a large-scale characterization ought to be feasible between the experiment's future running periods with some improvements to the current clock-monitoring system. In either case, f_{input}'s conservative uncertainty of 0.075 ppb is negligible compared to α's uncertainty. The systematic uncertainty assigned to ω_a—after unblinding—arising from the backend electronics is, thus, 2.1 ppb for Run 1.

6.2 Reconstruction Algorithm

The reconstruction algorithm supplies the foundation for this dissertation's ω_a-measurement. As described in Ch. 4, there were many decisions made as to how the algorithm behaves that have reasonable alternatives. The omnipresent concern is whether any of these choices end up biasing the final result for ω_a, which is why the algorithm was continually characterized using a realistic simulation while being developed. These characterization results have mostly already been presented throughout Ch. 4 to motivate the algorithm's default configuration used. As an example from Sec. 4.8, no ω_a-bias was detected when the predicted time and energy biases are coupled

with a coherent betatron oscillation as large as 25 mm—twice that measured in any data set in Fig. 5.18. Notwithstanding, potential systematic biases arising from the reconstruction were assessed using data from Run 1. In this section, the reconstruction's time correction (Sec. 4.4), fit algorithm (Sec. 4.5), and noise mitigation (Sec. 4.9) will, in particular, be evaluated. Evaluation of the energy calibration and the artificial dead time are reserved for the following sections. As a reminder, the reconstruction algorithm's cluster-partition step was not used for this dissertation's ω_a-measurement and, thus, cannot bias the final result.

6.2.1 Time Correction

The reconstruction's fit algorithm is built around simultaneously fitting multiple crystals' pulses, which are modeled to peak in unison. Consequently, the time alignment among each calorimeter's crystals is crucial for such fitting to be sensible. For this reason, each crystal's trace was aligned relative to crystal 0, following the procedures described in Sec. 4.4. In short, the relative alignment was achieved for each crystal i by subtracting its trace's sample times by its sync pulse's time and by $\delta_i = \delta_i^{\text{shower}} - \delta_i^{\text{fiber}}$. Here, δ_i^{shower} and δ_i^{fiber} are the delays originating from the laser fiber's length and the electromagnetic shower, respectively. As the fiber lengths were not precisely measured prior to installation, δ_i had to be empirically estimated (relative to δ_0) by comparing the times of pulses in adjacent crystals. The method used to do this was found to be robust against the data set analyzed, but, as expected, the contribution of δ_i^{shower} to δ_i depended on the pulse-selection criteria. As seen in Fig. 4.7, δ_i^{shower} can be largely eliminated by only using pulses with amplitudes within 10 % of each other. If this pulse-selection is not used, δ_i is observed to systematically vary across each crystal row by up to 130 ps. The δ_i-difference between any two *adjacent* crystals, though, is less than 40 ps.

Varying amounts of δ_i^{shower} in δ_i will primarily distort the reconstructed clusters' times. This

distortion in turn changes the reconstruction's time bias—as demonstrated in Fig. 4.4. In the extreme case, a different set of δ_i values can even cause two clusters—instead of one—to be formed for a single positron hit. However, the δ_i-offsets are fixed for all islands, so an ω_a-bias can only arise from the time-dependent distribution of positron hits caused by muon dynamics (Sec. 2.6). Tracker measurements of the coherent betatron oscillation (CBO), for instance, indicate that the mean horizontal hit position varies by as much as half of a crystal width (c.w.) near 30 μs after injection [126]. Further, as seen in Fig. 5.40, the mean hit position drifts slowly over the fill up to 0.007 and 0.02 c.w. horizontally and vertically, respectively. To predict the impact on ω_a, Monte Carlo simulations of such effects were performed incorporating different position-dependent time biases. For time biases between adjacent crystals of $\mathcal{O}(10 \text{ ps})$ and $\mathcal{O}(100 \text{ ps})$, the corresponding ω_a-bias was found to be less than 1 and 3 ppb, respectively.

Although it should be small, as an ω_a-bias can occur, ω_a's systematic uncertainty due to the set of δ_i values used must be assessed. To that end, the change in R, ΔR, when instead using the δ_i-values found with *no* pulse-amplitude selection was estimated. In principle, to test a different set of δ_i values, the reconstruction algorithm must be run again over the data set, which is very computationally intensive. Note, however, that the cluster time will be dominated by the crystal with the largest energy deposit, i.e., the central crystal. Therefore, as a sufficient approximation, each reconstructed cluster's time was shifted by the δ_i-difference (Fig. 4.7) for its central crystal, resulting in the ΔR values given in Tab. 6.2. In this procedure, though, a statistical difference is introduced in the new histogram that is fit, predominately from counts moving across the first time bin's lower edge. This can lead to a deviation that is characterized by Eq. 5.44 and complicates how to measure the ΔR arising from the expected ω_a-bias. The ΔR observed for the 1a and 1b data sets, for example, are consistent with being caused by a change in statistics, while those for the 1c and 1d data sets significantly deviate from their allowed deviations. All of the ΔR values, in

Weighting method	Data set			
	1a	1b	1c	1d
$w(E) = 1$	1.5 ± 4.2	3.7 ± 4.0	-3.6 ± 1.3	-2.2 ± 0.5
$w(E) = A(E)$	-1.8 ± 3.4	1.5 ± 2.5	-5.1 ± 0.7	-4.8 ± 0.6

Table 6.2: The change in R, in ppb, with substitution of the relative time correction with one determined without a pulse-amplitude requirement (Sec. 4.4). The uncertainty quoted is the allowed deviation from the change in statistics, calculated via Eq. 5.44. Note that, when using the different time correction, the fits' χ^2 typically increased by around 3.

any case, are small relative to R's statistical uncertainty and, therefore, are taken as conservative estimates of the systematic uncertainty from the set of δ_i values used.

Furthermore, each crystal's trace was aligned to the time of muon injection. This calorimeter-dependent time correction was empirically found from the time of flight of lost muons between adjacent calorimeters (Sec. 4.4). Without precise alignment data available, the assumption was made that the calorimeters are equidistant, which is surveyed to be accurate to about 1 cm. As a result, each calorimeter's time alignment may be offset by up to 0.03 ns. This should slightly change the asymmetry and phase of the muon-dynamics oscillations as in Eq. 2.32. A potential ω_a-bias would have to come from a distortion in the oscillations' decoherence envelopes relative to the models used. However, as this was not found in the internal consistency checks in Sec. 5.7.3 (and as the envelopes' bias will be evaluated in Sec. 6.8), no systematic uncertainty on ω_a due to the calorimeter's absolute time alignment will be assigned.

6.2.2 Fit Algorithm

A cornerstone of the reconstruction's fit algorithm is that the spatial clustering is done together with the pulse fitting. It is, therefore, important to check whether such clustering causes a bias to ω_a—even though no bias mechanism is known. To that end, the algorithm's cluster window was enlarged from a 3×3 to a 5×5 crystal region, causing more clustering to occur. For the 1a data

set, doing so resulted in 5 % fewer clusters (before the pileup correction) and, using the baseline method, changed R by -201.4 ppb. The allowed deviation from different statistics, however, is estimated via Eq. 5.44 to be 436.3 ppb, indicating that the best-fit R using a 5×5 window size is statistically consistent with that using the nominal, 3×3 window size. As more pileup is made when using a larger window, this agreement also partially checks the pileup correction (Sec. 5.4). In any case, as a statistically significant ω_a-bias was neither found nor expected, no systematic uncertainty is assigned to the fit algorithm's clustering method.

6.2.3 Noise Mitigation

As detailed in Sec 4.9, the reconstruction is robust against the noise signals observed as single-frequency wave packets. It is less robust, though, against the noise signals caused by the periodic temperature polling of the silicon photomultipliers (SiPMs). This is because they are closer in shape to the positron-pulse template, typically yielding clusters above the fit algorithm's energy threshold but—as seen in Fig. 4.41—no higher than 500 MeV. This observation was the primary reason for choosing the asymmetry-weighted method's lower energy threshold to be 1 GeV—well above the noise's endpoint. As a result, these noise clusters should not enter the constructed his-tograms, and, thus, no *direct* bias on the extracted R is expected. This is supported by the results from fits to data in energy ranges as low as 500 MeV, which yielded reasonable R and reduced-χ^2 values (Sec. 5.7). Moreover, in each data set's nominal fits, if a constant term is added to the model in Eq. 5.6, its best-fit value is consistent with zero to 1.6 standard deviations. This lack of a flat background also provides evidence that noise does not contaminate the fitted data.

Noise clusters will, however, pile up with those from positrons, but Monte Carlo simulations indicate that they pose a negligible net time/energy bias (Sec. 4.9). In any case, the SiPM's tem-perature polling is asynchronous to the accelerator's operation and in turn the fill—as verified in

Weighting method	Data set			
	1a	1b	1c	1d
$w(E) = 1$	268.6 ± 215.2	-299.0 ± 210.1	-80.0 ± 131.9	-114.0 ± 94.2
$w(E) = A(E)$	207.6 ± 159.2	-222.5 ± 189.7	-228.8 ± 119.1	-103.2 ± 85.1

Table 6.3: The change in R, in ppb, when only analyzing fills without clusters flagged by the reconstruction as noise (Sec. 4.9). The uncertainty quoted is the allowed deviation from the change in statistics, calculated via Eq. 5.44. Note that, for some of these fits, the stop time was changed to 633.056 μs to ensure that at least 30 counts in every time bin used in the χ^2.

the data (Fig. 4.41). This means that even any unexpected bias to the cluster times should be constant over the fill and, therefore, not introduce a bias mechanism to ω_a. Further, the pileup correction should partially correct for any such bias because noise and positron pulses will also be combined in its construction (Sec. 5.4). To check for a hidden bias anyway, any fill with a cluster flagged as noise by the reconstruction (Sec. 4.9) was vetoed, as a test, from the ω_a-measurement. For each data set, doing so removed about 1.7 % of the fills nominally used, resulting in the change in R given in Tab. 6.3. Estimating the allowed deviation via Eq. 5.44, the best-fit R values with and without such fills are consistent with expectation to less than 1.5 standard deviations. This suggests that there is indeed no hidden bias, and, for that reason, no systematic uncertainty on ω_a from noise clusters is assigned to any of the data sets.

6.3   Energy Calibration

As part of the reconstruction, each crystal–SiPM pair's energy scale is calibrated relative to one another. After the baseline calibration, to avoid a systematic bias to ω_a, the coherent variations across the muon fills in each SiPM's gain response are corrected; otherwise, the ω_a-measurement energy threshold is effectively time-dependent. More specifically, two effects—in-fill and short-term double-pulse—are corrected, which correspond to distinct physical phenomena. As discussed

Uncertainty source	$w(E) = 1$				$w(E) = A(E)$			
	1a	1b	1c	1d	1a	1b	1c	1d
In-fill gain amplitude	3.8	8.8	3.9	3.6	3.9	4.4	3.0	3.5
In-fill gain time constant	10.5	15.6	4.5	7.5	4.7	2.4	5.1	5.7
STDP gain amplitude	0.0	0.1	0.1	0.2	0.2	0.1	0.2	0.9
STDP gain time constant	0.2	1.0	0.5	0.6	0.6	0.4	0.7	0.8
Residual gain variation	0.7	13.5	6.1	5.4	4.4	17.4	19.6	7.3

Table 6.4: A summary of the ω_a-systematic uncertainties, in ppb, from the energy calibration for each data set in Run 1. The combined uncertainties are shown in Tab. 6.1.

in Sec. 4.6, each correction is applied at the crystal level using an analytical function of a time separation, whose parameters are found from a fit to measurements using the laser-calibration system. The exponential shape of each model used—given in Eqs. 4.33 and 4.34—is motivated by the resistor-capacitor circuits governing the gain recovery and, as seen in Fig. 4.28, describes the data well. Further, a fast Fourier transform (FFT) of the fit residuals does not reveal any significant peaks. As a result of such good agreement, in this section, ω_a's systematic uncertainty from each gain correction will be estimated based on its model-parameter uncertainties—with no other models assessed. Afterward, an ω_a-bias from a possible gain variation remaining after the reconstruction's gain corrections will be evaluated. A summary of the individual systematic uncertainties assigned to the energy calibration is provided in Tab. 6.4.

6.3.1 In-Fill Gain Correction

The in-fill effect describes the SiPM's bias-voltage recovery caused by the initially large flux of particles hitting the calorimeters at each fill's onset. To measure each crystal–SiPM pair's exact recovery, the laser was occasionally pulsed at different times after injection, allowing for the level of pulse-amplitude reduction to be mapped as a function of time. The model used to fit these laser data is given in Eq. 4.33, where the free parameters are the characteristic amplitude A^{IF}

Quantity	$w(E) = 1$				$w(E) = A(E)$			
	1a	1b	1c	1d	1a	1b	1c	1d
$\sigma_{A^{\mathrm{IF}}}$	21.4	17.2	15.3	11.0	21.3	17.2	15.3	11.0
$\sigma_{\tau^{\mathrm{IF}}}$	16.1	13.2	11.5	7.5	16.1	13.2	11.5	7.4
$\rho_{A^{\mathrm{IF}},\tau^{\mathrm{IF}}}$	-78.1	-81.5	-87.4	-87.6	-78.2	-81.5	-87.4	-87.7
$\sigma_{A^{\mathrm{STDP}}}$	1.9	1.9	1.9	1.9	1.9	1.9	1.9	1.9
$\sigma_{\tau^{\mathrm{STDP}}}$	3.4	3.4	3.4	3.4	3.4	3.4	3.4	3.4

Table 6.5: The average uncertainties (σ) and correlation coefficients (ρ), in percent, among the crystal–SiPM pairs, weighted by the number of counts each pair contributed to the baseline or asymmetry-weighted method's histogram. Here, A^{IF} and τ^{IF} are the parameters in Eq. 4.33, describing the in-fill gain effect; A^{STDP} and τ^{STDP} are the parameters in Eq. 4.34, describing the short-term double-pulse gain effect. Note that A^{STDP} and τ^{STDP} are taken to be uncorrelated.

and time constant τ^{IF}. As seen in Tab 6.5, the uncertainties on the best-fit values of A^{IF} and τ^{IF} are around 16 and 12 %, respectively. This in turn creates an uncertainty on how to stabilize the energy scale over the fill, posing a systematic uncertainty to the R extracted.

To assess R's sensitivity to A^{IF} and τ^{IF} separately, each crystal–SiPM pair's best-fit value of A^{IF} or τ^{IF} was artificially scaled, after which the ω_a-measurement was repeated. As illustrated in Fig. 6.1 for the 1c data set, R is approximately linearly biased as A^{IF} is scaled. The substructure found mostly in the baseline method results from the statistical fluctuations expected from varying the number of counts in the fitted histogram. A larger value of A^{IF}, for example, only increases the number of positrons above the ω_a-measurement's lower energy threshold. When weighting by the asymmetry, though, the change in counts is smaller since lower-energy positrons are weighted less (Fig. 5.27). To separate such statistical fluctuations from the systematic bias, the change in R for all of the artificial multipliers sampled from zero to two was linearly fit. As seen in Tab. 6.6, the resulting R-sensitivity is comparable between the weighting methods for three of the four data sets. For each ω_a-measurement done, its sensitivity is then multiplied by A^{IF}'s uncertainty in Tab. 6.5 to estimate the corresponding systematic uncertainty on R.
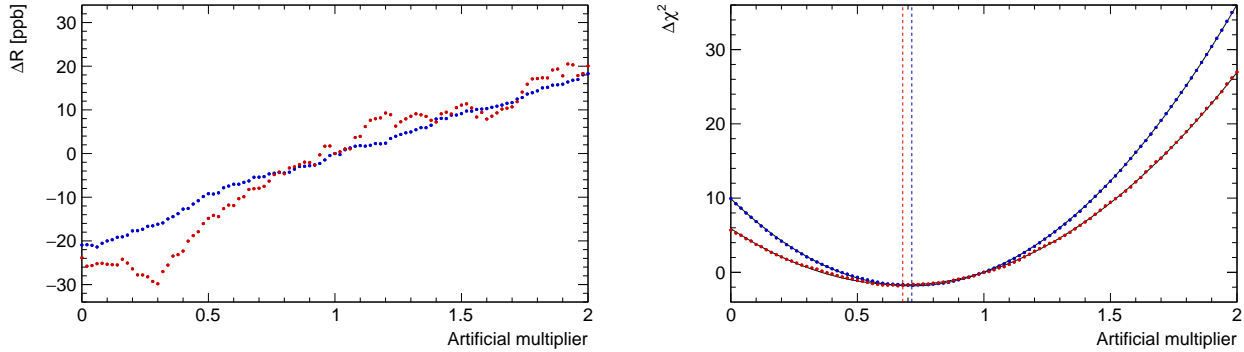
Figure 6.1: The change in R (left) and χ^2 (right) when artificially scaling the in-fill gain correction's amplitude parameter, A^{IF}, for all crystal–SiPM pairs in the 1c data set, using the baseline (red) and asymmetry-weighted (blue) method. A unit multiplier corresponds to the nominal A^{IF} values. The vertical, dashed lines denote the minima of parabolic fits to the changes in χ^2.

Among the data sets, the largest systematic uncertainty from A^{IF} is only 8.8 ppb—as seen in Tab. 6.4. This small uncertainty is partially expected because, as illustrated in Fig. 4.28a, the gain has largely recovered by the fit's start time of 30.2876 μs. However, as shown in Fig. 6.2, R's sensitivity to A^{IF} can increase by an order of magnitude when starting the fit slightly earlier or later. In general, Monte Carlo studies have shown that, for any uncorrected slow effect, R's sensitivity will be minimized when the fit starts on a node of the ω_a-oscillation [120]. This was the motivation behind choosing the specific start time nominally used (Sec. 5.7). When the fit starts one time bin earlier, for example, the statistical precision gained is outweighed by the systematic precision lost. In contrast, R's sensitivity to A^{IF} is relatively insensitive to the fit's stop time—as also shown in Fig. 6.2.

Further, as illustrated in Fig. 6.3, R is only linearly biased when τ^{IF} is scaled by more than around 0.7. In general, the nonlinear behavior found as τ^{IF} approaches zero is expected because the exponential model's value in the fit range becomes effectively removed. To estimate ω_a's systematic uncertainty from τ^{IF}, R's sensitivity to τ^{IF} was approximated by a linear fit to the change in R for artificial multipliers between 0.7 and 1.3—a range larger than τ^{IF}'s uncertainty in Tab. 6.5. For the 1c data set's asymmetry-weighted method, though, the linear fit was only per-

Sensitivity		$w(E) = 1$				$w(E) = A(E)$			
		1a	1b	1c	1d	1a	1b	1c	1d
dR/dA^{IF}	[ppb]	18.0	51.1	25.5	32.5	18.4	25.8	19.7	31.7
dR/dA^{STDP}	[ppb]	1.1	7.5	4.9	8.7	8.9	3.9	8.3	9.9
$dR/d\delta G$	[ppm]	$-$ 2.4	-41.3	-15.0	11.9	-15.2	-53.3	-48.3	-16.3

Table 6.6: The R-sensitivity per unit scale of the in-fill, short-term double-pulse, and residual gain correction's amplitude parameters—A^{IF}, A^{STDP}, and δG, respectively—for each data set.

formed for multipliers between 0.7 and 1.0. This one exception is needed because R's sensitivity changes sign when the best-fit τ^{IF} is increased—which is atypical compared R's sensitivity for the other data sets. As done for A^{IF}, these sensitivities are then multiplied by τ^{IF}'s mean uncertainties to find the corresponding systematic uncertainties on R (Tab. 6.4).

To properly combine ω_a's systematic uncertainty from A^{IF} and τ^{IF}, their correlation must be considered. As seen in Tab. 6.5, A^{IF} and τ^{IF} are estimated to be correlated by around -84 %. It is, therefore, conservative to assume that their resulting systematic uncertainties are uncorrelated. For that reason, ω_a's systematic uncertainties from A^{IF} and τ^{IF} are added in quadrature to calculate the total uncertainty. Moreover, note that the individual systematic uncertainties are also conservative estimates because A^{IF} or τ^{IF} are tacitly assumed to be correlated among all crystal–SiPM pairs. A closer total estimate, however, is unnecessary when it is compared to the size of each data set's statistical uncertainty. Notwithstanding, as reported in Tab. 6.1, the conservative approaches used across this section yield a total systematic uncertainty from the energy calibration that already reaches the experiment's target precision of 20 ppb.

