Available online at www.sciencedirect.com
ScienceDirect
Nuclear Instruments and Methods in Physics Research A 579 (2007) 1096–1116
NUCLEAR INSTRUMENTS & METHODS IN PHYSICS RESEARCH Section A
www.elsevier.com/locate/nima

Statistical equations and methods applied to the precision muon ($g-2$) experiment at BNL

G.W. Bennett[b], B. Bousquet[j], H.N. Brown[b], G. Bunce[b], R.M. Carey[a], P. Cushman[j], G.T. Danby[b], P.T. Debevec[h], M. Deile[l], H. Deng[l], W. Deninger[h], S.K. Dhawan[l], V.P. Druzhinin[c], L. Duong[j], E. Efstathiadis[a], F.J.M. Farley[l], G.V. Fedotovich[c], S. Giron[j], F. Gray[h], D. Grigoriev[c], M. Grosse-Perdekamp[l], A. Grossmann[g], M. Hare[a], D.W. Hertzog[h], X. Huang[a], V.W. Hughes[l,✠], M. Iwasaki[k], K. Jungmann[f], D. Kawall[l], M. Kawamura[k], B.I. Khazin[c], J. Kindem[j], F. Krienen[a], I. Kronkvist[j], A. Lam[a], R. Larsen[b