6.3.2 Short-Term Double-Pulse Gain Correction

The short-term double-pulse (STDP) effect describes the recovery of the SiPM's pixels after being fired due to a positron hit. To measure each crystal–SiPM pair's exact recovery, the laser
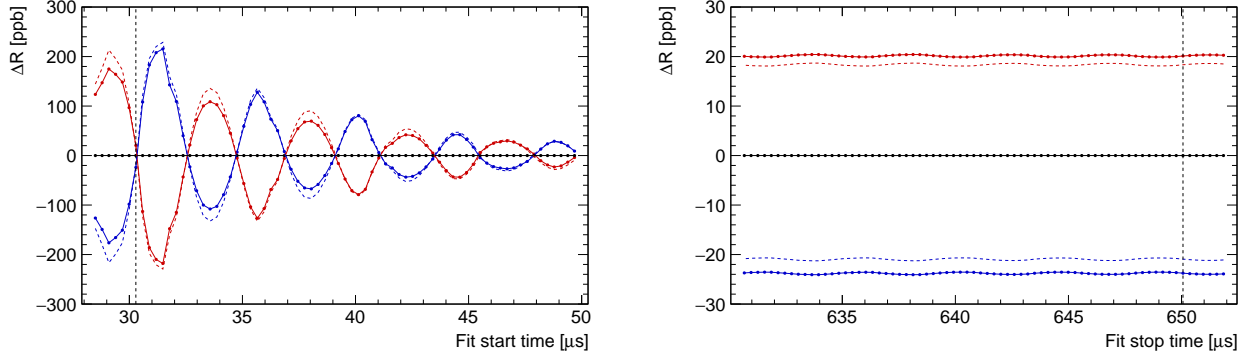
Figure 6.2: A scan over the fit start (left) and stop (right) time when artificially scaling by 0 (blue), 1 (black), and 2 (red) the in-fill gain correction's amplitude parameter, A^{IF}, for all crystal–SiPM pairs in the 1c data set, using baseline (solid) and asymmetry-weighted (dashed) method. The change in R is with respect to that when using the nominal A^{IF} values. The nominal fit start and stop times are denoted by vertical, dashed lines.
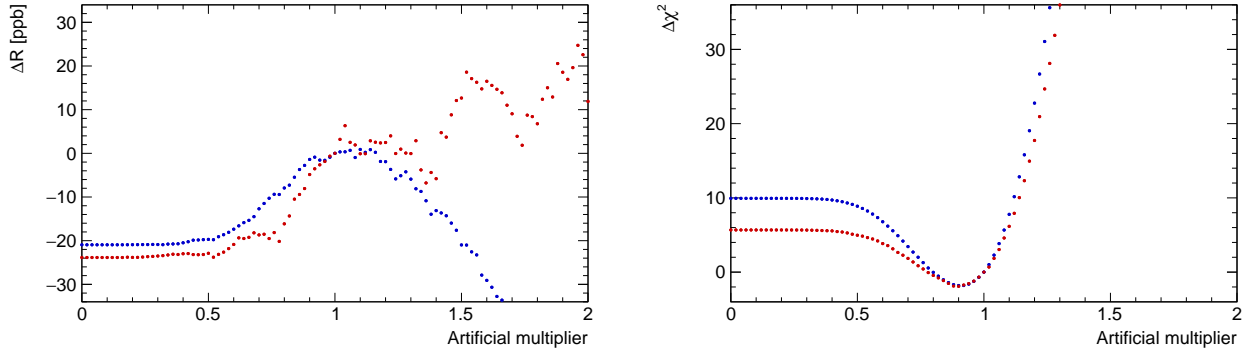

Figure 6.3: The change in R (left) and χ^2 (right) when artificially scaling the in-fill gain correction's timescale parameter, τ^{IF}, for all crystal–SiPM pairs in the 1c data set, using the baseline (red) and asymmetry-weighted (blue) method. For reference, an artificial multiplier of one corresponds to using the nominal τ^{IF} values.

was pulsed twice at varying time separations and intensities. As described in Sec. 4.6, this allowed the second pulse's attenuation to be mapped as a function of their time separation and the first pulse's reconstructed amplitude. The model used to correct for this STDP effect is given in Eq. 4.34, where the parameters are the amplitude A^{STDP} and time constant τ^{STDP}. These parameters' values were obtained from separate fits and will be treated as being uncorrelated. As seen in Tab. 6.5, the uncertainties on the best-fit values of A^{STDP} and τ^{STDP} are around 2 and 3 %, respectively. While much smaller than those in the in-fill gain correction, these uncertainty will,
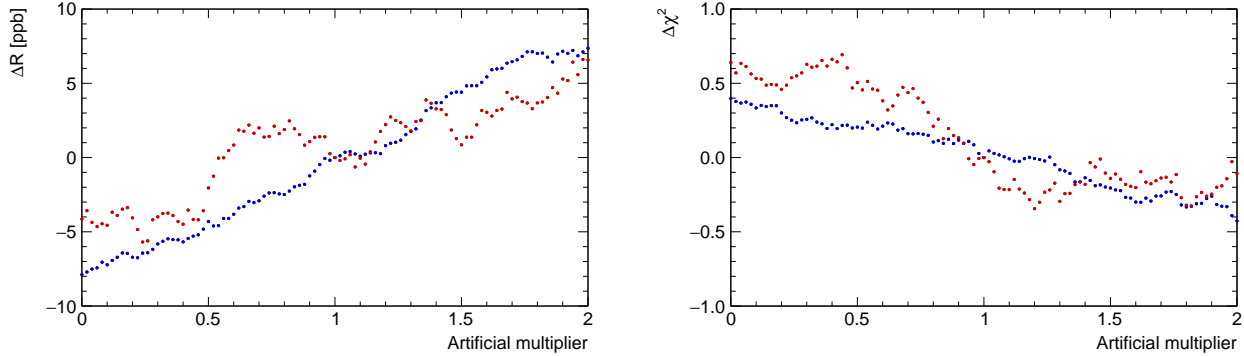
Figure 6.4: The change in R (left) and χ^2 (right) when artificially scaling the short-term double-pulse gain correction's amplitude parameter, A^{STDP}, for all crystal–SiPM pairs in the 1c data set, using the baseline (red) and asymmetry-weighted (blue) method. For reference, an artificial multiplier of one corresponds to using the nominal A^{STDP} values.

nonetheless, introduce a systematic uncertainty on the R extracted.

The systematic uncertainty on R from A^{STDP} and τ^{STDP} was estimated following the same procedure described above for that from A^{IF} and τ^{IF}, respectively. The only exception is that the 1c data set's asymmetry-weighted method did not need to be specially treated when assessing R's sensitivity to τ^{STDP}'s artificial scale. A summary of the uncertainty assessment for each data set and weighting method is given in Tab. 6.4. As representative examples, R's sensitivity to scaling A^{STDP} and τ^{STDP} is shown for the 1c data set in Figs. 6.4–6.6. As also shown for all data sets in Tab. 6.6, R is an order of magnitude less sensitive to the STDP's than the in-fill gain correction's parameters. In any case, R's sensitivity to A^{STDP} can be reduced by starting the fit slightly earlier, but, as shown in Figs. 6.2 and 6.5, the precision that would be gained is much less than the precision that would be lost from R's sensitivity to A^{IF}. In any case, with the nominal start time, the STDP effect only adds 1.3 ppb (at most) to ω_a's systematic uncertainty.

6.3.3   Residual Gain Variation

After applying the reconstruction's gain corrections, any residual gain variation over the fill will yield an energy-dependent bias of the fit result. As discussed in Sec. 5.7.3, such a bias has been
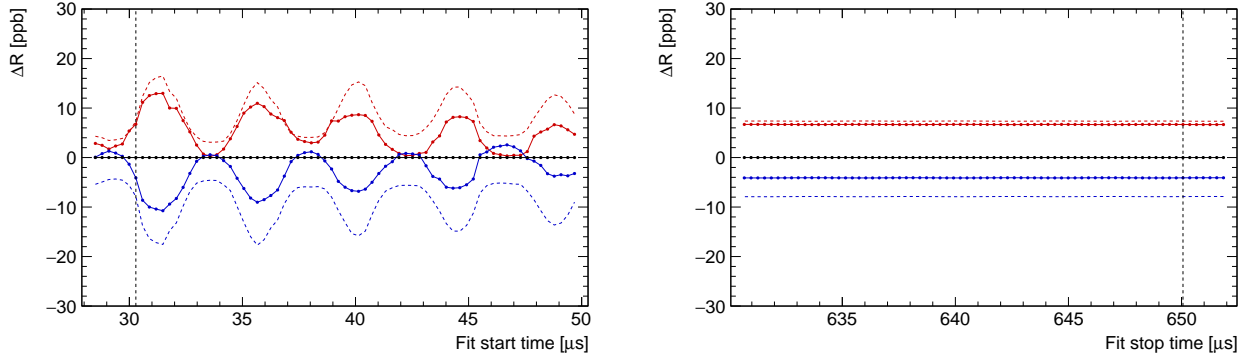
Figure 6.5: A scan over the fit start (left) and stop (right) time when artificially scaling by 0 (blue), 1 (black), and 2 (red) the short-term double-pulse gain correction's amplitude parameter, A^{STDP}, for all crystal–SiPM pairs in the 1c data set, using the baseline (solid) and asymmetry-weighted (dashed) method. The change in R is with respect to that when using the nominal A^{STDP} values. The nominal fit start and stop times are denoted by vertical, dashed lines.
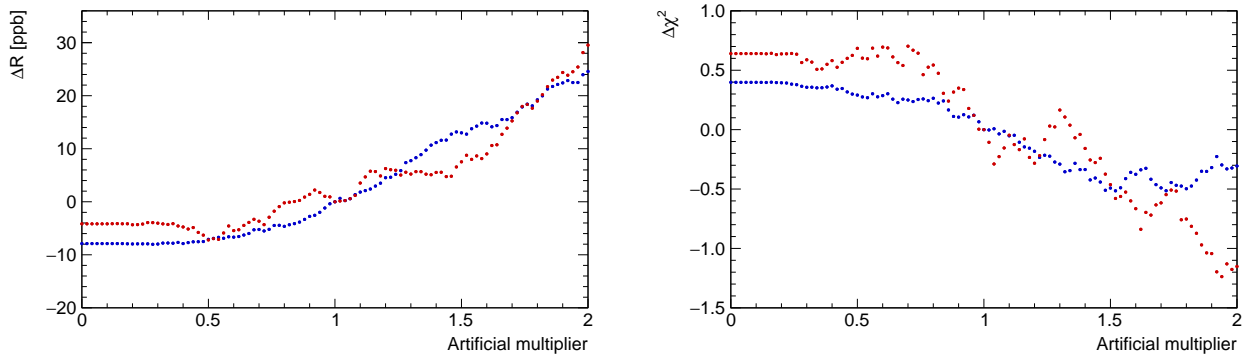

Figure 6.6: The change in R (left) and χ^2 (right) when artificially scaling the short-term double-pulse gain correction's timescale parameter, τ^{STDP}, for all crystal–SiPM pairs in the 1c data set, using the baseline (red) and asymmetry-weighted (blue) method. For reference, an artificial multiplier of one corresponds to using the nominal τ^{STDP} values.

identified in the best-fit κ_{loss} for every data set. While it should be independent of energy, κ_{loss} decreases with energy—even becoming negative in the 1b and 1c data sets, which is unphysical. This provides an indirect way to assess the possible inaccuracies of the in-fill and/or STDP gain correction—and their influence on R. It must be noted, though, that roughly a 10 % reduction in the pileup correction applied can also resolve κ_{loss}'s energy dependence. In general, κ_{loss} could be compensating for any uncorrected effect that varies slowly over the fill with a time dependence different than the muon loss. However, such a large inaccuracy in the pileup correction applied
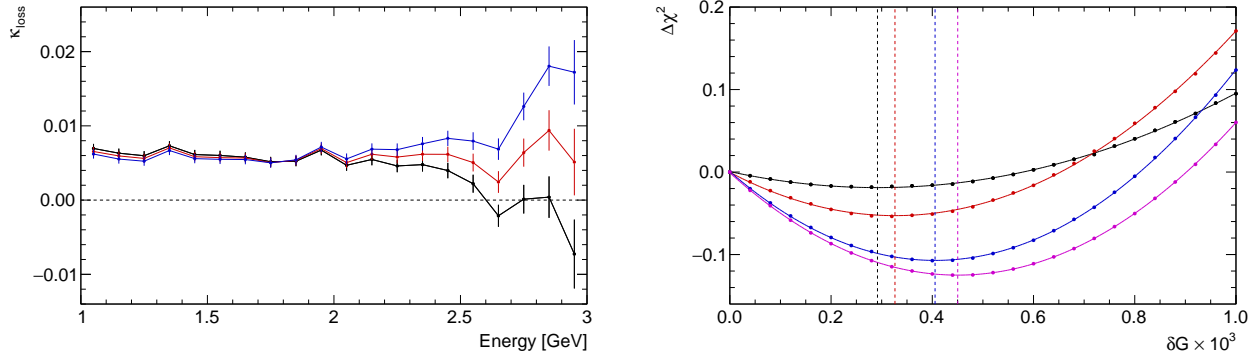
Figure 6.7: Left: The best-fit κ_{loss} for different energy ranges in the 1c data set after correcting for the residual gain variation with a δG value of 0 (black), 4×10^{-4} (red), and 8×10^{-4} (blue). Right: The change in χ^2 of a constant fit to $\kappa_{\mathrm{loss}}(E)$ when applying the residual gain correction with increasing δG values on the 1a (black), 1b (red), 1c (blue), and 1d (magenta) data set. The δG values at the minima of parabolic fits to $\chi^2(\delta G)$ are denoted by vertical, dashed lines.

is unreasonable, as will be demonstrated in Sec. 6.5.

Assuming a residual gain variation exists, it will most likely be proportional to the rate of positron hits at a given crystal, where some voltage is trying to be replenished in the calorimeter system. A more rigorous analysis suggests that the gain will vary as

$$G(t) = \delta G \cdot e^{-t/\tau} \left[1 + 0.22 \cdot \cos(\omega_a(R) \cdot t + \phi_0) \right] \tag{6.2}$$

up to the small perturbations from the muon dynamics [91]. Here, the asymmetry of 0.22 is the energy-weighted value that is expected from Sec. 2.5, and δG sets the residual gain variation's scale at injection. In addition, the values of τ, R, and ϕ_0 can be approximated by those from the nominal fit result in Sec. 5.7.

To remove this residual gain variation on average, each cluster's energy, E, can be replaced by $E(1 + G(t))$. As illustrated in Fig. 6.7, using this correction with δG on the order of only 10^{-4} is enough to have the best-fit κ_{loss} be energy-independent as expected. At the same time, this residual gain correction resolves the poor fit start time scans of N_0, τ, and κ_{loss} found in the 1d data set (Sec. 5.7.2). Furthermore, a gain perturbation as large as $5 \cdot 10^{-4}$ at injection only corresponds to that around $3 \cdot 10^{-4}$ when the fit starts at $30.2876\ \mu$s. This size of a gain variation
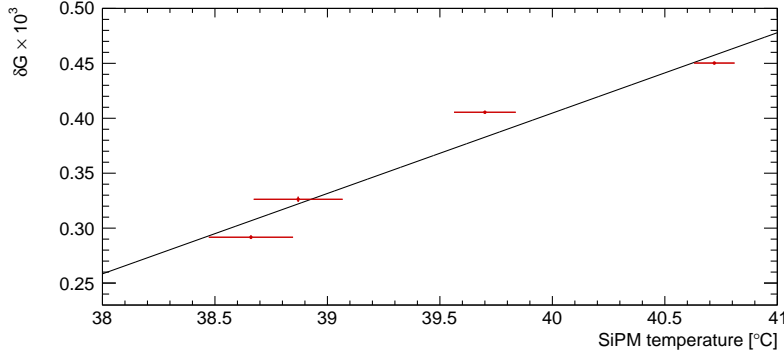
Figure 6.8: The optimal value of δG—which makes κ_{loss} energy-independent—against the mean SiPM temperature over the data set. From left to right, the points correspond to the 1a, 1b, 1c, and 1d data sets. A linear fit is overlaid, which has a slope of $0.0732(59)/^\circ\text{C}$.

is below the statistical uncertainty of the laser measurements [38]. An uncorrected, residual gain variation in the data is, therefore, plausible and, as such, is considered to be responsible for the $\kappa_{\text{loss}}(E)$ instability observed across the data sets.

As a result, the size of residual gain variation in each data set can be deduced by observing $\kappa_{\text{loss}}(E)$ for varying values of δG. To that end, for each data set, the above residual gain correction was applied using increasing values of δG up to 10^{-3}. For each trial value of δG, the best-fit value of κ_{loss} was found in 100-MeV energy bins from 1 to 3 GeV—as done in Sec. 5.7.3—and fit to a constant. As illustrated in Fig. 6.7, the constant fit's χ^2 for each value of δG was fit to a parabola. The value of δG at the parabolic fit's minimum—defined to be the optimal value of δG—is 2.9, 3.3, 4.1, 4.5 for the 1a, 1b, 1c, and 1d data set, respectively, in units of 10^{-4}.

Notice that the optimal value of δG increases with the time when the data set was taken. This could be related to the experimental hall's mean temperature, which increased over Run 1. As seen in Fig. 6.8, the optimal value of δG is roughly linear with the mean SiPM temperature over each data set. The residual gain variation, therefore, may originate from a deficiency in the STDP gain correction because the laser data to obtain its parameters were taken after Run 1. It has been recently measured that A^{STDP} increases with temperature by around 5 %/$^\circ$C on average,
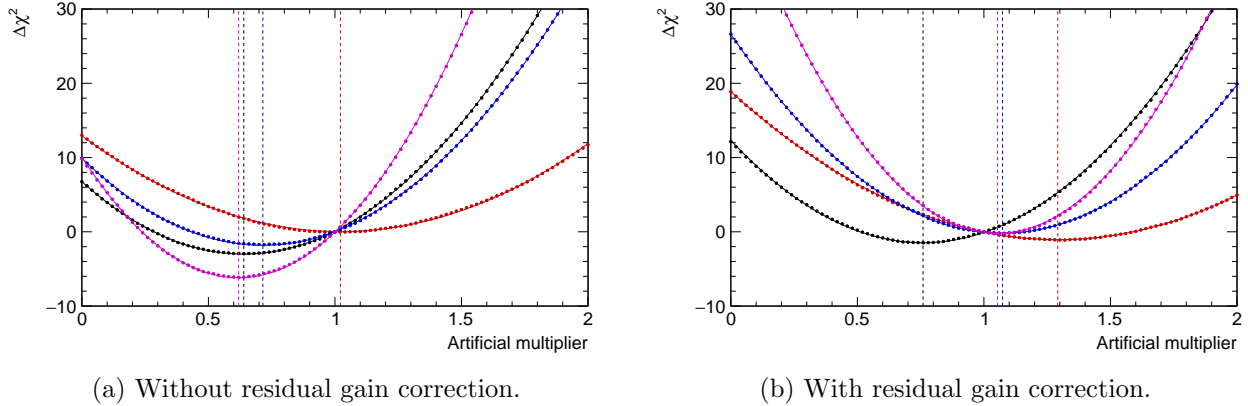
(a) Without residual gain correction.

(b) With residual gain correction.

Figure 6.9: The change in χ^2 when artificially scaling the in-fill gain correction's amplitude parameter, A^{IF}, for all crystal–SiPM pairs in the 1a (black), 1b (red), 1c (blue), and 1d (magenta) data set, using the asymmetry-weighted method, and when (a) applying and (b) not applying the residual gain correction. An artificial multiplier of one corresponds to the nominal correction. The artificial multiplier minimizing the χ^2 is denoted by a dashed line for each data set.

which is unaccounted for in the STDP gain correction used in this dissertation [133]. How best to resolve this temperature dependence of A^{STDP} is still being investigated, and it is unclear so far whether $\kappa_{\mathrm{loss}}(E)$ will be completely stabilized as a result.

Also note that, as depicted in Figs. 6.1 and 6.9a, A^{IF}'s artificial multiplier that minimizes the ω_a-fit's χ^2 is 1.3–2.5 standard deviations below the nominal value of one for most of the data sets, which all agree on an optimal scale around 0.64. However, as seen in Fig. 6.10, the artificial multiplier minimizing the χ^2 linearly scales with the value of δG in the residual gain correction above. When the optimal values of δG are used, as shown in Fig. 6.9b, the χ^2 minima among the data sets are shifted to be statistically consistent with an artificial multiplier of one. This observation further supports that a residual gain variation exists in the reconstructed data. If such variation originates from the STDP effect, as considered above, the A^{STDP} scale that minimizes the χ^2 of the ω_a fits should also be affected. As seen in Fig. 6.4, however, the χ^2 is too insensitive to discern whether it prefers a residual gain correction or not. Notwithstanding, any deficiency in the STDP gain correction will alter the in-fill gain correction at some level because the STDP gain correction
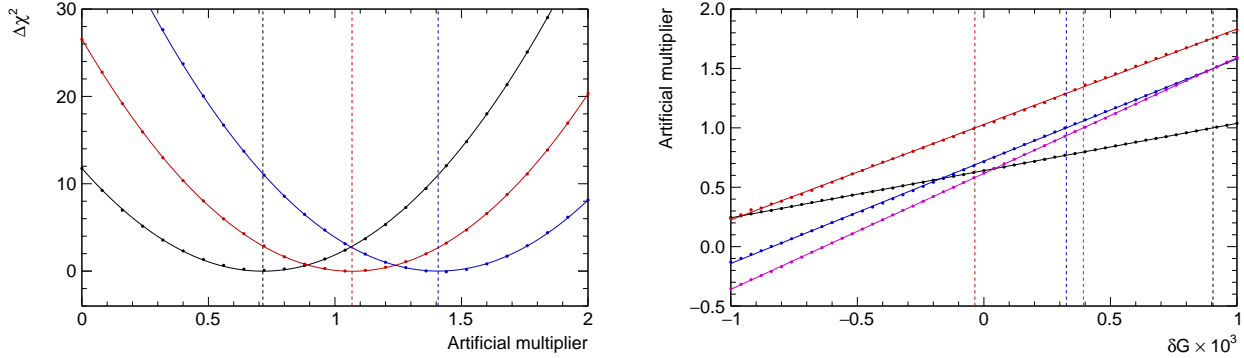
Figure 6.10: Left: The change in χ^2 when artificially scaling the in-fill gain correction's amplitude parameter, A^{IF}, for all crystal–SiPM pairs in the 1c data set, using the asymmetry-weighted method, when applying the residual gain correction with a δG of 0 (black), 4×10^{-4} (red), and 8×10^{-4} (blue). The artificial multiplier minimizing the χ^2 is denoted by a dashed line for each value of δG. Right: The artificial multiplier of A^{IF} minimizing the χ^2 when applying the residual gain correction with different δG values on the 1a (black), 1b (red), 1c (blue), and 1d (magenta) data set. The δG values causing the χ^2 minima to be at one are denoted by dashed lines.

is also applied to the in-fill laser pulses that are used (Sec. 4.6).

All together, a residual gain variation approximately following the functional form in Eq. 6.2 is well motivated and is able to explain the unexpected features identified in Ch. 5. However, for the purpose of this dissertation, the possibility of a residual gain variation will be treated only as a systematic uncertainty on R—i.e., with no correction made to the best-fit R. Without the physical mechanism conclusively identified, it would otherwise be difficult to confidently estimate the uncertainty on the residual gain correction. To that end, the change in R was found for δG up to 10^{-3}, using the average residual gain correction above. As shown in Fig. 6.11, R's sensitivity to δG is approximately linear. To separate the statistical fluctuations from the systematic bias, the change in R was linearly fit to δG. The systematic uncertainty on R from the residual gain variation is then estimated as the linear fit's slope (Tab. 6.6) multiplied by the optimal δG value found above. For some data sets, as seen in Tab. 6.4, this systematic uncertainty is the largest of those assessed in this section. The total systematic uncertainty on R from the reconstruction's energy calibration, though, is at most only 23 ppb.
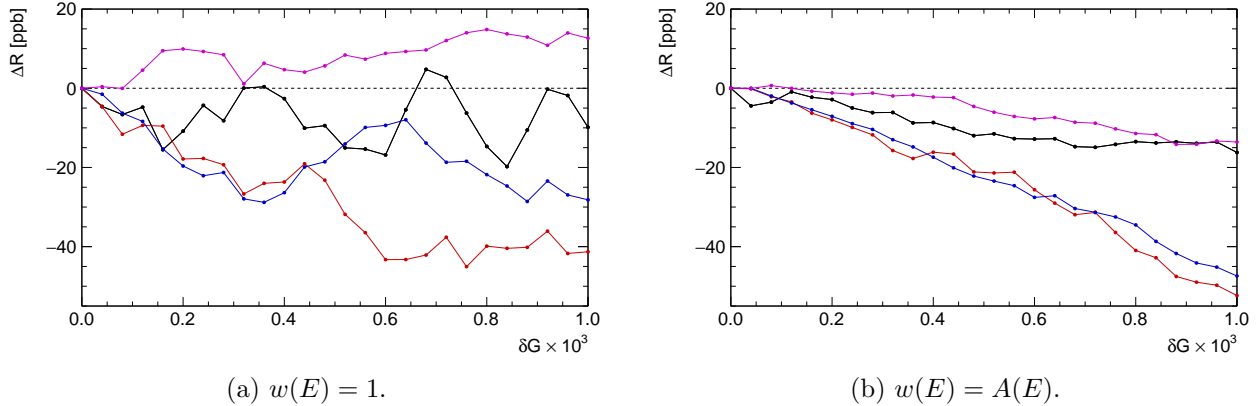
(a) $w(E) = 1$.

(b) $w(E) = A(E)$.

Figure 6.11: The change in R when applying the residual gain correction with increasing values of δG for the 1a (black), 1b (red), 1c (blue), and 1d (magenta) data set, using the (a) baseline and (b) asymmetry-weighted method.

6.4 Artificial Dead Time

If uncorrected, pileup—when two or more positrons arrive at a calorimeter too close in time and space to be individually resolved—poses a systematic ω_a-bias (Sec. 2.6). The amount of pileup in the reconstructed data that needs correcting depends, in part, on the reconstruction algorithm used. In particular, the reconstruction's ability to resolve two positron hits close in time can be characterized in terms of its dead time—the time separation below which the positrons are not always resolved. As discussed in Sec. 4.8, this dissertation's reconstruction is estimated to have a dead time near 1.5 c.t. A lower bound on this value is set by the fit algorithm's artificial dead time (ADT) of 1 c.t.—the time separation below which only one cluster is allowed (Sec. 4.5). To reduce the amount of pileup, the ADT chosen to be as low as reasonably possible. However, as the ADT decreases, the occurrence of *false pileup*—i.e., two close-in-time, resolved clusters that are not both decay positrons—will increase. It is, therefore, important to characterize the false pileup allowed by an ADT of 1 c.t. and to assess whether any systematic bias on R is caused by this choice of ADT, which will be discussed in turn in this section.

(a) $r = 0$ c.w.

(b) $r = 1$ c.w.
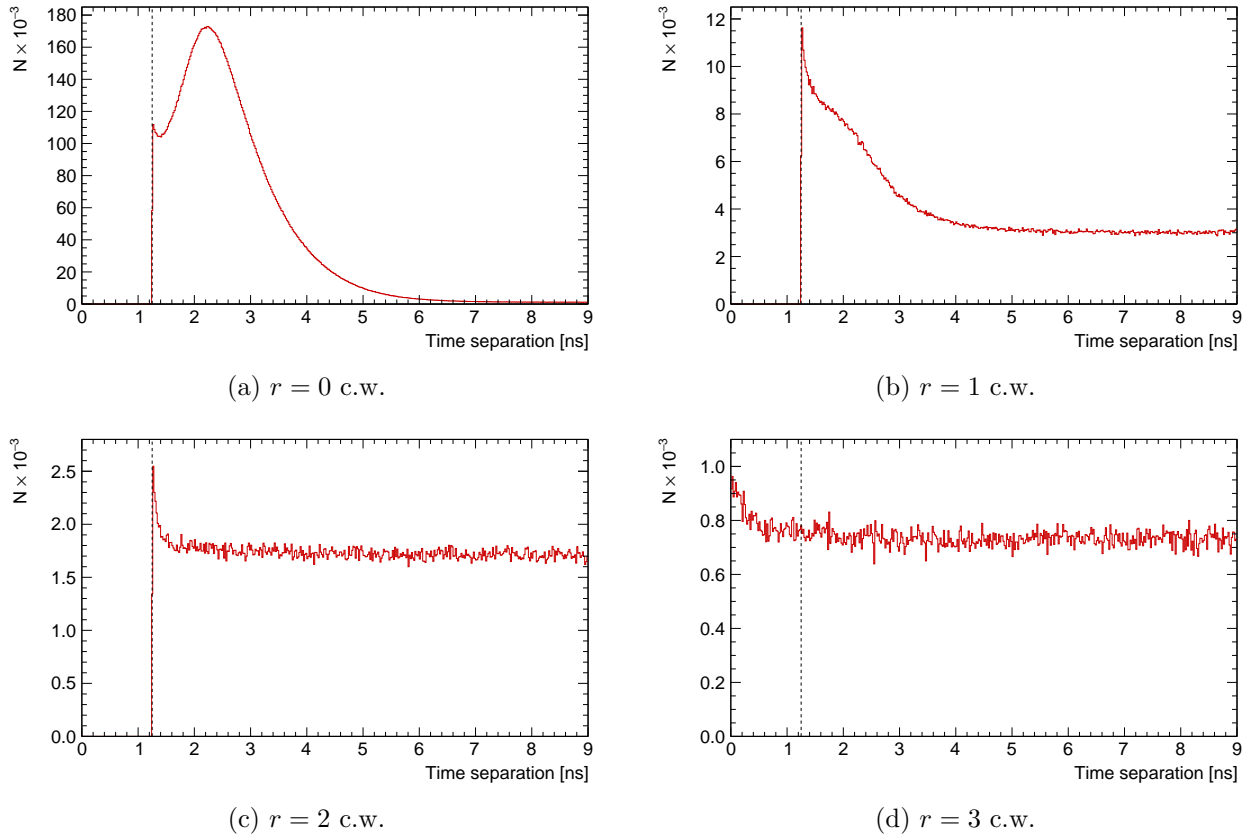
(c) $r = 2$ c.w.

(d) $r = 3$ c.w.

Figure 6.12: The time-separation distribution of cluster pairs in the 1a data set at increasing spatial separations (r) with a combined energy above 1.7 GeV. The reconstruction's artificial dead time is denoted by the dashed, vertical lines.

6.4.1 False Pileup

The probability for decay-positron pairs should be constant across time separations beyond the dead time (and small relative to the cyclotron period). The presence of false pileup can, thus, be identified as any excess above a flat distribution in the time-separation distribution of cluster pairs. Further, considering the baseline method, the cluster pairs of interest are those that, if combined, would pass the 1.7-GeV energy threshold and enter the histogram that is fit for ω_a. Therefore, to characterize any relevant false pileup, time-separation distributions were made using this energy cut for increasing spatial separations, as seen in Fig. 6.12. A spatial separation, r, of two crystal widths (c.w.), for example, refers to a cluster pair where the second cluster's position
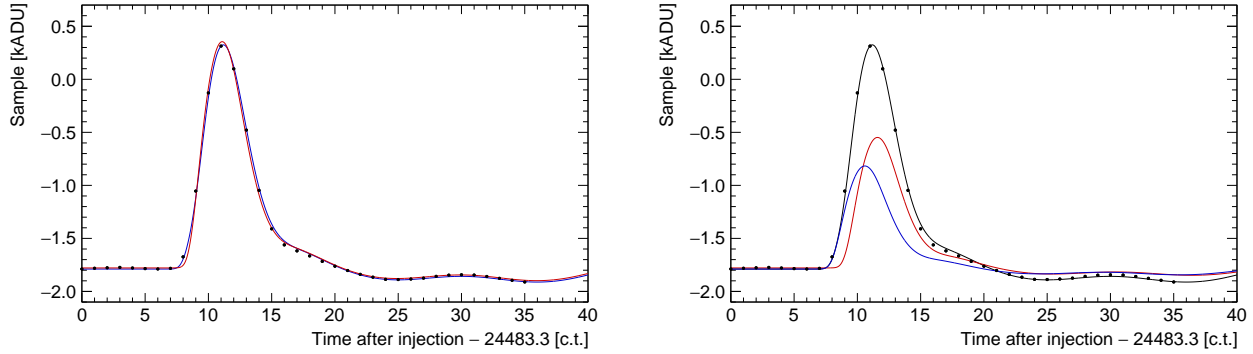
Figure 6.13: An example island in the 1a data set where one positron hit may have been reconstructed as two. Left: A fit to one (red) and two (blue) clusters with a time separation of 1.01 c.t. Right: The two-cluster fit (black) and its underlying clusters (blue, red).

is in a crystal two away horizontally or vertically from the first cluster's position.

For $r \leq 2$ c.w., there is a sudden increase at time separations between roughly 1.5 ns and the ADT of 1.25 ns. This fractional enhancement is also observed to be larger when using smaller energy thresholds [134]. As depicted in Fig. 6.13, this excess is predominantly due to a "positron split-up" phenomenon, where one positron is reconstructed as two by mistake because of a slightly wider pulse shape than that of the template. As discussed in Sec. 4.1, such distorted pulse shapes can arise simply from the positron's incident position/angle on the crystal face.

Furthermore, for $r = 3$ c.w., the cluster pair's 3×3 crystal windows no longer overlap, and, therefore, their time separation is not restricted by the ADT. In this case, as seen in Fig. 6.12d, an uptick is observed for very short time separations. This may be instead caused by the decay positron interacting with material in the ring before hitting the calorimeter. For example, bremsstrahlung is known from the ring's `Geant4`-simulation (Sec. 4.8) to occur in the quadrupole and kicker plates, resulting in such false pileup [135]. Moreover, this uptick could also be caused by fluctuations of the electromagnetic shower in the calorimeter, where a small amount of energy is deposited in a distant crystal and is reconstructed separately.

More notably, a broad distribution is found for $r \leq 1$ c.w., centered about 2.3 ns. As seen in

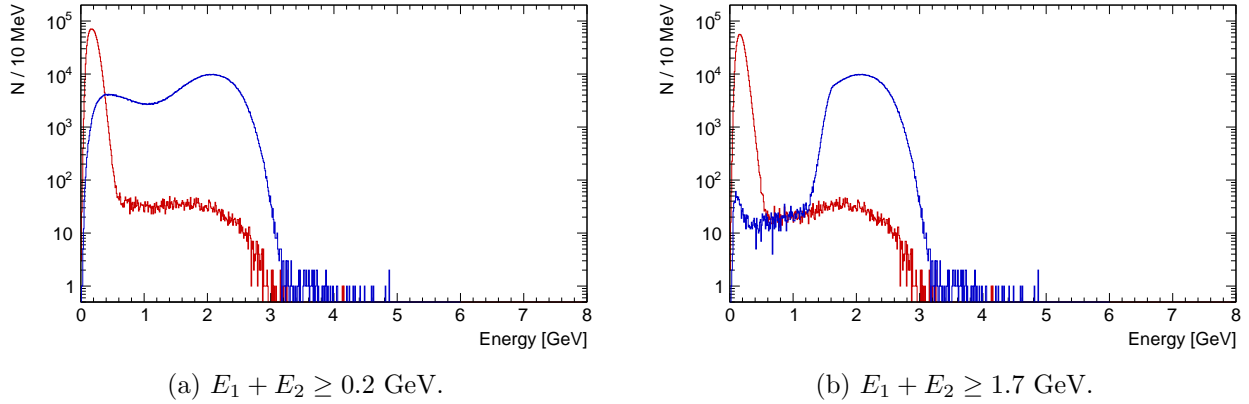(a) $E_1 + E_2 \geq 0.2$ GeV.

(b) $E_1 + E_2 \geq 1.7$ GeV.

Figure 6.14: The energy distribution of the first (blue) and second (red) cluster in time with an energy sum above (a) 0.2 and (b) 1.7 GeV. These cluster pairs from the 1a data set are selected to be separated spatially by $r = 0$ c.w. and temporally between 2.3 and 2.4 ns.

Fig. 6.14b, in this time-separation range, the first and second cluster in time are predominantly high- and low-energy, respectively. When reviewing these traces, a class of cluster pairs can be identified where a large pulse is immediately followed by a small "after-pulse" that accounts for a tail that is broader than the template's, as depicted in Fig. 6.15. As in positron split-up, such broadening can likely be due to the incident position and/or angle of the decay position on the crystal face. In these after-pulse cases, however, it can be seen that the first of the two pulses, as compared to the one-pulse fit, actually results in an amplitude that more closely matches the waveform's height. Therefore, even though a spurious cluster is obtained, the energy resolution of the high-energy positron is likely to be marginally smaller and, thus, better.

As shown in Fig. 6.14a, if the energy-sum cut is lowered to 200 MeV, a second peak appears around 400 MeV in the first pulse's energy distribution. This low-energy peak is thought to be caused by radiative muon decay, $\mu^+ \rightarrow e^+ + \gamma + \nu_e + \bar{\nu}_\mu$. Emitted tangentially, a γ will span a chord of only about 7 ° before pair producing in the outer vacuum chamber wall. For it to strike the same calorimeter as the decay positron, the pair-produced positron must be low in energy, so that it curls more in the magnetic field. Further, the pair-produced positron will be more
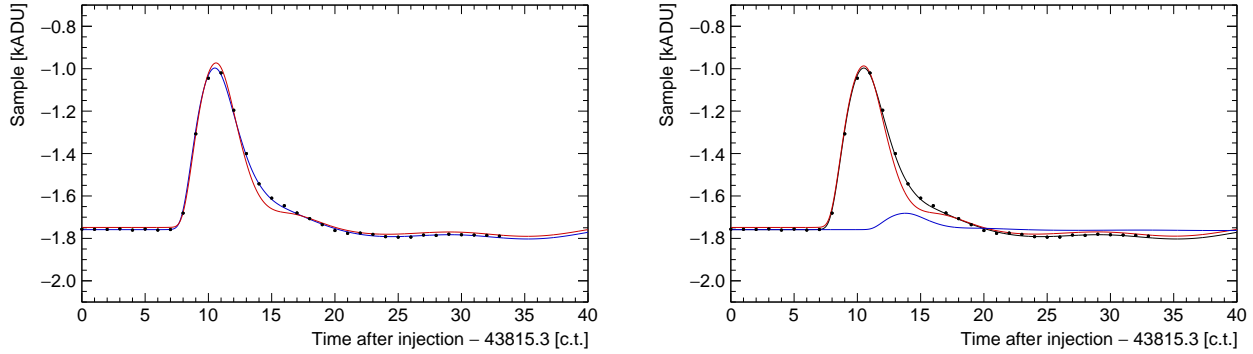
Figure 6.15: An example island in the 1a data set where an after-pulse cluster was reconstructed. Left: A fit to one (red) and two (blue) clusters with a time separation of 3.3 c.t. Right: The two-cluster fit (black) and its underlying clusters (blue, red).

likely to hit the side of the calorimeter than its front face due to the geometry, meaning that such false pileup will predominantly be observed in the crystal column closest to the storage region—a prediction supported by simulation and data. With the 200-MeV cut, the first column's time-separation distribution exhibits a large, narrow peak between 2.3 and 2.4 ns, which is not found in those of the other columns [134]. As illustrated in Fig. 6.16, the relatively sharp transition in the number of cluster pairs in this time-separation range between the first two crystal columns supports that such false pileup is primarily due to radiative muon decay—not differences in the columns' distribution of hit positions and angles.

In addition, the spatial distribution of cluster pairs separated by less than 5 ns is unique for each calorimeter [134]. For some calorimeters, these cluster pairs for $r = 0$ c.w. can even largely reside in a single crystal. In general, the spatial patterns simply reflect the crystals' relative energy calibration constant, found from the lost muons' energy signature (Sec. 4.6). Aside from variations in the crystals' light transmission, the calibration constant is a proxy for the crystal trace's gain and thereby for the reconstruction's threshold. This means that a lower constant will result in a lower energy threshold to fit a second cluster, resulting, for example, in more pairs with an after-pulse cluster. If combined among the calorimeters, though, the spatial distribution roughly
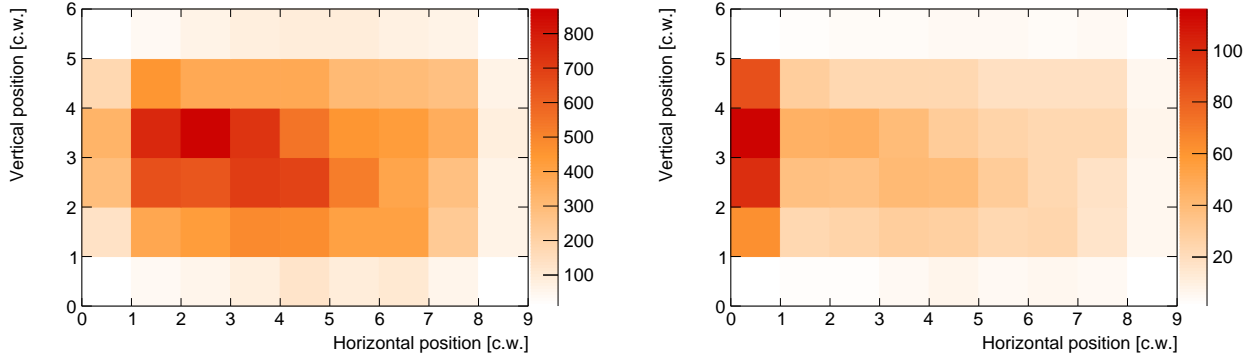
Figure 6.16: The spatial distribution across the calorimeter of cluster pairs, divided by 1000, with $r = 0$ c.w. in the 1a data set. The cluster pairs are selected to have an energy sum above 1.7 GeV and a time separation below 5 ns (left) or an energy sum above 0.2 GeV and a time separation between 2.3 and 2.4 ns (right). Lower horizontal positions are closer to the storage region.

approximates the positron hit distribution, as illustrated in Fig. 6.16. The only exception is a notable lack of cluster pairs in the column closest to the storage region—most likely because the reconstruction has a slightly higher threshold to add a new cluster when its central crystal in on the calorimeter's boundary (Sec. 4.5).

With respect to the crystal face, the spatial distribution of cluster pairs with an energy sum above 1.7 GeV and a time separation below 5 ns is non-uniform for each crystal [134]. The first cluster in these pairs most often strikes near the crystal's horizontal edge farthest from the center of the muon storage region. After that, it is most likely to hit close to the crystal's corners. For such incident positions, the electromagnetic shower is more likely to be spread across multiple crystals, with the struck crystal sampling more of the shower's tail. From simulation, crystals seeing primarily the shower's tail have a wider distribution of Čerenkov photons at the SiPM (Sec. 4.1). This causes the pulse shape to be broader than the template shape, increasing the chance for an after-pulse cluster—in agreement with what has been observed above.

Moreover, as shown in Fig. 6.17, a large variation exists in the amount of false pileup among the calorimeters. As expected, there is a local excess near $1/(160 \text{ MHz}) = 6.25$ ns for calorimeter 9, with the 200-MeV energy cut. The omnipresent, 160-MHz noise signal in this calorimeter

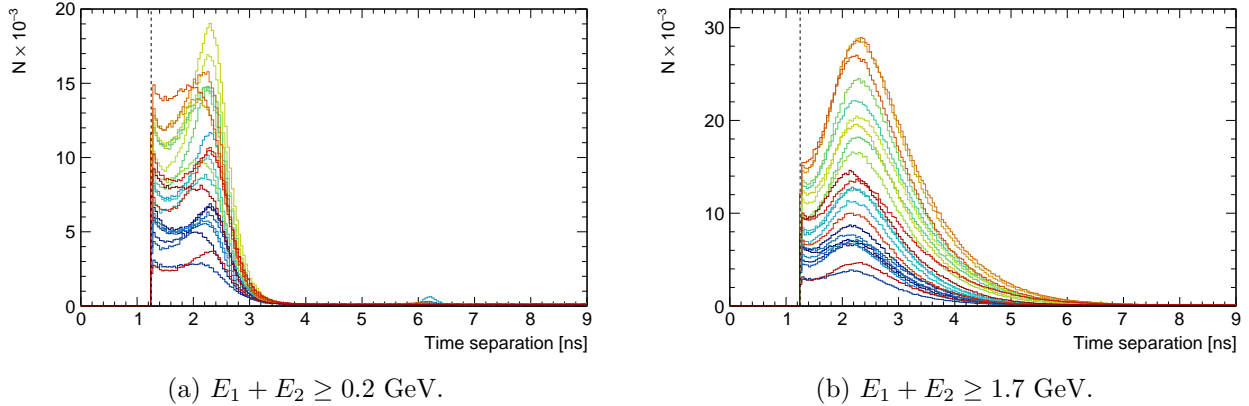(a) $E_1 + E_2 \geq 0.2$ GeV.

(b) $E_1 + E_2 \geq 1.7$ GeV.

Figure 6.17: The time-separation distribution of cluster pairs in the 1a data set with $r = 0$ c.w. and an energy sum above (a) 0.2 and (b) 1.7 GeV for all of the calorimeters, where each one is different color from red to blue in order of the rainbow. The reconstruction's artificial dead time is denoted by the dashed, vertical lines.

is known to be larger than in the others, causing a larger fraction of noise-noise cluster pairs to pass the reconstruction's noise mitigation strategy (Sec. 4.9). In contrast, when using a 1.7-GeV energy cut—where these noise-noise pairs are not present—each calorimeter's distribution appears approximately the same except for a scale factor. As discussed above, some of this variation is attributable to differences in the calorimeters' mean relative energy calibration constant. Most of the variation, however, is thought to be caused by differences in the amount of material—such as quadrupole and kicker plates—shadowing the calorimeters [134].

All together, many phenomena can produce false pileup, resulting in complex time-separation distributions. In particular, false pileup can occur between a decay positron and a noise cluster, an after-pulse, a pair-produced positron, a shower fluctuation, or a lost muon. It can also happen due to decay positron split-up or two noise clusters. However, regardless of its source, false pileup does not pose an inherent bias to ω_a as long as the false-pileup and false-pileup-free spectra have the same time dependence into the fill. In that case, including false pileup will simply yield a functional form with the same frequency—but a new asymmetry, phase, and normalization. Of the false-pileup categories identified, those from a single incident positron should follow the same

time distribution as the incident positrons themselves. An exception may arise, though, via the positron/after-pulse occurrence's position dependence because the positrons' hit distribution will change slightly as mean hit position drifts over the fill (Fig. 5.40).

Further, the only false-pileup category identified not from a decay positron alone is between a decay positron and a lost muon or between two noise clusters. The former process should occur at the muon-decay rate times the muon-loss rate, while the latter process should be uniform in time. These are not of concern, however, because any pileup with a lost muon will be rectified by the pileup correction and because noise clusters are characterized to be at low enough energies that, even if combined, they will not pass the ω_a-measurement's energy threshold (Sec. 4.9).

To test these predictions, the false pileup's time dependence was estimated by histogramming cluster pairs found between 1 and 4 c.t. for increasing spatial separations, as shown in Fig. 6.18. To remove the ω_a-oscillation and increase each time bin's statistics, each histogram was binned approximately in the ω_a-period, $1/(0.2291 \text{ MHz})$. For $r \geq 2$ c.w., where there is minimal false pileup, true pileup dominates the distribution, following the pileup rate of close to $e^{-t/(32.22\ \mu s)}$ (Sec. 5.4). As the spatial separation decreases, the false pileup's fraction increases to where, for $r = 0$ c.w., the distribution is closer to the muon-decay rate of $e^{-t/(64.44\ \mu s)}$. This is also found to a lesser extent using a time-separation window between 6 and 9 c.t. [134]. The precise time dependence, however, will slightly depend on the muon dynamics.

Accounting for muon loss (and binning out the ω_a-oscillation), the time dependence of cluster pairs close in time can be modeled by the sum of the true and false pileup rates:

$$N(t) = N_1 \cdot \Lambda^2(t) \cdot e^{-t/(32.22\ \mu s)} + N_2 \cdot \Lambda(t) \cdot e^{-t/(64.44\ \mu s)} \tag{6.3}$$

where $\Lambda(t)$ is the muon-loss correction (Sec. 5.5). Here, the free parameters, N_1 and N_2, reflect the abundance of true and false pileup, respectively. As illustrated in Fig. 6.18, for the cluster pairs relevant to the baseline method, the time distribution is well described by this simple model in

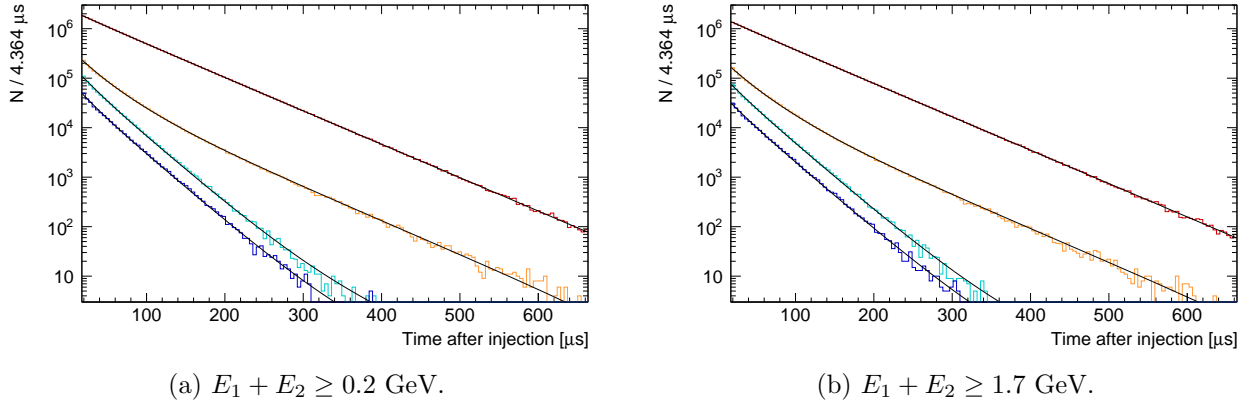(a) $E_1 + E_2 \geq 0.2$ GeV.   (b) $E_1 + E_2 \geq 1.7$ GeV.

Figure 6.18: The time distribution of the number of cluster pairs with an energy sum above (a) 0.2 and (b) 1.7 GeV, a time separation between 1 and 4 c.t., and a spatial separation, r, of 0 (red), 1 (orange), 2 (cyan), and 3 (blue) c.w. For each distribution, a fit to Eq. 6.3 is overlaid (black).

Eq. 6.3. It should be mentioned that the muon-loss correction yields a noticeable improvement in the $r \geq 1$ c.w. fits. Also, even when the energy-sum cut is lowered to 200 MeV so that the radiative muon decay's false pileup contributes, there is no evident deviation from the model. The false-pileup and muon-decay rates, thus, appear to have the same time dependence over the fill. As a result, the presence of false pileup is not expected to bias ω_a.

6.4.2 Artificial Dead Time Scan

If the ADT is increased, then the cluster pairs from false pileup will be merged together, thereby reducing the amount of false pileup. For an ADT larger than 1 c.t., the number of cluster pairs that would be merged can be read off of the time-separation distribution, such as in Fig. 6.12. Thus, a systematic bias to ω_a from false pileup can also be directly searched for by scanning the reconstruction's ADT from 1 to 3 c.t.—below which the majority of the false pileup occurs. Increasing the ADT further will primarily cause only true-pileup pairs to be combined.

In addition, when the ADT increases beyond the typical pulse width, as shown in Fig. 6.19, a single-cluster fit visibly becomes a poor fit to the data. As it increases even further, the individual pulse peaks become resolved, leading to a nonsensical fit. In this extreme, the largest pulse is fit
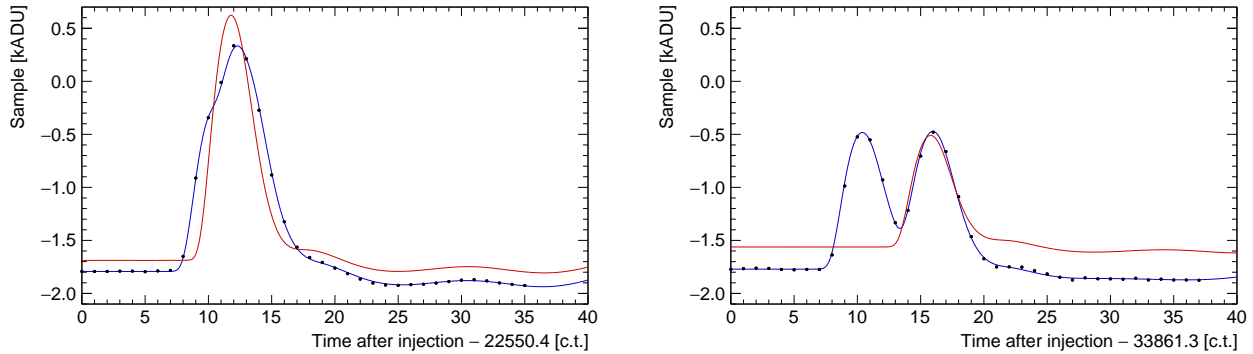
Figure 6.19: A fit to one (red) and two (blue) clusters with a time separation of 2.7 (left) and 5.6 (right) c.t. These islands are from the 1a data set.
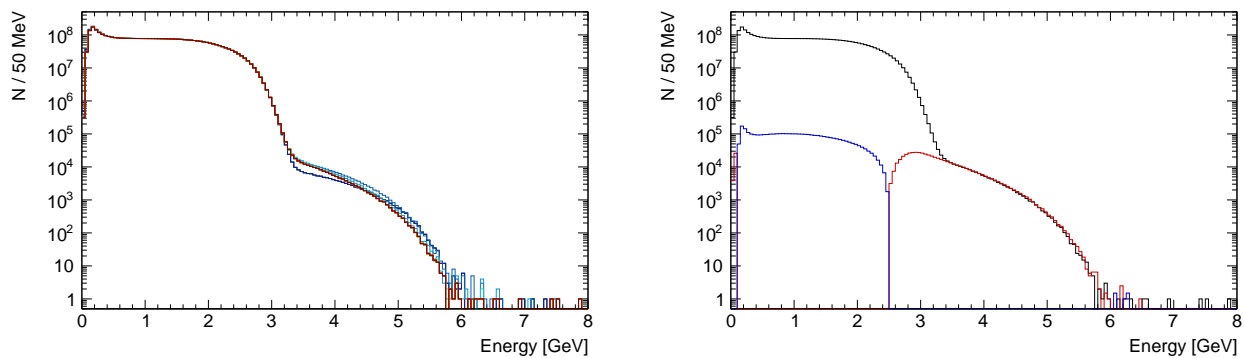

Figure 6.20: Left: The energy distribution of clusters reconstructed using an ADT ranging from 1 to 10 c.t. in steps of 1 c.t., which each one is a different color from red to blue in order of the rainbow. Right: The measured (black) and pileup (blue/red) energy distribution for an ADT of 10 c.t. Positive is red; negative is blue. These data are from the 1a data set.

with a raised pedestal to try to account for the other nearby pulses, which the ADT prevents from being included. Moreover, a larger ADT will introduce more true pileup and in turn test the limits of the pileup-correction method—together with any bias from false pileup. As shown in Fig. 6.20, the reconstructed energy distribution's high-energy tail—a non-physical regime consisting only of pileup—stops increasing when the ADT is above 3 c.t. This is expected since, when nonsensical fits begin to dominate, the energy contribution from nearby pulses is lost. In any case, this dissertation's pileup-correction method combines nearby islands that are passed through the same algorithm for a per-event modeling of the reconstruction's behavior—however bizarre (Sec. 5.4). As seen in Fig. 6.20, this method can fully account for the high-energy tail's shape even for an

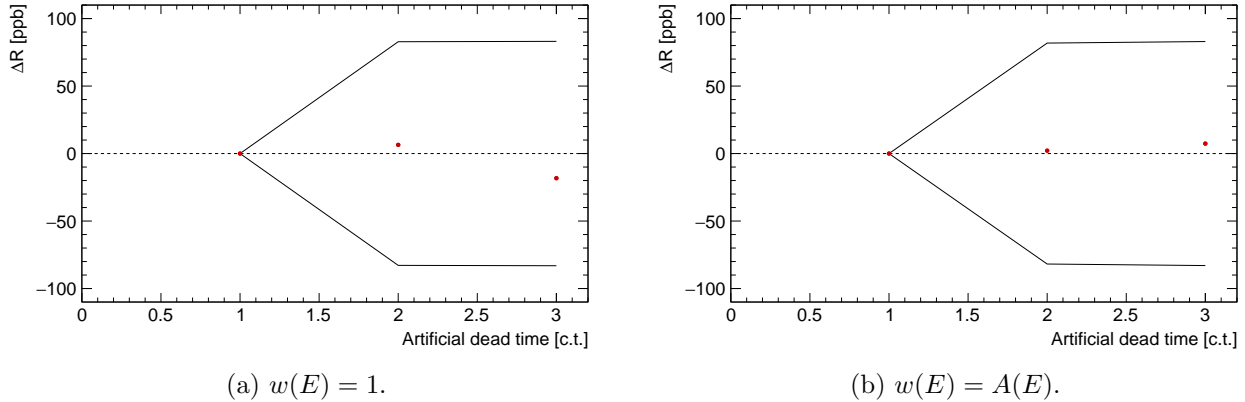(a) $w(E) = 1$.   (b) $w(E) = A(E)$.

Figure 6.21: The change in R for the (a) baseline and (b) asymmetry-weighted method with the 1a data set when reconstructed using different artificial dead times. The bands (black) show the allowed drift due to statistical differences (Eq. 5.44).

ADT of 10 c.t. It should be noted, though, that the method's sample-trimming option must be disabled if the ADT is above about 5 c.t.

Therefore, a scan over the reconstruction's ADT is expected to primarily test for a systematic bias to ω_a from false pileup. To that end, the 1a data set was reconstructed multiple times using an ADT of 1, 2, and 3 c.t. As shown in Fig. 6.21, for both the baseline and asymmetry-weighted method, the change in R for each ADT remained well within the allowed statistical drift due to different statistics. This allowed drift was calculated according to Eq. 5.44, using an estimate for f_{21} [134]. Based on such good agreement, no systematic bias to ω_a from false pileup is evident. A more stringent test can be made if repeated on larger data sets, but reconstructing all of the data sets in Run 1 several times is computationally impractical. However, after the nominal reconstruction, such tests can be approximated—as will be discussed next.

6.4.3 Time Clustering

After the nominal reconstruction (with an ADT of 1 c.t.), a larger ADT can be approximated by artificially combining clusters below some time separation. To do that, each cluster is combined with any following clusters less than the desired ADT by summing their energies and taking an
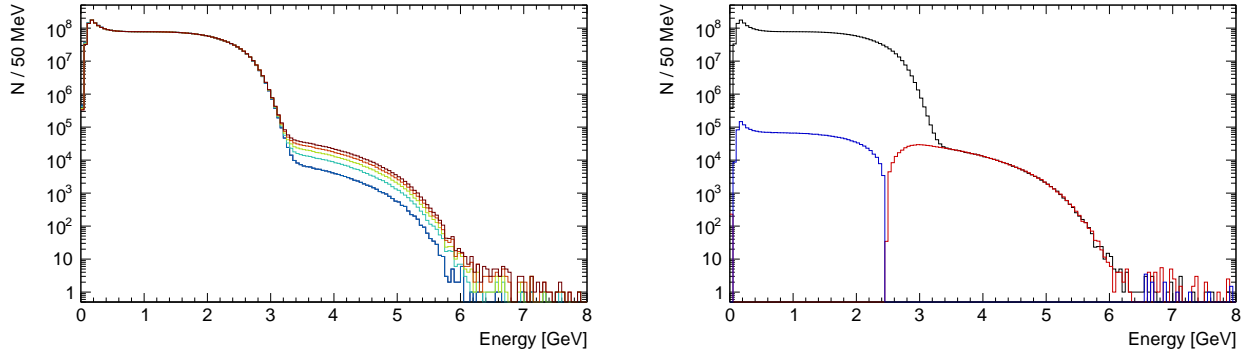
Figure 6.22: Left: The energy distribution after time clustering using an ADT ranging from 1 to 10 c.t. in steps of 1 c.t., which each one is a different color from red to blue in order of the rainbow. Right: The measured (black) and pileup (blue/red) energy distribution for an approximated ADT of 3 c.t. Positive is red; negative is blue. These data are from the 1a data set.
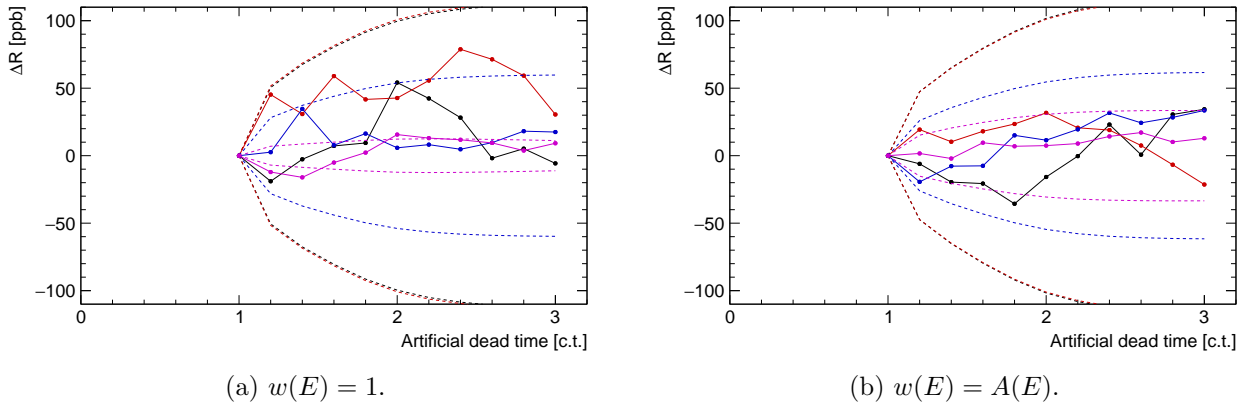

(a) $w(E) = 1$.

(b) $w(E) = A(E)$.

Figure 6.23: The change in R for the (a) baseline and (b) asymmetry-weighted method with the 1a (black), 1b (red), 1c (blue), and 1d (magenta) data set when time clustering with different artificial dead times. The dashed bands show the data sets' allowed statistical drifts (Eq. 5.44).

energy-weighted time. As each cluster is spatially defined by a 3×3 crystal window, this *time clustering* procedure is applied only to clusters whose central crystals are adjacent. As a result, the measured energy distribution's high-energy tail, as shown in Fig. 6.22, linearly scales with this procedure's ADT. Therefore, such time clustering appears to mimic the reconstruction's behavior for an ADT up to around 3 c.t. As also shown in Fig. 6.22, by applying this procedure to the pileup correction's clusters too, this dissertation's pileup correction method accounts for the pileup increase very well—without any tweaks—for every ADT tested.

Uncertainty source	$w(E) = 1$				$w(E) = A(E)$			
	1a	1b	1c	1d	1a	1b	1c	1d
Covariance matrix	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Amplitude	0.9	0.8	0.5	0.7	0.7	0.3	0.4	0.7
Time/energy bias	0.0	0.0	0.0	0.0	0.2	0.2	0.1	0.1
Pileup rate error	20.9	23.0	19.8	17.8	0.4	0.7	0.2	0.1
Unseen pileup	3.7	0.4	1.6	0.7	1.2	2.8	0.6	1.1

Table 6.7: A summary of the ω_a-systematic uncertainties, in ppb, from the pileup correction for each data set in Run 1. The combined uncertainties are shown in Tab. 6.1.

To further look for any systematic bias to R from false pileup, this procedure was applied to each data set using an (approximated) ADT from 1 to 3 c.t. in steps of 0.2 c.t. As presented in Fig. 6.23, for both the baseline and asymmetry-weighted method, the change in R for each ADT remained in agreement with the allowed statistical drift (Eq. 5.44) even for the 1d data set—the largest data set in Run 1. Together with the results above, a systematic bias to ω_a from false pileup is neither expected nor identified. Therefore, no systematic uncertainty is assigned to this dissertation's ω_a-measurement due to the reconstruction's small ADT of 1 c.t.

6.5 Pileup Correction

The calorimeter, backend electronics, and reconstruction algorithm were largely designed to mitigate the level of pileup. All together, for each data set, the pileup contamination was limited to less than 0.08 % of the total number of clusters (Sec. 5.4). Even this amount of pileup, though, was enough to need a correction. To review briefly, the pileup correction developed is based on artificially creating pileup events by combining the waveforms of islands. The correction is then formed from the difference of the distributions of clusters reconstructed in the uncombined islands and those in the combined island. In this section, the advantages and disadvantages of this approach will first be considered. The possible sources of systematic uncertainty on ω_a from the

pileup correction will then be estimated. However, the systematic uncertainty from pileup involving clusters below the reconstruction's energy threshold—known as unseen pileup—will be reserved for the next section. A summary of the individual systematic uncertainties assigned to the pileup correction is provided in Tab. 6.7.

6.5.1 Pileup Correction Approach

As mentioned in Sec. 5.4, the pileup correction's approach avoids modeling challenges by replacing a model with the reconstruction itself. The positrons used are sampled from the same observed time and space distribution used to obtain ω_a, resulting in as close to a perfect model as can be expected. Any nonlinearities, crystal-to-crystal differences, muon-dynamics changes, or anything else are all incorporated. For instance, the reconstruction's resolution does not need to be carefully understood and modeled around crystals having a partially unresponsive SiPM, an example of which is shown in Fig. 6.24. Further, this modeling approach is robust against changes to the reconstruction and/or simulation, which, from a practical view, is especially convenient.

A key drawback to this approach is that the noise level is inflated by the square root of the number of islands combined. Consequently, the reconstructed clusters' energy resolution will be increased slightly—but acceptably considering the ω_a-measurement's modest energy resolution requirement (Sec. 2.5). In addition, the reconstruction's threshold to add a new cluster will be inflated by the same percentage, which, when combining two islands, increases it by roughly 27 analog-to-digital units. The lowest energy cluster that can be observed, therefore, will be slightly higher, making the pileup correction incorrect at energies around the reconstruction's threshold. However, such low-energy clusters do not enter the ω_a-measurement and, thus, do not pose a problem either. The influence of clusters around (and below) the reconstruction's threshold, on the other hand, will be assessed as a systematic uncertainty in Sec. 6.6.
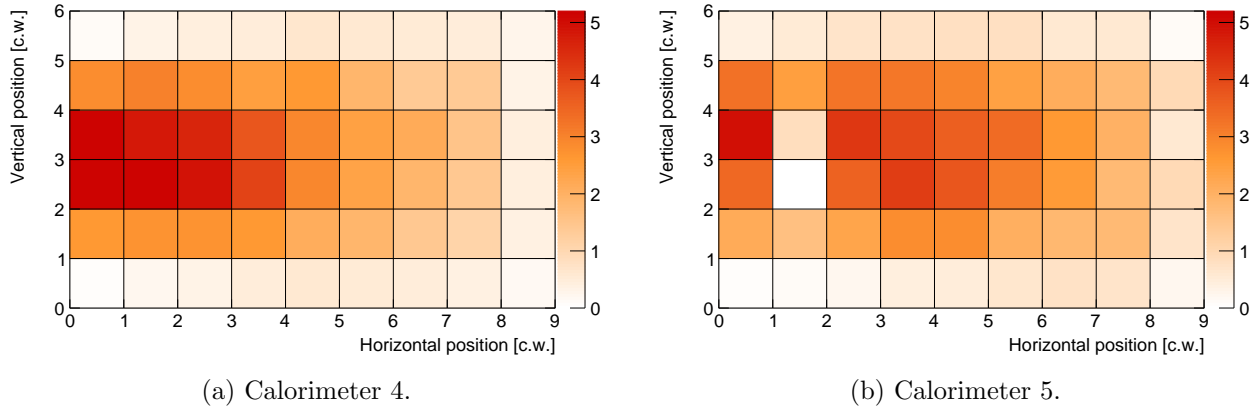
(a) Calorimeter 4.

(b) Calorimeter 5.

Figure 6.24: The spatial distribution of clusters, in percent, reconstructed with an energy above 1.7 GeV in calorimeter (a) 4 and (b) 5, using the 1c data set. At least one crystal–SiPM pair in calorimeter 5 is observed to be partially unresponsive throughout Run 1.

Another drawback is that the SiPM-pixel gain recovery is not modeled when combining the two islands. In a true pileup event, the probability for a Čerenkov photon to hit a pixel recovering from an earlier photon hit is enhanced. However, since the SiPM employed has 16 channels with 3584 pixels per channel, this probability is negligibly small.

Furthermore, the approach suffers from the pileup rate error discussed in Sec. 5.4, which will be assessed as a systematic uncertainty in Sec. 6.5.5. An alternative approach, which removes this error, is to combine islands from two different fills—instead of from the same fill—to artificially create known pileup events in an analogous way. In this approach, for every cluster in the first fill, a pileup candidate is formed with a cluster in the second fill if they are within δt and overlap spatially. This alternative approach, however, has two prominent limitations. First, the absolute time offset between two crystals that are connected to different digitizers varies between fills if a sub-clock tick slip occurs in only one of them. When combining the discrete samples, the waveform for one of the crystals will, therefore, have to be interpolated by up to half a clock tick for their absolute times to match. To that end, a simple parabolic interpolation between groups of three adjacent samples was tried, but it noticeably distorts the pileup energy spectrum.
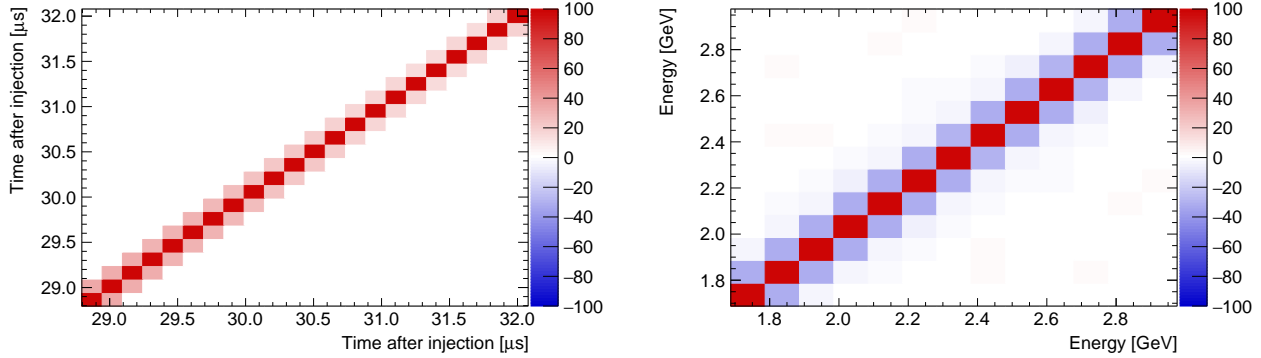
Figure 6.25: The pileup correction's correlation matrix, in percent, among time bins using the baseline method (left) and among energy bins using the baseline weighting (right) for the 1a data set. Here, the measured number of counts, N_m, in Eq. 5.22 has been neglected. The inter-bin correlations are, thus, only those due to the pileup subtraction.

Second, the positron rate between the fills may not be equal. When the beam returns, for example, after being briefly sent to the switchyard every minute for FNAL's test beam facility, the number of clusters reconstructed drops by about 40 % for the next eight fills. The pileup rate, following approximately the square of the positron rate, will then not necessarily average out because the sum of the squares does not always equal the square of the sums, i.e., $\langle \rho(t)^2 \rangle \neq \langle \rho(t) \rangle^2$. Together, these two limitations outweigh the pileup rate error incurred in the same-fill approach, but, even so, both methods yield similar results. All in all, the approach taken in Sec. 5.4 has a number of conceptual and practical advantages and no significant drawbacks.

6.5.2 Covariance Matrix

By construction, the pileup correction is correlated with the reconstructed data, and this correlation must be known to apply the correct bin uncertainties in the χ^2 in Eq. 5.43. As discussed in Sec. 5.4, the correlation can be computed exactly by keeping track of the number of clusters that participate in the pileup construction according to Eqs. 5.22 and 5.23. For the ω_a-measurement in Ch. 5, however, only the covariance matrix's diagonal elements were used, and the triple-pileup correction's contribution was entirely neglected. The influence of these choices on the best-fit R
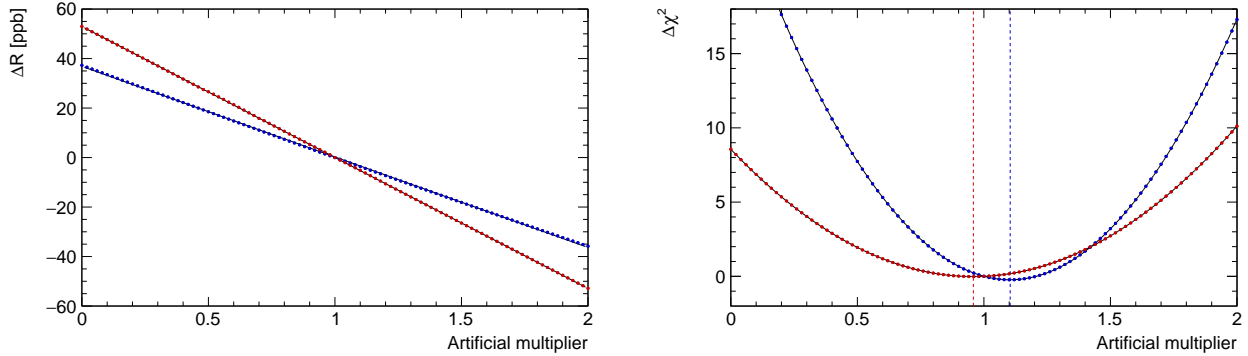
Figure 6.26: The change in R (left) and χ^2 (right) when artificially scaling the pileup correction's amplitude in the 1c data set, using the baseline (red) and asymmetry-weighted (blue) method. A unit multiplier corresponds to the nominal pileup correction. The vertical, dashed lines denote the minima of parabolic fits to the changes in χ^2.

must, therefore, be evaluated. To provide context, using these choices, the correlations due to pileup change both R and its statistical uncertainty, σ_R, on the order of 0.1 ppb when using the baseline or asymmetry-weighted method. The added changes from neglecting the aforementioned contributions are, thus, likely negligible—but, nonetheless, should be verified.

Because the pileup candidates are formed from clusters about a cyclotron period apart—the same width as the time histogram's bins—adjacent time bins will, by construction, be partly correlated with each other. Moreover, this off-diagonal component should not average out when randomizing the fills' injection times (Sec. 5.6). As illustrated in Fig. 6.25, the time-bin correlations caused by the pileup correction are found to oscillate between 15 and 35 % in phase with the ω_a-oscillation. This simple model was incorporated into the χ^2 in Eq. 5.43 and found to change R and σ_R on the order of 0.001 and 0.01 ppb, respectively. Therefore, the systematic uncertainty on ω_a from neglecting the inter-bin correlations is indeed safely negligible.

In addition, neglecting the inter-bin correlations, the triple-pileup correction's contribution to the covariance matrix was computed according to the procedure described in Sec. 5.4.3. This test was found to change both R and σ_R on the order of 0.001 ppb—the same order of magnitude as the changes found by repeated minimizer calls in the ω_a-fit (Sec. 5.7). Accordingly, the systematic

uncertainty on ω_a from neglecting those contributions to the time-bin uncertainties in Eq. 5.43 is likewise safely negligible. All together, these fit results were the reasons behind the simplifications used to obtain the covariance matrix. It should be noted, though, that, while more computationally intensive, there is no fundamental limitation to calculating the full covariance matrix and incorporating it into the χ^2 if needed for future data sets.

6.5.3 Amplitude

The pileup construction is based on statistical sampling of the reconstructed clusters and, therefore, may differ from the true pileup distribution by up to their separate statistical fluctuations. To assess its corresponding ω_a-bias, the amplitude of the pileup correction was artificially scaled from zero (i.e., no pileup correction) to two (i.e., twice the pileup correction). Note that, for each artificial multiplier sampled, the appropriate ξ-factor was applied to the covariance matrix used in Eq. 5.22. Also note that this is a very conservative assessment since it assumes that the differing statistical fluctuations are coherent among all of the histogram's time bins.

As shown in Fig. 6.26, R is linearly biased when the pileup correction's amplitude is scaled. This allowed for R's sensitivity to be obtained from a linear fit. Across the data sets, as seen in Tab. 6.8, R is found to be around 28 % less sensitive to the pileup correction's amplitude for the asymmetry-weighted method as compared to the baseline method. Further, as also shown in Fig. 6.26, the artificial multiplier minimizing the fit's χ^2 agrees with one to within 1.4 standard deviations among the data sets. This agreement is striking since, as discussed in Sec. 6.3.3, the pileup correction can partially compensate for a residual gain variation. For instance, the instability observed in $\kappa_{\text{loss}}(E)$ can be resolved by reducing the pileup correction's amplitude by around 10 %. However, if such a large reduction is applied, the measured and pileup energy distributions would no longer agree beyond the endpoint of the decay-positron spectrum.

Sensitivity	$w(E) = 1$				$w(E) = A(E)$			
	1a	1b	1c	1d	1a	1b	1c	1d
$\mathrm{d}R/\mathrm{d}A^{\mathrm{pileup}}$ [ppb]	-93.4	-80.0	-52.9	-69.5	-74.3	-32.2	-36.6	-70.3
$\mathrm{d}R/\mathrm{d}t^{\mathrm{pileup}}$ [ppb/ns]	0.9	1.0	0.9	0.8	5.0	5.6	5.0	4.5

Table 6.8: The R-sensitivity to artificially scaling the pileup correction's amplitude, A^{pileup}, and shifting the pileup correction in time, t^{pileup}, for each data set and weighting method.

Because such high energies are non-physical, all of the observed clusters in this regime must be pileup and, thus, serve as a stringent test of the pileup correction's accuracy. In particular, as illustrated in Fig. 5.5, the amplitude's accuracy can be assessed by taking the ratio of the pileup to the measured spectrum. A constant fit to this ratio for energies above 3.8 GeV yields a best-fit value consistent with one to less than a standard deviation—which, for the smallest data set, is 0.76 %. Therefore, extrapolating from this regime, the pileup-correction amplitude is generously estimated to be accurate to 1 % for each data set. Accordingly, the systematic uncertainty on ω_a can then be taken as 1 % of its sensitivity to scaling the amplitude. As shown in Tab. 6.7, this estimate is found to be less than 1 ppb for each data set and weighting method. Consequently, a less conservative treatment is considered to be unnecessary for Run 1.

6.5.4 Time/Energy Bias

As described in Sec. 5.4, the pileup correction is built from a linear combination of single, double, and triple distributions. Each of these distributions is in turn built from single and/or pileup clusters that are reconstructed. These clusters' times and energies, as shown in Fig. 4.36, can be biased by up to 30 ps and 20 MeV, respectively. In general, these biases pose an ω_a-bias if they differ between the clusters nominally reconstructed and those reconstructed as part of the pileup construction. However, based on the discussion in Sec. 6.5.1, any such differences are anticipated to be minimal. This expectation is supported by the reconstruction's artificial dead time scan in
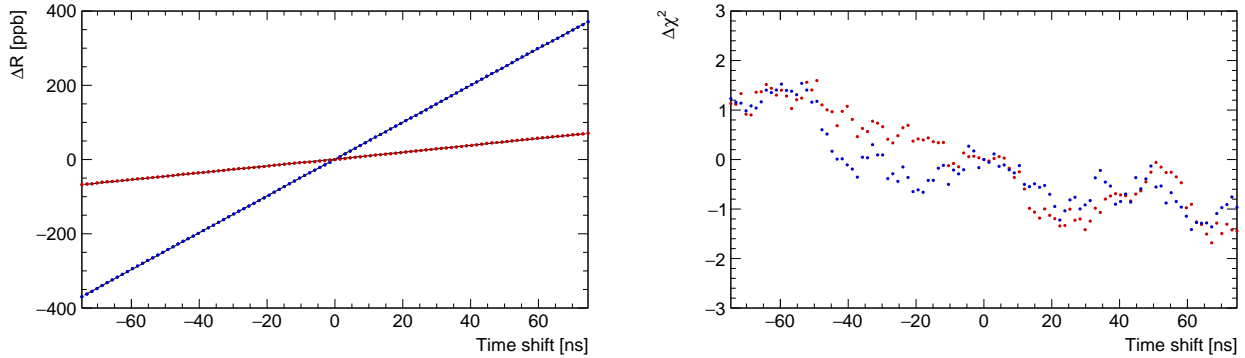
Figure 6.27: The change in R (left) and χ^2 (right) when artificially shifting the pileup correction in time for the 1c data set, using the baseline (red) and asymmetry-weighted (blue) method. A time shift of zero corresponds to using the nominal pileup correction.

Sec. 6.4.2, where the clusters' biases were magnified.

Notwithstanding, an upper bound of the corresponding ω_a-bias can be assessed by considering every cluster used in the pileup construction to be coherently biased. To that end, if each cluster's energy was increased by 20 MeV, the resulting pileup correction would be approximately equivalent to scaling the pileup correction's amplitude. A potential difference in the energy bias, therefore, will already be treated as part of the conservative assessment in Sec. 6.5.3. More of a concern is a difference in the time bias, particularly for the asymmetry-weighted method because, as seen in Fig. 5.6, the pileup correction's shape is dominated by an oscillation at ω_a.

To assess R's sensitivity for each method, the time of each cluster used in the pileup correction was shifted by up to half of a cyclotron period. As shown in Fig. 6.27, R is observed to be linearly biased by such a time shift. The R-sensitivity found from a linear fit for each weighting method and data set is reported in Tab. 6.8. As expected, R's sensitivity using the asymmetry-weighted method is around five times larger than that using the baseline method—approximately equal to the ratio of the oscillation amplitudes for the two methods in Fig. 5.6. An upper bound of ω_a's systematic uncertainty from a different time bias in the pileup construction is then found by multiplying the R-sensitivity by 30 ps. As seen in Tab. 6.7, this conservative estimate is below 0.1 and

0.3 ppb for the baseline and asymmetry-weighted method, respectively, for each data set. Since this is already so small, more stringent bounds will not be estimated.

6.5.5 Pileup Rate Error

In the pileup-correction approach, pileup candidates are formed, in part, from clusters that are within 5 c.t. of $\Delta t = 119.36$ c.t. ≈ 149.2 ns. As discussed in Sec. 5.4.1, Δt was set as low as reasonably possible so that the positron rates sampled at time t and $t + \Delta t$ are as close as possible. To improve this assumption—which is formally given in Eq. 5.7—the reconstructed cluster times were also shifted by $\Delta t/2$. If the positron rate is only due to exponential decay, this pileup-rate assumption would be exactly true. However, the muon precession—and, to a smaller extent, the muon loss and dynamics—introduce a difference between the pileup rate sampled, $\rho(t - \Delta t/2) \times \rho(t + \Delta t/2)$, and the pileup rate assumed, $\rho(t)^2$. This causes an under- or over-estimate of the pileup correction used, which will primarily oscillate at ω_a. As a result, the pileup correction's shape will be distorted, leading to a systematic bias on the best-fit R.

For each data set, the expected distortion can be closely estimated from the nominal best-fit parameters, $\boldsymbol{p}$, in Sec. 5.7. At some time t, the positron rate should follow $N(t \mid \boldsymbol{p})$ in Eq. 5.36. Therefore, the pileup-rate error can be calculated by subtracting what was wanted, $N(t)^2$, from what was done, $N(t - \Delta t/2) \cdot N(t + \Delta t/2)$. As shown in Fig. 6.28, this error oscillates approximately at ω_a between -0.6 and $+0.3$ % of what was wanted. Also as expected, the fractional error over an ω_a-period is roughly constant across the entire fill. The pileup-rate error can, thus, be largely removed by multiplying each time bin in the pileup correction (before being applied) by

$$C_2(t) = \frac{N(t)^2}{N(t - \Delta t/2) \cdot N(t + \Delta t/2)} \tag{6.4}$$

where $N(t)$ is evaluated using the data set's nominal fit result. Further, an analogous error arises from the pileup candidates used in the triple-pileup correction (Sec. 5.4.2). As seen in

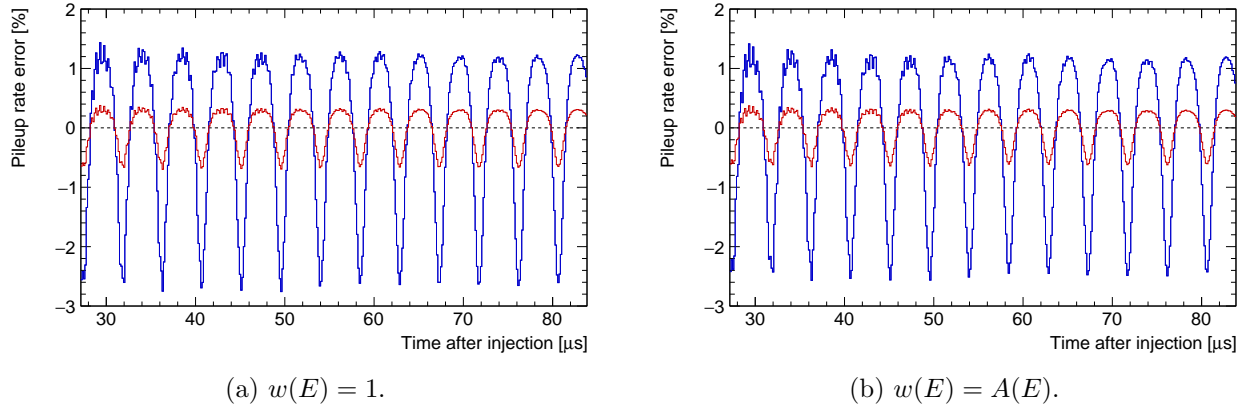(a) $w(E) = 1$.

(b) $w(E) = A(E)$.

Figure 6.28: The pileup-rate error for pileup candidates formed from two (red) and three (blue) clusters, using $N(t)$ with the best-fit parameters found from the 1c data set via the (a) baseline and (b) asymmetry-weighted method in Sec. 5.7.

Fig. 6.28, it has a similar shape, oscillating between roughly -2.5 and $1.2\ \%$ of what was wanted. Thus, the triple-pileup correction's shape will be distorted more, but, given the size of the triple-pileup correction, the distortion's biasing of R is expected to be small. Nevertheless, this correction's pileup rate error can also be largely eliminated by multiplying each time bin in the triple-pileup correction (before being applied) by $C_2(t)$ using Δt instead of $\Delta t/2$.

The change in R when applying both pileup-rate corrections is shown in Tab. 6.7. The baseline method's R for each data set is found to be about 20 times more sensitive to the pileup-rate error as compared to the asymmetry-weighted method. As with R's sensitivity to a time shift in Sec. 6.5.4, this difference arises from the difference in the time distributions for the two methods (Fig. 5.6). The bias to R even using the baseline method, though, remains small compared to R's statistical uncertainty. The change in R is, therefore, taken as the systematic uncertainty from the pileup-rate error. For future data sets, if such an uncertainty ever becomes unacceptably large, a correction similar to the one described above can be made for the final result. In this case, R's systematic uncertainty will arise from any inaccuracy in the specific correction used.

6.6   Unseen Pileup

A particle striking the calorimeter is referred to as unseen (seen) when its energy is below (above) the reconstruction's threshold to be detected. This leads to the phenomenon of *unseen pileup*, defined as when two or more particles arrive at the calorimeter close in time and space but at least one of them is unseen. The pileup correction in Sec. 5.4, though, cannot account for such pileup because it is built upon seen clusters. As will be first established in this section, neglecting unseen pileup poses an ω_a-bias, and, thus, a systematic uncertainty on ω_a from it must be assessed. To that end, this section's approach is to build the pileup correction using only clusters above an artificially high energy threshold, enhancing ω_a's bias by faking more unseen pileup. A realistic systematic uncertainty for such an artificial threshold will then be extracted with simulation and an understanding of the reconstruction's true energy threshold. The result from this approach is summarized for each data set in Tab. 6.7.

6.6.1   Time/Energy Bias Expectation

The presence of unseen pileup will bias ω_a only if the unseen cluster introduces a net time and/or energy bias to the seen cluster. If there is no time separation between them, the seen cluster's time will be unchanged but its energy will be augmented by the unseen cluster's. If, however, they are offset in time from each other, the seen cluster's energy will be lowered as the fitted pedestal will be falsely raised to account for the extra signal. In this case, the seen cluster's time will be biased towards the unseen cluster's. As—for small time separations—the unseen cluster is equally likely to occur before or after the seen cluster, one may anticipate such time and energy biases to cancel as a first approximation, yielding no systematic bias to ω_a.

To estimate the size of any incomplete cancellation, a Monte Carlo simulation was developed that imitated an unseen pileup event [136]. The procedure was to generate an analog trace from a
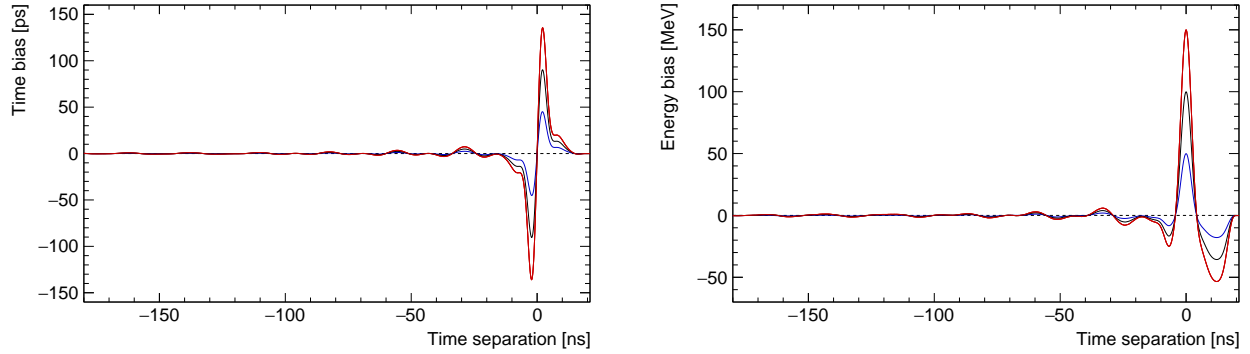
Figure 6.29: A 1.7-GeV seen cluster's time (left) and energy (right) bias at different time separations with an unseen cluster energy of 50 (blue), 100 (black), and 150 (red) MeV.

seen and unseen pulse, each of known amplitude and time, using a representative template. The amplitude was scaled so that a 4-GeV positron's pulse maps to the waveform digitizer's full bit-range—as was approximately true in Run 1. The waveform was then sampled at discrete times in steps of 1.25 ns, with the first time sampled offset by a random number drawn from a uniform distribution between 0 and 1.25 ns—so as not to sample the pulses always at the same time for each trial. To include realistic noise, each sample value was shifted by a random number drawn from a normal distribution with a width of four analog-to-digital units (ADU), and the total value was rounded off to the nearest integer. Afterward, the digitized trace was fit using the reconstruction's fit algorithm with a 1×1 crystal window (Sec. 4.5). The time and energy bias is then defined as the difference from the seen pulse's true time and energy, averaged over 10^4 trials.

The unseen pulse's energy was swept from zero to 200 MeV—roughly the reconstruction's highest energy threshold, as will be seen in Sec. 6.6.2—in steps of 10 MeV. For pulses of each unseen energy, their times were swept from 225 ns before to 27 ns after the seen pulse's time in steps of 0.125 ns, spanning the time range in which the unseen pulse's template overlaps with the fit region. For a 1.7-GeV seen cluster, the largest absolute time and energy bias observed is about 200 ps and 200 MeV, respectively [136]. As shown in Fig. 6.29, the bias directions agree with the above expectation, with a net bias close to zero. Note that the bias oscillations at large, negative time
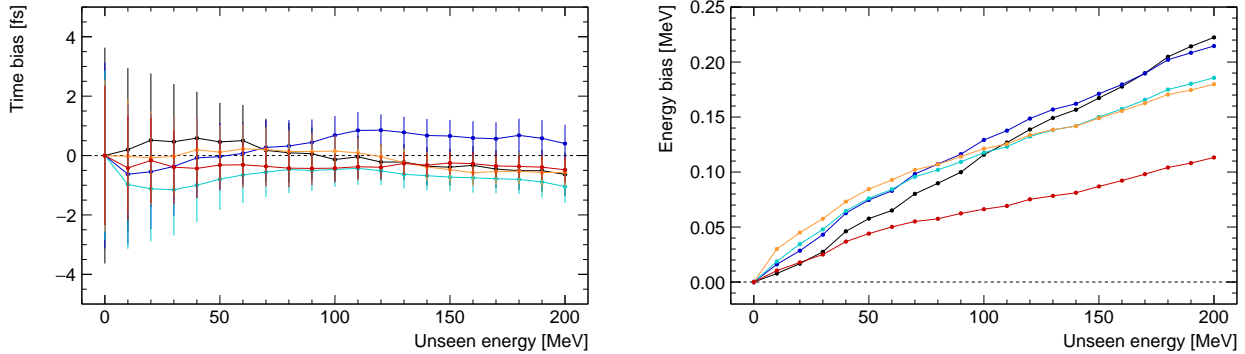
Figure 6.30: The net time (left) and energy (right) bias of a 1.7 (black), 1.9 (blue), 2.1 (cyan), 2.3 (orange), and 2.5 (red) GeV seen cluster for different unseen cluster energies.

separations correspond to those in the tail of the unseen pulse's template.

The amount of incomplete cancellation is estimated by a Riemann sum over the sampled time separations for each unseen energy. As shown in Fig. 6.30, the seen pulse's time is unbiased, while the seen pulse's energy is linearly biased by about 0.1 % of the unseen energy. For an unseen energy as large as 200 MeV, however, the net energy bias is only 0.25 MeV. As also shown in Fig. 6.30, repeating the above procedure with seen pulses of energies increasingly higher than 1.7 GeV, the net time and energy bias remains negligible and (roughly) linear, respectively. The net energy bias' slopes, though, decrease with increasing seen energy—most likely because of the increasing energy resolution per ADU.

Such a non-zero net energy bias can introduce an ω_a-bias by mimicking an uncorrected gain variation that has the time dependence of pileup. While the in-fill gain correction (Sec. 4.6) should remove part of the bias (as unseen pulses will also overlap with the laser pulses used), such bias mitigation should be minimal because the amplitude of the laser pulses is large. Furthermore, an even smaller net energy bias is expected since the unseen and seen clusters must overlap in space—as well as time—to introduce a non-zero energy bias. Notwithstanding, the above Monte Carlo simulation has established that the presence of pileup with an undetected cluster can plausibly lead to an ω_a-bias by changing, on average, the reconstructed-cluster energies across the fill.
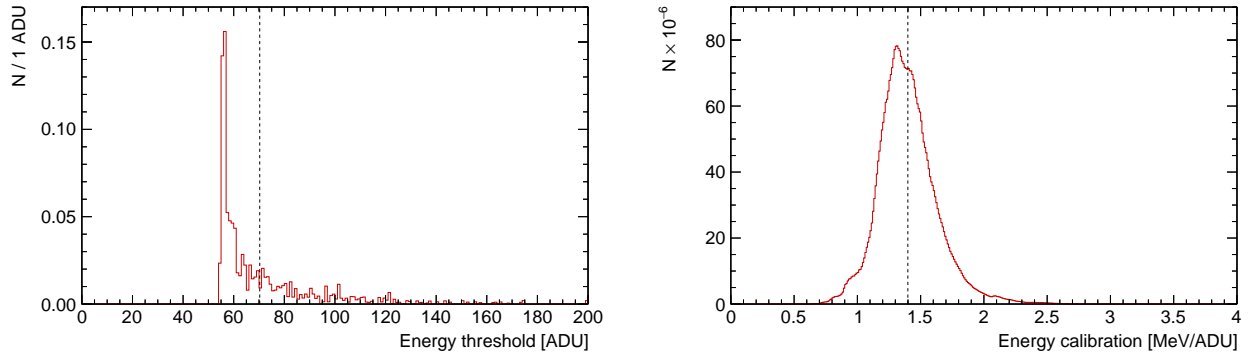
Figure 6.31: Left: The distribution of the reconstruction's uncalibrated energy threshold to both fit and keep a cluster. These data are from the 1a data set, where each possible cluster location is weighted by its relative hit rate. For each location, all of the deposited energy is assumed to be in the central crystal. Right: The distribution of energy calibrations in the 1a data set, empirically found by the ratio of each cluster's amplitude sum to calibrated energy. The distributions' mean values are denoted by dashed lines.

Consequently, a systematic uncertainty on ω_a arising from it will be estimated.

6.6.2 Energy Threshold

In the low-energy regime, the number of particles that are detected depends on the range of energy thresholds in the reconstruction used. This detection efficiency was characterized in simulation in Sec. 4.8. However, it will not be exactly the same as in the measured data due to differences in the gain corrections applied across Run 1. For that reason, each data set's efficiency as a function of energy needs to be characterized separately.

As detailed in Sec. 4.5, the reconstruction's energy threshold is initially set by fit algorithm in two steps. Every new cluster must first pass a threshold to estimate if there is enough activity to warrant a fit attempt. Each fitted cluster must then pass a different threshold to determine if it is distinguishable from a noise cluster. Both of these thresholds depend on the crystals' noise levels and thereby on the region of the calorimeter. In general, the corresponding spread of thresholds to keep a cluster has been found to be wider than that of thresholds to fit a cluster, but both thresholds' distributions have a mean around 60 ADU [136]. As shown in Fig. 6.31, however, the
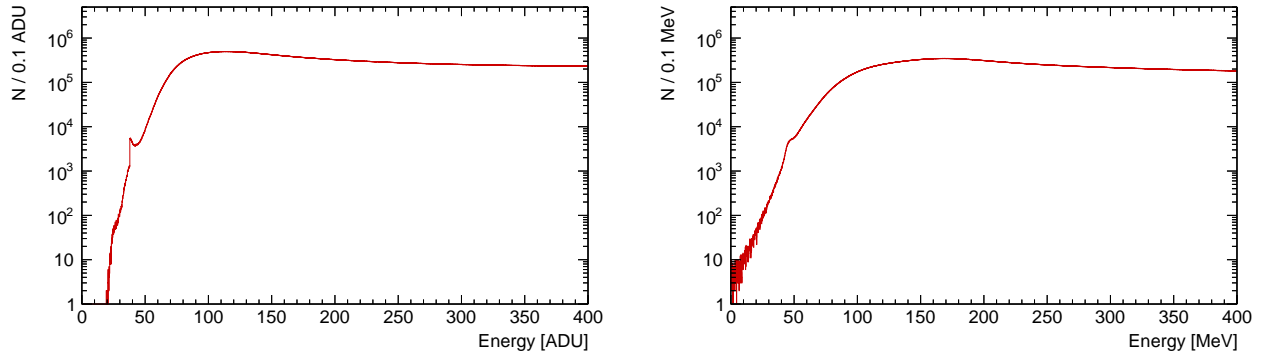
Figure 6.32: The uncalibrated (left) and calibrated (right) energy distribution for the 1a data set. The local excess near 40 ADU and 45 MeV is predominately caused by the 160-MHz noise signal in calorimeter 9 (Sec. 4.9).

effective energy threshold—i.e., the maximum of these two thresholds for every possible cluster location—has a higher mean threshold of roughly 70 ADU. This is consistent with the observed knee in the accepted clusters' (uncalibrated) energy distribution, as seen in Fig. 6.32.

Further, when a cluster is accepted, its energy is relatively calibrated with the crystals' calibration constants and corrected for the various gain effects. As part of the ω_a-measurement, the cluster's absolute energy is next calibrated following the procedure in Sec. 5.3. All together, the calibrated energy distribution, as shown in Fig. 6.32, rolls over near 100 MeV, with cluster energies observed down to 1 MeV. Also note that lost muons introduce a small local excess to this distribution, as can be subtly seen, at about 150 MeV (Sec. 4.6).

In addition, the uncalibrated and calibrated energy distributions exhibit a narrow low-energy spike around 40 ADU and 45 MeV, respectively, which is predominately from calorimeter 9. This local excess is expected because, as mentioned in Sec. 6.4.1, the 160-MHz noise signal in this calorimeter is known to be larger than in the others, causing a larger fraction of noise clusters to pass the reconstruction's thresholds. From a Monte Carlo simulation similar to that in Sec. 6.6.1, the presence of such a wave packet is found not to introduce a net time or energy bias and, as a result, can be neglected with regard to its unseen pileup with detected clusters.
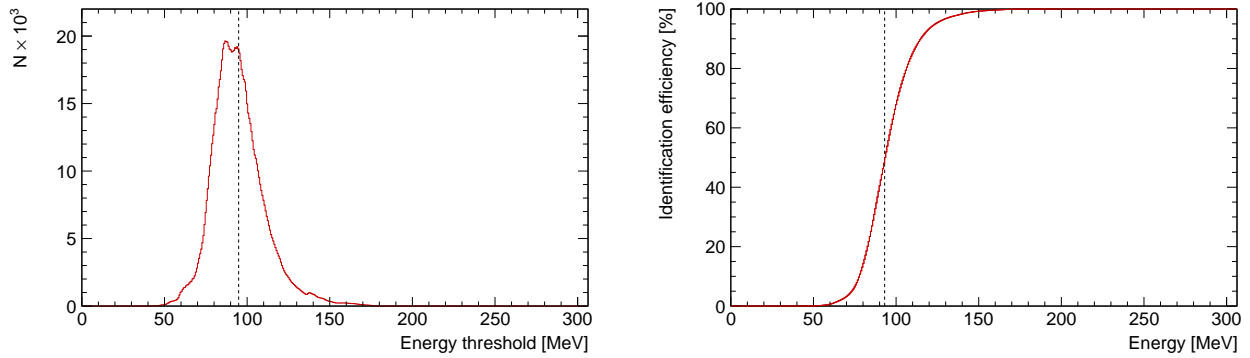
Figure 6.33: Left: The energy-threshold distribution in the 1a data set, found by the convolution of the distributions in Fig. 6.31. The mean threshold is denoted by the dashed line. Right: The efficiency of identifying clusters at different energies in the 1a data set, found by the cumulative distribution function of the normalized energy-threshold distribution. The cluster energy with a 50 % identification efficiency is denoted by a dashed line.

Moreover, the total energy calibration's distribution—empirically found by dividing each cluster's amplitude sum by its calibrated energy—is also shown in Fig. 6.31. The mean calibration is found to range between 1.4 and 1.7 MeV/ADU among the data sets, and the distribution is wider than the relative energy calibration's distribution mostly due to the long-term gain effect [136]. The long-term gain correction, in particular, introduces calibration constants down to zero along with a wide high-calibration tail for calorimeter 5. The large values for this calorimeter are primarily caused by crystal 25, whose gain significantly dropped sometime after being installed—as seen in Fig. 6.24. Overall, the longer-term gain correction increases the energy, on average, by around 20 %. The in-fill and short-term double-pulse gain corrections, on the other hand, do not significantly alter the total energy calibration distribution.

The reconstruction's effective energy threshold can then be calibrated by convoluting the two distributions in Fig. 6.31. Among the data sets, the resulting energy-threshold distribution, as seen in Fig. 6.33, has a mean between 95 and 118 MeV and a width of about 20 MeV. Further, each energy-threshold distribution can be turned into an efficiency by recognizing that a cluster detected at one threshold will, by definition, also be detected at a higher threshold. With that in

mind, the efficiency of detecting a cluster of a given energy for each data set was obtained via the cumulative distribution function of the normalized energy-threshold distribution. As depicted in Fig. 6.33, the cluster energy with a 50 % identification efficiency ranges between 93 and 115 MeV among the data sets. This energy is higher than that found in simulation (Sec. 4.8), as expected due to the long-term gain effect caused by the temperature drift over Run 1. In any case, having benchmarked the reconstruction's identification efficiency for each data set, the expected energy distribution of unseen particles can now be estimated.

6.6.3   Simulation

As unseen pileup involves undetected particles, the low-energy regime of particles that hit a calorimeter must be characterized in simulation. To that end, the same full-scale simulation employed in Sec. 4.8 was used to track secondary particles from 90 million muon decays in the storage ring. The resulting distribution of low-energy particles that hit anywhere on a calorimeter is shown in Fig. 6.34. Below 400 MeV, the expected decay positrons account for only about 1.1 % of all such particles, with most decay positrons having an energy less than 10 MeV. Of those hits not attributed to decay positrons, about 46.9 % are electrons, 46.5 % are photons, and 6.6 % are positrons—all of which are from decay positrons interacting with material, such as kicker or quadrupole plates. Aside from some small structure in the positrons' distribution, the number of particles hitting the calorimeter increases rapidly at low energies starting around 100 MeV.

To obtain the energy distribution of the particles that should be reconstructed, the simulation's energy distribution for all particles (Fig. 6.34) can be weighted by the characterized efficiency (Fig. 6.33). This predicted distribution, as illustrated in Fig. 6.35, is fairly consistent with that actually measured in Fig. 6.32. For example, both energy distributions begin to roll over at around 100 MeV. However, below about 40 MeV, more clusters are found in data than in simulation.
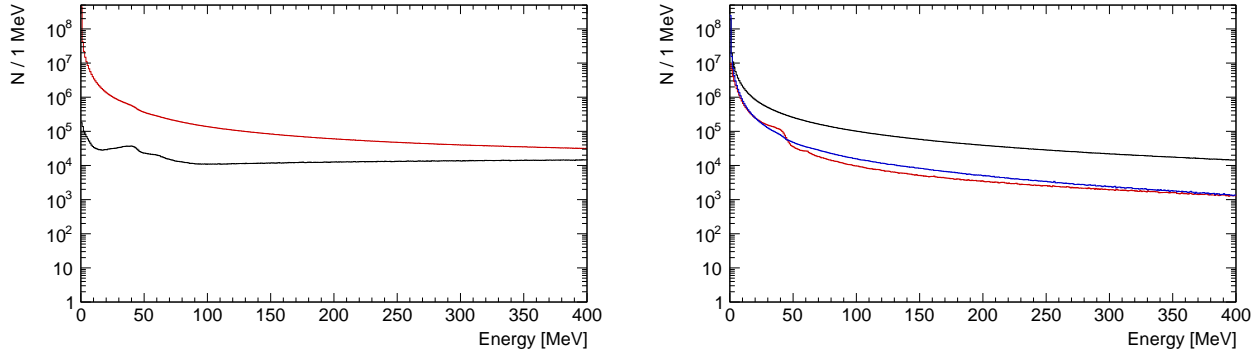
Figure 6.34: The energy distribution of different particles that hit a calorimeter in the full-scale simulation (Sec. 4.8). Left: The distribution of all particles (red) and decay positrons (black). Right: The distribution of gammas (black), non-decay positrons (red), and electrons (blue).

Some excess is expected, though, because the simulation used neglects such phenomena as noise, lost muons, and false pileup. The measured energy is also smeared by a non-zero energy resolution, being about 25–26 MeV for cluster energies of 50–100 MeV (Eq. 4.38).

6.6.4 Systematic Uncertainty Estimation

With an estimate of the unseen particles' energy distribution in Sec. 6.6.3, the systematic uncertainty from their pileup with seen particles can be assessed. The approach, in general, is to artificially increase the amount of unseen pileup by not correcting for seen pileup below an artificially high energy threshold, thereby increasing any ω_a-bias from it. To do that, in the pileup construction (Sec. 5.4), pileup candidates were excluded if any of their matched islands has a total cluster energy below an artificial energy threshold, E_a. For example, when $E_a = 300$ MeV, this procedure reduces the number of two- and three-window matches by about 10 and 20 %, respectively. It also causes the best-fit R to shift by $\mathcal{O}(1$ ppb) (with a commensurate allowed statistical drift). As long as E_a is reasonably higher than the energy thresholds found in the data set (Fig. 6.33), this change in R could be interpreted as the unseen pileup's systematic bias.

However, as seen in Fig. 6.30, the influence from such artificially higher-energy unseen clusters
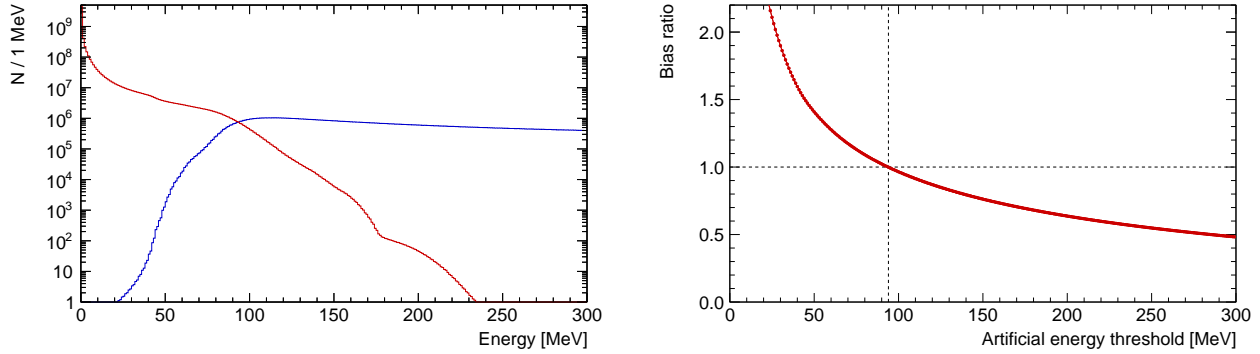
Figure 6.35: Left: The energy distribution of seen (blue) and unseen (red) clusters predicted for the 1a data set, found by weighting the simulation's distribution in Fig. 6.34 (scaled by 10 for about the same statistics as the 1a data set) by the characterized efficiency in Fig. 6.33. Right: The bias ratio (Eq. 6.5) at different artificial energy thresholds for the 1a data set.

should be larger than that from the true, lower-energy unseen clusters. To find a more realistic bound on the unseen pileup's ω_a-bias, the simulation's unseen energy distribution can be used. In particular, since the seen cluster's energy bias scales roughly linearly with unseen energy, the energy distribution of unseen clusters can likewise be linearly weighted by their energy to create a proxy for their contribution to ω_a's bias at a given energy. Integrating this proxy then yields a metric for the total amount of ω_a-bias anticipated from the unseen-energy distribution. Further, the ratio of such a metric between the simulation's distribution of unseen clusters and all clusters up to E_a then defines a *bias ratio* given by

$$\eta = \int_0^\infty E \cdot \left[\widetilde{N}(E)\right] \mathrm{d}E \bigg/ \int_0^{E_a} E \cdot \left[N(E) + \widetilde{N}(E)\right] \mathrm{d}E \tag{6.5}$$

where $N(E)$ and $\widetilde{N}(E)$ are the seen and unseen energy distributions, respectively, as depicted in Fig. 6.35. Note that, in the integrals above, the PbF$_2$ crystals' Čerenkov threshold of 102 keV is considered to be negligible [38].

For a given E_a, this bias ratio can be multiplied by the change in R to yield a more realistic systematic uncertainty from unseen pileup. As presented for the 1a and 1c data sets in Fig. 6.36, this uncertainty estimate is roughly constant for a wide range of E_a from 10 to 290 MeV. It must
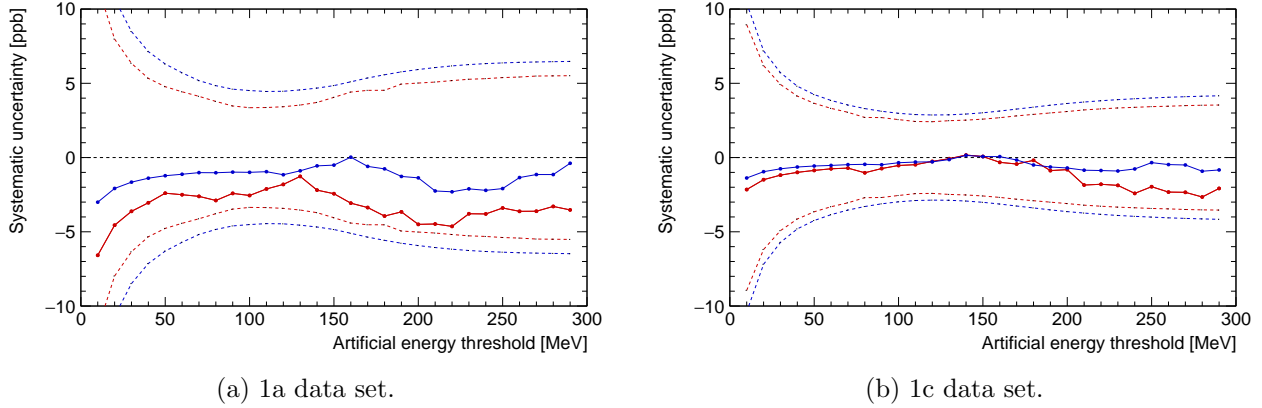
(a) 1a data set.

(b) 1c data set.

Figure 6.36: An estimate of the systematic uncertainty on ω_a from unseen pileup when using different artificial energy thresholds in the (a) 1a and (b) 1c data set, using the baseline (red) and asymmetry-weighted (blue) method. The dashed bands for each weighting method show the allowed drift due to statistical differences (Eq. 5.44).

be noted, however, that the uncertainties among the E_a values are correlated because a pileup candidate that was neglected at one E_a will, by definition, be also neglected at any higher E_a. In any case, the small variation in the systematic uncertainty for decreasing E_a—when the simulation's low-energy distribution is relied upon most—boosts confidence in the simulation and, in general, in this section's approach. It must also be noted that, for every data set, all of these estimates are less than 1.1 standard deviations of the allowed deviation due to different statistics (Eq. 5.44). But, as a systematic bias is expected, a conservative view will be taken by attributing the change in R entirely to an ω_a-bias. To partially separate the statistical from systematic bias (and rely somewhat less on the simulation), this dissertation's systematic uncertainty is taken as the average of the uncertainties found for E_a between 150 and 300 MeV in steps of 10 MeV, which is reported for each data set and weighting method in Tab. 6.7.

Two key assumptions, however, went into these systematic uncertainty estimates. First, there is an uncertainty in the absolute energy calibration, which, as discussed in Sec. 5.3, was only intended to be approximate. This absolute calibration is essential for a direct comparison between the measured energy and the simulation's true energy. Based on the energy spectrum's endpoint,

Uncertainty source	$w(E) = 1$				$w(E) = A(E)$			
	1a	1b	1c	1d	1a	1b	1c	1d
Coincidence window	0.9	2.5	1.3	0.2	0.7	2.6	1.2	0.4
Statistics	1.5	1.0	0.7	1.0	1.6	1.0	0.7	1.0
Detection efficiency	1.9	1.5	6.6	2.6	1.1	2.1	6.1	2.5

Table 6.9: A summary of the ω_a-systematic uncertainties, in ppb, attributed to muon loss for each data set in Run 1. The combined uncertainties are shown in Tab. 6.1. Note that, as discussed in Sec. 6.7.5, a separate systematic uncertainty arising from potential phase correlations is conservatively projected to be 100 ppb for each data set in Run 1 until the dedicated analyses are completed (and a possible correction is applied).

though, the calibration constants applied are certainly accurate to 10 %. This bound leads to a maximum uncertainty in the method's estimate of only 0.2 ppb. Second, the ω_a-bias was assumed to be linear with unseen energy. As seen in Fig. 6.30, the energy bias does not scale perfectly linearly with unseen energy for all seen energies, but it can be fairly well modeled by incorporating a quadratic term. Of the seen energies analyzed, the mean quadratic coefficient is about -2×10^{-6} with a maximum value found of -4×10^{-6}. Such coefficients lead to a shift in the method's systematic uncertainty estimate of about 0.2 ppb. Together, these key assumptions do not significantly change this section's final estimates.

6.7 Muon Loss Correction

After injection, a muon can exit the storage region via a number of mechanisms—as expounded in Sec. 2.4. The rate of this muon loss was partially mitigated by the quadrupole system's scraping procedure and the choice of field index. Ultimately, of the muons stored at the fit start time of 30.2876 μs, up to 1 % are estimated to have been lost over the fill (Fig. 5.11). To correct for its ω_a-bias, the time distribution of lost muons was constructed by finding triple coincidences among adjacent calorimeters (Sec. 5.5). Then, by looking in nearby time windows, about a 20 %

correction was subtracted to account for accidental and higher-order coincidences. The corrected spectrum, $L(t)$, is incorporated into the fit model in Eq. 5.36 with an overall scale factor, κ_{loss}, describing the detection efficiency of lost muons. As κ_{loss} can float in the fit, the size of its correlation to R is already absorbed in the statistical uncertainty found via the χ^2-minimization. In this section, the remaining sources of ω_a's systematic uncertainty from the uncertainties and limitations in the muon-loss correction will be evaluated. A summary of this section's results for each data set and weighting method is provided in Tab. 6.9.

6.7.1 Covariance Matrix

Because no energy cut is applied when constructing the accidental background in $L(t)$, some positrons will appear in both the muon-loss correction and the ω_a-fit. This introduces a small correlation that, in principle, requires a covariance matrix akin to the one found for the pileup correction (Sec. 5.4.3). By neglecting such a correlation in Ch. 5, a systematic bias to the best-fit R is possible and needs assessment.

For each data set, comparing the relative sizes of the terms in Eq. 5.26, the accidental background is found to be dominated by occurrences of a true double coincidence and an unrelated positron hit. The rate of this background, therefore, should approximately follow the lost muon's rate times the decay positron's rate. As there are fewer lost muon than positron hits, this rate is expected to be lower than that of two positron hits. Further, as the characteristic time of flight of $\alpha \approx 6$ ns is small relative to the boosted muon lifetime, this rate should also be lower than that of double pileup—but possibly larger than that of triple pileup. Moreover, as discussed in Sec. 6.5.2, incorporating the pileup correction's covariance matrix due to double and triple pileup changes the best-fit R by $\mathcal{O}(0.1\text{ ppb})$ and $\mathcal{O}(0.001\text{ ppb})$, respectively. Hence, neglecting the correlations induced by the accidental subtraction should introduce a bias of less than 0.1 ppb and can there-

fore be safely ignored. Accordingly, no systematic uncertainty on ω_a from it is assigned.

6.7.2 Coincidence Window

As discussed in Sec. 5.5.1, particles other than muons can be stored in the ring and subsequently lost. In particular, coincidences from lost deuterons and protons have been empirically found in each data set to varying extents (Fig. 5.9). To exclude these contaminants—and, more generally, any lost particle heavier than the muon—only coincidences with a time separation of $\delta t \in (\alpha - 1.2 \text{ c.t.}, \alpha - 0.2 \text{ c.t.})$ were used to construct $L(t)$. This δt-cut, however, has two main downsides. First, the statistics loss increases the muon-loss correction's statistical uncertainty—which will assessed directly in Sec. 6.7.3. Second, pileup between a lost muon and positron can pull the cluster time outside the δt-window, causing the lost muon to be missed. This leads to a time-dependent distortion of $L(t)$ and, thus, a potential ω_a-bias.

The influence of the chosen δt-window on the best-fit R, therefore, must be assessed. To that end, the upper bound for δt was scanned from $\alpha - 0.2$ c.t. to $\alpha + 1.2$ c.t. in steps of 0.1 c.t. The change in R and χ^2 for each one is shown in Fig. 6.37. For the 1a, 1b, and 1c data set, the χ^2 is fairly insensitive to the time cut. In contrast, for the 1d data set, the χ^2 significantly increases as the upper time cut is increased. This is expected because the 1d data set is observed to have by far the largest deuteron and proton contamination, causing an obvious distortion to $\Lambda(t)$—as shown in Fig. 5.12. Notwithstanding, for every data set, R does not change by more than 6 ppb among the different δt-windows when using either weighting method. In other words, R appears to be fairly insensitive to the tradeoff between the aforementioned effects. Thus, the pileup concern when using a smaller δt-window, along with any residual deuteron or proton contamination, is estimated to be minimal. Without additional input from simulation on the relative ω_a-bias expected from these two effects, the systematic uncertainty on ω_a from the δt-window used in Ch. 5
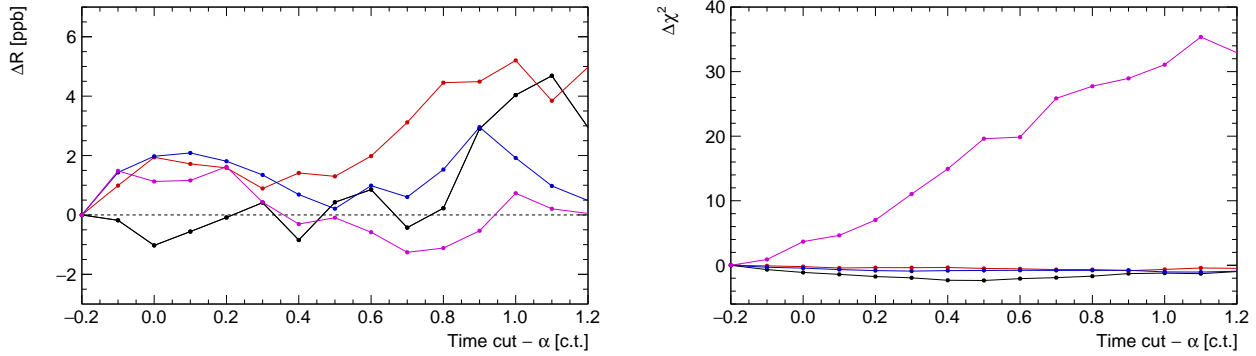
Figure 6.37: The change in R (left) and χ^2 (right) when increasing the δt-window's upper bound from the value chosen for this dissertation's ω_a-measurement. These results are for the baseline method, but those for the asymmetry-weighted method are very similar.

is taken to be the average of the changes to R in the above scan. The result for each weighting method and data set is provided in Tab. 6.9.

6.7.3 Statistics

The distribution of lost muons, $L(t)$, is found by statistically sampling the reconstructed clusters for coincidences among adjacent calorimeters. Consequently, the muon-loss correction will always have a statistical uncertainty from the finite size of the data set analyzed. This uncertainty can also be enhanced by choosing a small δt-window. Further, as $L(t)$ is formed from 20 underlying distributions (Eq. 5.26)—each with Poisson fluctuations—such a statistical uncertainty could be larger than one might intuit.

To assess the corresponding ω_a-bias, for each underlying distribution, the number of counts in each time bin was replaced by a random number drawn from a Poisson distribution with a mean of the measured number of counts. Afterward, the ω_a-fit was performed as usual. This procedure was repeated using 1000 unique random seeds of the random number generator—the same generator used across Ch. 5. The resulting distribution of R is presented in Fig. 6.38, which closely follows a normal distribution. Among the data sets, as shown in Tab. 6.9, this distribution's population
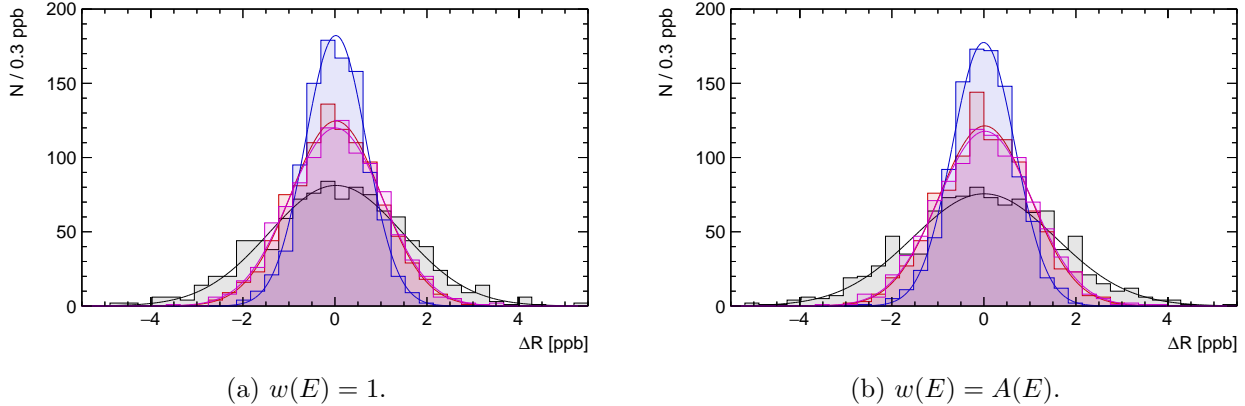
(a) $w(E) = 1$.

(b) $w(E) = A(E)$.

Figure 6.38: The distribution of the change in R from its nominal value in Tab. 5.6 for 1000 unique seeds of the random number generator used to add statistical fluctuations in the underlying distributions forming $L(t)$ in Eq. 5.26. These data are for the 1a (black), 1b (red), 1c (blue), and 1d (magenta) data set, using the (a) baseline and (b) asymmetry-weighted method. For each data set, a fit to a normal distribution is overlaid.

standard deviation ranges from 0.7 to 1.6 ppb, which is taken to be the systematic uncertainty on ω_a from statistical fluctuations in $L(t)$.

6.7.4 Detection Efficiency

A key assumption in deriving the muon-loss correction (Eq. 5.29) is that the approach's detection efficiency of lost muons remains constant over the fill—i.e., that κ_{loss} is constant in time. The extent to which this is true largely depends on the mechanism by which muons are lost. After the quadrupole system's scraping procedure, simulations indicate that muons are predominately lost due to the storage ring's circular—instead of square—collimators [137]. More precisely, the to-be-lost muons' horizontal- and vertical-betatron motion (Sec. 2.4) cause them to move occasionally across the edge of the collimator's aperture. The copper collimators, though, are only located in a few azimuthal positions. These to-be-lost muons, therefore, can survive for up to thousands of turns around the ring before eventually scattering off a collimator and being lost. Thus, as the collimators are fixed in place, the probability over a cyclotron period that such a lost muon will

create a triple coincidence should indeed remain constant across the fill.

However, as discussed in Sec. 5.6, the recovery of the quadrupole plates from their scraping to nominal voltages was prolonged in Run 1. This caused the field index and, thus, the muons' betatron motion to vary even after 30 μs into the fill (Figs. 5.18 and 5.40). This in turn might cause the distribution of where the to-be-lost muons are scattered to vary slightly over the fill. As a result, the probability of a lost muon being detected, on average, may change for at least two reasons. First, the lost muon's typical trajectory could change such that it can only traverse one or two calorimeters. Second, the lost muon's typical trajectory could change which crystals it traverses. As characterized in Sec. 6.6.2, the energy threshold for some crystal–SiPM pairs is above the lost muons' mean energy (Fig. 4.27). In the 1c data set, for example, 1.5 and 0.2 % of the crystals are expected to have a calibrated energy threshold above 150 and 170 MeV, respectively. Therefore, if a lost muon's trajectory shifts slightly, it is possible that it would no longer be identified as a triple coincidence.

To ascertain whether the prolonged scraping is actually a concern, more simulations are required. Until such time, imagined scenarios like those above motivate trying to assess any associated ω_a-bias by allowing a time dependence in κ_{loss}. To that end, κ_{loss} was replaced in the nominal fit model by $\kappa_{\mathrm{loss}} + \xi_1 \cdot t$, where ξ_1 is a new free parameter. In the 1b and 1c data set, the best-fit ξ_1 is statistically consistent with zero, but, in the others, it is a few standard deviations away. The change in R, though, is at most only 6.6 ppb, as shown in Tab. 6.9. It must be noted, however, that any change in R could be caused by another slowly varying phenomenon, such as the residual gain variation (Sec. 6.3.3). Without a predicted model for $\kappa_{\mathrm{loss}}(t)$, adding further terms will likely only increase the chance of fitting to something other than a changing lost-muon detection efficiency. Therefore, for this dissertation, the systematic uncertainty on ω_a from assuming κ_{loss} is a constant will conservatively be taken as the change in R when fitting with ξ_1.

6.7.5 Phase Correlations

In the muon-loss correction, the muons that are lost are tacitly assumed to have the same average spin direction as those observed by decay. A pernicious systematic ω_a-bias will arise, therefore, if the probability for a muon to be lost correlates with its initial phase at injection, ϕ_0. In this case, as muons are lost over the fill, the average initial phase of the remaining, stored muons will change, i.e., $\mathrm{d}\langle\phi_0\rangle/\mathrm{d}t \neq 0$. This poses a direct bias to ω_a according to Eq. 2.31, which is *not* corrected for in $\Lambda(t)$. The dedicated analyses to estimate the size of such a bias, however, are still ongoing for the Run 1 data sets as of this writing (Dec. 2019). In general, these analyses entail simulations of the experiment's muon production, injection, and storage to understand how the muons' phase space at injection correlated with the probability of being lost.

As a concrete example, such an ω_a-bias will occur if the loss probability is found to be momentum-dependent, causing the average momentum of stored muons to change across the fill—i.e., $\mathrm{d}\langle p\rangle/\mathrm{d}t \neq 0$. This bias arises because, as $\omega_c \propto 1/\gamma$ (Eq. 2.1), higher-momentum muons will spend a relatively longer time circulating around FNAL's delivery ring (Sec. 2.2). As a result, their spin directions will have precessed more in the delivery ring's magnetic field before being injected into the experiment's storage ring. A correlation, therefore, exists between the injected muons' initial phase and momentum—i.e., $\mathrm{d}\langle\phi_0\rangle/\mathrm{d}\langle p\rangle \neq 0$—which has been confirmed in simulation and data [137, 138]. Together, the fractional ω_a-bias can be approximated by

$$\frac{\Delta\omega_a}{\omega_a} = \frac{1}{\omega_a} \cdot \frac{\mathrm{d}\langle\phi_0\rangle}{\mathrm{d}\langle p\rangle}\frac{\mathrm{d}\langle p\rangle}{\mathrm{d}t} \tag{6.6}$$

to leading order. The magnitude of the ω_a-bias will therefore, in part, depend on the lost muons' time distribution, $L(t)$, along with its absolute scale determined by κ_{loss}. Accordingly, it is necessary that the best-fit κ_{loss} is not compensating for any uncorrected effect, such as a residual gain variation, so that it has physical meaning. For this reason, it must be noted that the best-fit κ_{loss}

changes from its nominal value found in Ch. 5 between 6 and 22 % among the data sets when applying the optimal residual gain correction (Sec. 6.3.3).

In the above example, a worst-case scenario—where all lost muons are assumed to come from the high- or low-momentum tail of the distribution—is estimated to bias ω_a by up to 200 ppb [91, 138]. However, preliminary results from the aforementioned, ongoing analyses suggest that the ω_a-bias is more realistically of $\mathcal{O}(10$ ppb$)$ [137]. Based on these available results, for the purpose of this dissertation, the systematic uncertainty on ω_a from such phase correlations will be conservatively projected to be 100 ppb for every data set. Once these analyses are finished, if the uncertainty is unacceptably large, a correction for the bias will be made. The systematic uncertainty on ω_a would then be reduced to that from the correction's uncertainty. Even in the above worst-case scenario, though, this systematic uncertainty is small compared to the statistical uncertainty on ω_a for Run 1 (Tab. 5.6).

6.8 Muon Dynamics

As discussed in Secs. 2.4 and 5.6, the muon dynamics in the storage region can introduce, via the calorimeters' acceptance, oscillations at characteristic frequencies in the number of reconstructed positrons. Each of these frequencies was either removed by randomizing the muon fills' injection times or modeled in the fit function. As expounded in Sec. 5.6, the modeling used was motivated by describing the muons' spatial distribution in terms of its raw moments. Using the calorimeter and tracker data as guides, the form of each moment was postulated to explain every evident frequency in the fit residuals' FFT. Note that, as all of the added parameters float in the fit, their correlation to R is absorbed in the statistical uncertainty given in Tab. 5.6. Any dynamics absent from the analysis are considered to have a negligible bias on ω_a for the size of the data sets in Run 1. In this section, the dominant sources of systematic uncertainty on ω_a from such treat-

Uncertainty source	$w(E) = 1$				$w(E) = A(E)$			
	1a	1b	1c	1d	1a	1b	1c	1d
CBO frequency change	3.9	9.6	11.0	9.6	4.7	11.7	12.6	11.1
CBO decoherence	23.8	4.8	1.6	12.1	25.0	4.0	5.2	8.7
CBO time constants	7.8	44.9	6.5	2.7	3.2	52.8	7.5	3.3

Table 6.10: A summary of the ω_a-systematic uncertainties, in ppb, attributed to muon dynamics for each data set in Run 1. Note that, as described in Sec. 6.8.3, an additional systematic uncertainty from phase correlations' correction will be added once the related analyses are completed. The combined uncertainties are shown in Tab. 6.1.

ment of the muon dynamics will be assessed. A summary of this section's results for each data set and weighting method is provided in Tab. 6.10.

6.8.1 Time Randomization

The oscillation at the cyclotron frequency f_c—referred to as the fast-rotation signal—was partially removed by binning the time histogram in the approximate cyclotron period (Fig. 5.15). However, the residual fast-rotation signal is large enough to noticeably bias the best-fit R among the calorimeters (Fig. 5.17). Therefore, the signal was further smeared out by artificially randomizing each fill's injection time according to a uniform distribution between $-149.2/2$ and $+149.2/2$ ns (Sec. 5.6.1). The implementation of the RANLUX random number generator with the highest luxury level employed (across this dissertation's analysis) is widely used in the physics community and is known to have excellent performance. Thus, the generator itself is not a concern.

Moreover, this time-randomization approach is limited by the implicit assumption that the cyclotron period is exactly 149.2 ns. As shown in Fig. 5.16, the muon momentum spread results in cyclotron periods ranging from about 148.9 to 149.7 ns. To check for a systematic ω_a-bias from assuming a fixed period, for each fill i, a random cyclotron period, $T_{c,i}$, was drawn based on the measured distribution in Fig. 5.16. Each fill's injection time was then offset by a random num-

ber drawn from a uniform distribution between $-T_{c,i}/2$ and $+T_{c,i}/2$. Note that the 1b data set's cyclotron-period distribution has not been measured as of this writing (Dec. 2019), but, as seen in Fig. 5.16, the other data sets' distributions have comparable spreads. Therefore, as an approximation, the 1a data set's distribution was used for the 1b data set's test. As shown in Tab. 6.11, the average change in R over 2000 seeds of the random number generator is consistent with zero to less than 1.1 standard deviations among the data sets. This is a stringent test, with the largest data set's standard deviation being less than 2 ppb. In addition, no statistically significant variation in the average change in R is found across the calorimeters.

As another check, the time-randomization method was changed to offset the clusters' times rather than the fills' times. More specifically, for each cluster, its reconstructed time was shifted by a random number drawn from uniform distribution between $-149.2/2$ and $+149.2/2$ ns. In this approach, some of the natural cancellation of muon-dynamics oscillations between calorimeters 180 ° apart (Fig. 5.33) will no longer be preserved. Nonetheless, any oscillation near the cyclotron frequency ought to be smeared out more—compared to randomizing the fills' injection times. As shown in Tab. 6.11, using the same 2000 seeds previously mentioned, the average change in R is consistent with zero to less than 1.5 standard deviations for each data set. Therefore, as all three time-randomization methods yield consistent results, no systematic uncertainty on ω_a from the choice of method will be assigned.

Further, when using the time-randomization approach, the need for a time-bin width approximately equal to the cyclotron period is removed—as demonstrated in Fig. 5.15. Accordingly, the best-fit R should be unbiased by a different choice of time-bin width. However, as discussed in Sec. 5.6, different choices will shift the observed frequencies—some of which might no longer be aliased. The modeling of the muon dynamics, therefore, would need be changed appropriately. Also, the best-fit R's sensitivities—such as to gain inaccuracies—are likely to depend on the bin-

Difference	Weighting method	Data set			
		1a	1b	1c	1d
$R_{\text{cluster}} \quad - R_{\text{fill}}$	$w(E) = 1$	0.7 ± 4.0	-6.2 ± 4.2	0.5 ± 2.6	0.8 ± 1.8
$R_{\text{cluster}} \quad - R_{\text{fill}}$	$w(E) = A(E)$	-1.1 ± 3.6	-4.8 ± 3.9	0.6 ± 2.4	1.4 ± 1.7
$R_{\text{distribution}} - R_{\text{fill}}$	$w(E) = 1$	1.1 ± 3.9	-4.3 ± 4.2	-1.0 ± 2.6	-0.5 ± 1.9
$R_{\text{distribution}} - R_{\text{fill}}$	$w(E) = A(E)$	2.4 ± 3.6	-3.3 ± 3.9	-2.3 ± 2.4	-0.2 ± 1.7

Table 6.11: The mean difference in R over 2000 seeds of the random number generator among when time randomizing each fill's injection time between $-149.2/2$ and 149.2 ns, R_{fill}; when time randomizing each fill's injection time between $-T_{c,i}/2$ and $T_{c,i}/2$ where $T_{c,i}$ is randomly drawn for each fill from the data set's measured distributions of cyclotron periods, $R_{\text{distribution}}$; and when time randomizing each cluster's time between $-149.2/2$ and 149.2 ns, R_{cluster}.

ning (Figs. 6.1 and 6.4). All together, to identify if two R values found with different time-bin widths are consistent, a full assessment of their systematic uncertainties is required.

Nonetheless, a consistency test was done based only on R's statistical uncertainty by changing the time-bin width from 149.2 to 149.0 ns for the 1a data set. Using the baseline method, the resulting change in R is consistent with zero to 1.1 standard deviations of the allowed statistical drift. Further, for the 1c data set, the time-bin width was changed to $149.2/10$ ns. This required that the model be appropriately updated and that the fit stop time be smaller to ensure that more than 30 counts still remained in every time bin in the χ^2. However, even with such complications, the change in R using the baseline method is consistent with zero to 0.2 standard deviations of the allowed statistical drift. Together, these tests further verify that the time-randomization method used in Sec. 5.6 sufficiently smears out the fast-rotation signal.

6.8.2   Coherent Betatron Oscillation

By far the largest modulation from muon dynamics is the coherent betatron oscillation (CBO), as evidenced by Fig. 5.13. Among the data sets, including the CBO modulations on N_0, A_0, and ϕ_0 shifts the best-fit R as much as 0.5 ppm. This shift is larger than that when removing the gain,

pileup, or muon-loss correction, motivating a careful assessment of the systematic uncertainty on ω_a from any modeling inaccuracy. In general, imperfect modeling of the CBO (and the vertical oscillations) can be introduced by the frequency's time dependence, the decoherence envelope, or any terms that are absent. These potential sources of uncertainty will be discussed in turn below and together will be shown to contribute less than 55 ppb to ω_a's uncertainty.

As discussed in Sec. 5.6.3, the only parameters fixed in the model were A_1, τ_1, A_2, and τ_2, which collectively define the time dependence of ω_{CBO} in the fill (Eq. 5.41). For each data set, their values—given in Tab. 5.5—were found by averaging those from fits to the two straw trackers' data. As a result, each parameter has associated uncertainty that is not already accounted for in the fit's statistical uncertainty. In addition, note that, while the best-fit values of A_1, τ_1, A_2, and τ_2 are fairly precise, there is a comparable or larger deviation in their central values between the fits from each of the two trackers. To assess R's sensitivity to the $\omega_a(t)$-model, the ω_a-measurement was repeated with the values found from each tracker. For each data set, the changes in the best-fit R have opposite signs for the two trackers—but equal magnitudes to about 1 ppb. To be conservative, the estimate of the systematic uncertainty on ω_a from fixing the time dependence of ω_{CBO}, reported in Tab. 6.10, is taken as the larger of the two deviations.

Moreover, as explained in Sec. 5.6.2, the amplitudes of the muon ensemble's oscillations diminish over the fill. Motivated by tracker measurements, the analytical form of each oscillation's decoherence envelope, $D_i(t)$, was modeled by an exponential. However, as seen in Fig. 5.18, the amplitude of the CBO caused by the centroid's motion appears to level off after about 250 μs into the fill. As shown in Fig. 6.39, this behavior can be easily accounted for by adding a constant to the envelope model. Further, simulations indicate that, as faster muons lap slower muons, the phases of the muons' oscillations will begin to interfere constructively again. This should lead to a periodic modulation of the ensemble's decoherence envelope, which, as shown in Fig. 6.39, can
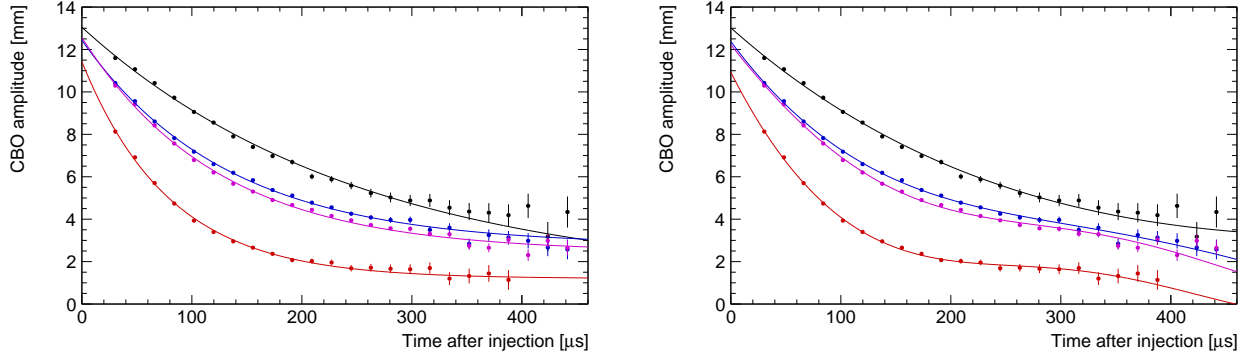
Figure 6.39: The amplitude of the f_{CBO}-oscillation for the 1a (black), 1b (red), 1c (blue), and 1d (magenta) data set. These data are based on the positrons' mean horizontal position measured by the tracker in front of calorimeter 18 [125]. A fit to the model in Eq. 6.7 (left) and Eq. 6.8 (right) is overlaid for each data set.

account for the measured amplitudes after 250 μs. These alternatives to the form of $D_1(t)$ set a reasonable bound on its potential modeling inaccuracy. To assess R's sensitivity, $e^{-kt/\tau_{\mathrm{CBO}}}$ was, therefore, replaced in Eqs. 5.27–5.40 by

$$D_k(t) = e^{-kt/\tau_{\mathrm{CBO}}} + C_e, \text{ or} \tag{6.7}$$

$$D_k(t) = e^{-kt/\tau_{\mathrm{CBO}}} \left[1 + A_e \cdot \cos\left(\omega_e \cdot t + \phi_e\right)\right] \tag{6.8}$$

for $k \in \{1, 2\}$, where the values of C_e, A_e, ω_e, and ϕ_e are fixed to the averages of their best-fit values from fits to the two straw trackers' data. The relative sign and magnitude of the change in R between these alternatives varied among the data sets. To be conservative, the systematic uncertainty on ω_a from the modeling inaccuracy of the decoherence envelope's form is taken as the largest change in R's magnitude. This result—which is reported in Tab. 6.10—is expected not to be strongly correlated with the above systematic uncertainty from $\omega_{\mathrm{CBO}}(t)$ since τ_{CBO} and ω_0 are correlated by less than 2 % in the fit. Furthermore, these systematic uncertainties are about 1 % of the change in R when neglecting the CBO altogether. Assuming this relationship holds for the vertical oscillations, the systematic uncertainty from the decoherence envelope's form in $N_y(t)$ is expected to be less than 0.1 ppb and, therefore, negligible.
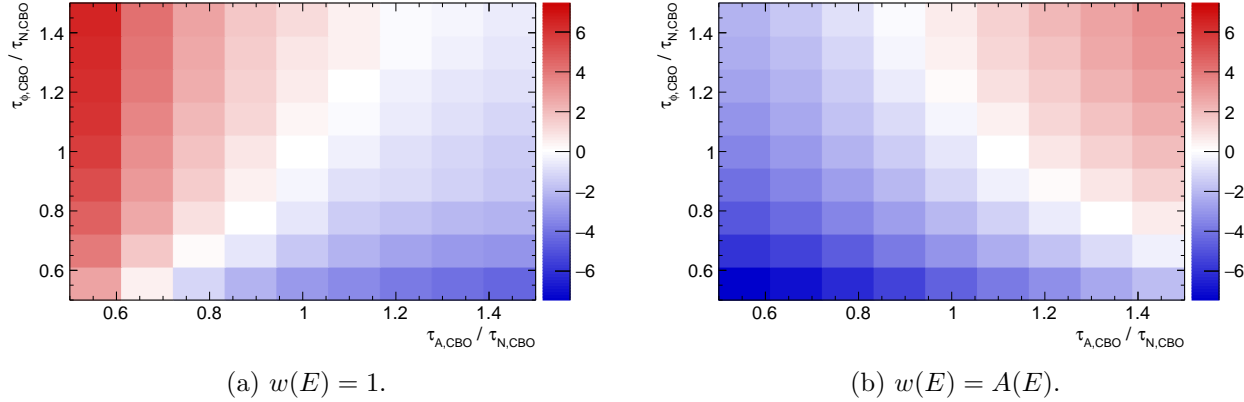
(a) $w(E) = 1$. (b) $w(E) = A(E)$.

Figure 6.40: The change in R when varying the ratios among the decoherence envelope's characteristic time constant in $N_x(t)$, $\tau_{N,\text{CBO}}$, in $N_A(t)$, $\tau_{A,\text{CBO}}$, and in $N_\phi(t)$, $\tau_{\phi,\text{CBO}}$, for the 1c data set, using the (a) baseline and (b) asymmetry-weighted method.

In addition, when using the exponential form, the characteristic time constant, τ_{CBO}, was assumed to be the same for the decoherence envelopes in $N_x(t)$, $A_x(t)$, and $\phi_x(t)$. Tracker measurements combined with the calorimeter acceptances predicted in simulation, though, indicate that the time constant in $A_x(t)$ and $\phi_x(t)$ may be lower than that in $N_x(t)$ by as much as a half [140]. This motivates letting τ_{CBO} in $N_x(t)$, $A_x(t)$, and $\phi_x(t)$ float separately in the fit. However, for the sizes of the data sets in Run 1, the CBO modulations on A_0 and ϕ_0 are too small for a reliable fit result. When allowed to float, the best-fit time constants in $A_x(t)$, and $\phi_x(t)$ increase to unreasonable values—perhaps to account for a residual gain variation (Sec. 6.3.3). For this reason, the potential inaccuracies in these time constants will be treated as a systematic uncertainty on ω_a. To that end, the ω_a-measurement was repeated—with the nominal fit parameters—fixing the ratios of the time constants in $A_x(t)$ and $\phi_x(t)$ to that in $N_x(t)$. As an example, the change in R when varying these ratios for the 1c data set is illustrated in Fig. 6.40. The corresponding, systematic uncertainty on ω_a—which is reported in Tab. 6.10—is then conservatively estimated as the largest change in R's magnitude for such ratios between 0.5 and 1.5.

The remaining, potential source of modeling inaccuracy is whether all of the relevant terms

to describe the muon dynamics were included in Eq. 5.36. In general, the fit residuals' FFT was used as a guide to account for the statistically significant frequencies in the data. When the oscillations were then enhanced by fitting each calorimeter's data, for example, no deficiencies in the modeling of the CBO modulations were revealed (Sec. 5.7.3). This supports the claim that any dynamics absent from the fit have a negligible ω_a-bias for the data-set sizes in Run 1. However, for the reasons discussed in Sec. 5.6.3, the $i = 2$ and $j = 1$ term in $N_x(t)$ was omitted from the final fit function—despite being directly observed in the tracker data. The primary consequence is that the decoherence envelope in $N_x(t)$ will be slightly distorted. When this term is added to the model with two new fit parameters, the χ^2 decreases by only about three among the data sets. The fit, therefore, appears to be insensitive to such changes to the envelope model. For that reason, the ω_a-bias from any distortions introduced by omitting the $i = 2$ and $j = 1$ term is expected to be accounted for in the above systematic uncertainty estimates.

6.8.3   Phase Correlations

As seen in Fig. 5.40, the reconstructed positrons' mean hit position is found to drift slowly downward over the fill. Further, as shown in Fig. 5.38, the initial phase at injection, ϕ_0, is found to be systematically correlated with the positrons' vertical hit position. Together, an uncorrected, time-dependent phase is possible, posing a direct systematic bias to ω_a via Eq. 2.31. This, more generally, holds for any muon-dynamics effect that couples to ϕ_0 through the calorimeter's acceptance. An illustrative example is the CBO modulation on ϕ_0, which could be observed in the fit residual's FFT, leading to the $\phi_x(t)$-model in the fit function.

As another example, the positrons' mean hit position is also found to drift slowly towards the storage region. This drift is expected to occur—even if the quadrupole system had not contained damaged/defective resistors in Run 1. Because there is a spread in the injected muons'

momenta, there is also a corresponding spread in their boosted lifetimes. This means that the stored muons' mean momentum increases over the fill. Hence, the stored muons' mean equilibrium radius and, in turn, drift time increases over the fill as well (Ch. 2). As shown in Sec. 2.6, such a time-dependent drift time poses a direct ω_a-bias, which has been estimated to be around 12 ppb [91]. The mean equilibrium radius, on the other hand, is only estimated to increase by about 200 μm over the fill. Therefore, the majority of the measured horizontal drift (Fig. 5.40) is attributed to the quadrupole system's resistors in Run 1.

To assess the ω_a-bias from these types of phase–acceptance effects that are not modeled in the fit function, a combination of calorimeter/tracker measurements and detailed simulations are required. To that end, several dedicated analyses are being developed to understand the correlations among the initial phase at injection, decay position, vertical hit position, etc. Very preliminary results appear to suggest that the total ω_a-bias from these types of effects could be as large as 0.4 ppm [141]. As this is comparable with R's statistical uncertainty, a correction for such bias is expected to be made. The systematic uncertainty on ω_a would then be estimated from the correction's uncertainty. As of this writing (Dec. 2019), it is too early to estimate a reasonable size of the correction or its uncertainty. This dissertation's results in Tab. 5.6, therefore, have not yet been corrected for these types of systematic effects. For a reference value, using the upper bound estimate of 0.4 ppm, the systematic uncertainty from the phase correlations will be 100 ppb if the future correction can be known to 25 %—which is probably achievable.

6.9   Electric Field and Pitch Correction

As mentioned in Sec. 5.8, the electric field and pitch corrections—both of $\mathcal{O}(100\ \text{ppb})$—will be applied to this dissertation's results in Tab. 5.6 once their dedicated analyses are finished. When that occurs, a systematic uncertainty on ω_a will be introduced by the corrections' uncertainties.

Example sources of the corrections' uncertainties include any misalignment and voltage errors of the quadrupole plates, which can distort the muons' closed orbit. The tracker measurements and simulations used can also introduce uncertainty. Preliminary results for some of the Run 1 data sets suggest that the total systematic uncertainty on ω_a from the electric field and pitch corrections will be around 60 and 20 ppb, respectively [131, 132, 142].

Together with this chapter's other assessments, the total systematic uncertainty on ω_a after applying all of the expected corrections is estimated to be around 160 ppb for each data set in Run 1. The systematic uncertainty is, therefore, well under control when compared to the statistical precision (Tab. 5.6). However, after taking data for four years, the experiment's systematic ω_a-precision goal is 70 ppb. Many of this dissertation's estimates are conservative, though, especially those from the phase correlations related to muon loss and dynamics. The systematic uncertainty on ω_a from the fitted anomalous-precession signal in the calorimeter—the physical observable—is only 33 ppb for the asymmetry-weighted method. If the combined uncertainty from the corrections after the fit can be improved to 60 ppb, the experiment's systematic uncertainty target will be almost reached for every data set in Run 1. In any case, the techniques established in this chapter to assess this dissertation's ω_a-measurement will provide a valuable foundation when analyzing the Muon $g-2$ experiment's future data sets.

BIBLIOGRAPHY

[1] J. Jackson, *Classical Electrodynamics* (John Wiley & Sons, 1999).

[2] G. Uhlenbeck, S. Goudsmit, Nature **117**, 264 (1926).

[3] L. Thomas, Nature **117**, 514 (1926).

[4] J. Sakurai and J. Napolitano, *Modern Quantum Mechanics* (Addison-Wesley, 2010).

[5] J. Schwinger, Phys. Rev. **73**, 416 (1948).

[6] M. Peskin and D. Schroeder, *An Introduction to Quantum Field Theory* (CRC Press, 1995).

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

[8] P. Mohr *et al.*, Rev. Mod. Phys. **88**, 35009 (2016).

[9] J. Abdallah *et al.* (DELPHI Collaboration), Eur. Phys. J. C **35**, 159 (2004).

[10] F. Jegerlehner and A. Nyffeler, Phys. Rep. **477**, 1 (2009).

[11] T. Aoyama *et al.*, Phys. Rev. D **97**, 36001 (2018).

[12] T. Aoyama *et al.*, Phys. Rev. Lett. **109**, 111808 (2012).

[13] A. Keshavarzi *et al.*, Phys. Rev. D **97**, 114025 (2018).

[14] C. Gnendiger *et al.*, Phys. Rev. D **88**, 53005 (2013).

[15] D. Hanneke *et al.*, Phys. Rev. A **83**, 52122 (2011).

[16] R. Bouchendira *et al.*, Phys. Rev. Lett. **106**, 80801 (2011).

[17] R. Parker *et al.*, Science **360**, 191 (2018).

[18] A. Kurz *et al.*, Phys. Rev. D **93**, 53017 (2016).

[19] S. Laporta, Phys. Lett. B **772**, 232 (2017).

[20] M. Davier *et al.*, Eur. Phys. J. C **77**, 827 (2017).

[21] A. Kurz *et al.*, Phys. Lett. B **734**, 114 (2014).

[22] M. Tanabashi *et al.* (Particle Data Group), Phys. Rev. D **98**, 30001 (2018).

[23] T. Blum *et al.*, Phys. Rev. Lett. **121**, 22003 (2018).

[24] R. Jackiw and S. Weinberg, Phys. Rev. D **5**, 2396 (1972).

[25] G. Colangelo *et al.*, Eur. Phys. J. Web Conf. **175**, 1025 (2018).

[26] M. Hoferichter *et al.*, Phys. Rev. Lett. **121**, 112002 (2018).

[27] H. Meyer and H. Wittig, Prog. Part. Nucl. Phys. **104**, 46 (2019).

[28] G. Colangelo *et al.*, Phys. Lett. B **735**, 90 (2014).

[29] R. Garwin *et al.*, Phys. Rev. **118**, 271 (1960).

[30] G. Charpak *et al.*, Phys. Rev. Lett. **6**, 128 (1961).

[31] J. Bailey *et al.*, Phys. Lett. B **28**, 287 (1968).

[32] J. Bailey *et al.*, Nucl. Phys. B **150**, 1 (1979).

[33] G. Bennett *et al.* (E821 Collaboration), Phys. Rev. D **73**, 72003 (2006).

[34] M. Davier *et al.*, Eur. Phys. J. C **71**, 1515 (2011).

[35] F. Jegerlehner and R. Szafron, Eur. Phys. J. C **71**, 1632 (2011).

[36] K. Hagiwara *et al.*, J. Phys. G: Nucl. Part. Phys. **38**, 85003 (2011).

[37] F. Jegerlehner, Eur. Phys. J. Web Conf. **166**, 22 (2018).

[38] J. Grange *et al.* (E989 Collaboration), in *Technical Design Report*, 2015.

[39] P. Langacker, *The Standard Model and Beyond* (CRC Press, 2010).

[40] A. Czarnecki and W. Marciano, Phys. Rev. D **64**, 13014 (2001).

[41] M. Blanke *et al.*, J. High Energy Phys. **5**, 13 (2007).

[42] T. Appelquist and B. Dobrescu, Phys. Lett. B **516**, 85 (2001).

[43] J. Miller *et al.*, Annu. Rev. Nucl. Part. Sci. **62**, 237 (2012).

[44] C. Adam *et al.*, Eur. Phys. J. C **71**, 1520 (2011).

[45] P. Fayet, Phys. Rev. D **75**, 115017 (2007).

[46] J. Batley *et al.* (NA48/2 Collaboration), Phys. Lett. B **746**, 178 (2015).

[47] G. Mohlabeng, Phys. Rev. D **99**, 115001 (2019).

[48] C. Chen *et al.*, Phys. Rev. D **93**, 35006 (2016).

[49] W. Marciano *et al.*, Phys. Rev. D **94**, 115033 (2016).

[50] D. Stratakis *et al.*, Phys. Rev. Accel. Beams **20**, 11001 (2019).

[51] G. Danby *et al.*, Nucl. Instr. Meth. A **457**, 151 (2001).

[52] D. Sweigart and R. Bjorkquist, E989 Internal Document No. 3245 (2015).

[53] M. Smith, Ph.D. Thesis, University of Washington, 2017.

[54] J. Crnkovic, E989 Internal Document No. 9191 (2019).

[55] E. Konopinski, Annu. Rev. Nucl. Sci. **9**, 99 (1959).

[56] L. Alonzi *et al.*, Nucl. Instr. Meth. A **824**, 718 (2016).

[57] J. Kaspar *et al.*, J. Instrum. **12**, 1009 (2017).

[58] W. Gohn, PoS ICHEP, 174 (2016).

[59] A. Fienberg *et al.*, Nucl. Instr. Meth. A **783**, 12 (2015).

[60] K. Khaw *et al.*, Nucl. Instr. Meth. A **945**, 162558 (2019).

[61] S. Charity, Ph.D. Thesis, University of Liverpool, 2018.

[62] G. Bennett *et al.*, Nucl. Instr. Meth. A **579**, 1096 (2007).

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

[64] F. Farley, Phys. Lett. **42**, 66 (1972).

[65] PCI Industrial Computer Manufacturers Group, MTCA.0 R1.0 Specification (2006).

[66] PCI Industrial Computer Manufacturers Group, AMC.0 R2.0 Specification (2006).

[67] D. Sweigart, E989 Internal Document No. 6497 (2017).

[68] E. Hazen *et al.*, J. Instrum. **8**, C12036 (2013).

[69] B. Taylor, IEEE Trans. Nucl. Sci. **45**, 821 (1998).

[70] A. Racz, Proc. 6 Workshop on Electronics for LHC Experiments, 405 (2000).

[71] M. Pesaresi *et al.*, J. Instrum. **10**, C03036 (2015).

[72] American National Standards Institute, VITA 57 FPGA Mezzanine Card (2008).

[73] D. Sweigart *et al.*, computer code `gm2ccc` (Redmine, 2018).

[74] D. Sweigart, E989 Internal Document No. 3678 (2017).

[75] D. Sweigart, E989 Internal Document No. 4497 (2017).

[76] C. Larrea *et al.*, J. Instrum. **10**, C02019 (2015).

[77] D. Sweigart, E989 Internal Document No. 8783 (2017).

[78] D. Sweigart, E989 Internal Document No. 4768 (2016).

[79] D. Sweigart, PoS ICHEP, 845 (2016).

[80] Texas Instruments, SLAS946B Datasheet, Rev. B (2016).

[81] D. Sweigart, E989 Internal Document No. 17305 (2019).

[82] D. Sweigart *et al.*, computer code `gm2wfdfw-mfpga` (Redmine, 2018).

[83] D. Sweigart *et al.*, computer code `gm2wfdfw-cfpga` (Redmine, 2018).

[84] D. Sweigart, computer code `gm2monitor` (Redmine, 2018).

[85] INCITS, in *Fibre Channel—Methodologies for Signal Quality Specification*, 2017.

[86] C. Green *et al.*, J. Phys. Conf. Ser. **396**, 22020 (2012).

[87] D. Sweigart *et al.*, computer code `gm2reconeast` (Redmine, 2018).

[88] D. Sweigart, E989 Internal Document No. 10051 (2018).

[89] D. Sweigart, E989 Internal Document No. 9527 (2017).

[90] G. Plunkett and A. Fienberg, E989 Internal Document No. 2369 (2015).

[91] A. Fienberg, Ph.D. Thesis, University of Washington, 2019.

[92] F. Gray, Ph.D. Thesis, University of Illinois at Urbana-Champaign, 2003.

[93] E. Jones *et al.*, computer code `SciPy` (GitHub, 2018).

[94] A. Anastasi *et al.*, Nucl. Instr. Meth. A **824**, 716 (2016).

[95] A. Anastasi *et al.*, Nucl. Instr. Meth. A **788**, 43 (2015).

[96] J. Hempstead and A. Fienberg, E989 Internal Document No. 16841 (2019).

[97] G. Guennebaud, in *Eigen: A C++ Linear Algebra Library*, 2013.

[98] K. Labe, E989 Internal Document No. 17695 (2019).

[99] H. Harter, Biometrika **48**, 151 (1961).

[100] F. James and M. Winkler, in *MINUIT User's Guide*, 2004.

[101] W. Press *et al.*, *Numerical Recipes* (Cambridge University Press, 2007).

[102] A. Lucia, in *Encyclopedia of Optimization* (Springer, 2001).

[103] P. Wieschollek, computer code `CppOptimizationLibrary` (GitHub, 2016).

[104] G. Akopjanov *et al.*, Nucl. Instr. Meth. **145**, 441 (1977).

[105] T. Awes *et al.*, Nucl. Instr. Meth. A **311**, 130 (1992).

[106] D. Sweigart, E989 Internal Document No. 3780 (2016).

[107] K. Labe, E989 Internal Document No. 17692 (2019).

[108] E. Bottailco *et al.*, E989 Internal Document No. 17827 (2019).

[109] P. Girotti *et al.*, E989 Internal Document No. 17668 (2019).

[110] S. Grant, E989 Internal Document No. 17566 (2019).

[111] K. Labe, E989 Internal Document No. 15325 (2018).

[112] D. Sweigart, E989 Internal Document No. 17248 (2019).

[113] A. Lyon *et al.*, computer code `gm2ringsim` (Redmine, 2017).

[114] S. Agostinelli *et al.*, Nucl. Instr. Meth. A **506**, 250 (2003).

[115] T. Arvanitis and A. Lyon, J. Phys. Conf. Ser. **513**, 22023 (2014).

[116] D. Sweigart, E989 Internal Document No. 3346 (2015).

[117] D. Sweigart, E989 Internal Document No. 11100 (2018).

[118] F. Gray *et al.*, E989 Internal Document No. 19907 (2019).

[119] K. Labe, E989 Internal Document No. 10995 (2018).

[120] C. Polly, Ph.D. Thesis, University of Illinois at Urbana-Champaign, 2005.

[121] L. Lönnblad, Comput. Phys. Commun. **84**, 307 (1994).

[122] J. Crnkovic *et al.*, E989 Internal Document No. 9191 (2019, to be published).

[123] A. Keshavarzi *et al.*, E989 Internal Document No. 15304 (2018).

[124] J. Crnkovic *et al.*, E989 Internal Document No. 2783 (2015).

[125] J. Mott, E989 Internal Document No. 15376 (2018).

[126] V. Tishchenko and B. Morse, E989 Internal Document No. 16228 (2019).

[127] V. Tishchenko and B. Morse, E989 Internal Document No. 16234 (2019).

[128] T. Halewood-Leagas and J. Price, E989 Internal Document No. 12894 (2018).

[129] W. Morse and V. Tishchenko, E989 Internal Document No. 16622 (2019).

[130] W. Morse and V. Tishchenko, E989 Internal Document No. 20279 (2019).

[131] A. Chapelain *et al.*, E989 Internal Document No. 18097 (2019).

[132] J. Mott *et al.*, E989 Internal Document No. 18094 (2019).

[133] P. Girotti, E989 Internal Document No. 19709 (2019).

[134] D. Sweigart, E989 Internal Document No. 16964 (2019).

[135] A. Fienberg, E989 Internal Document No. 3170 (2015).

[136] D. Sweigart, E989 Internal Document No. 17170 (2019).

[137] M. Syphers, E989 Internal Document No. 20294 (2019).

[138] H. Binney, E989 Internal Document No. 20297 (2019).

[139] M. Syphers, E989 Internal Document No. 11544 (2018).

[140] V. Tishchenko, E989 Internal Document No. 17362 (2019).

[141] V. Tishchenko, E989 Internal Document No. 20568 (2019).

[142] D. Rubin *et al.*, E989 Internal Document No. 18085 (2019).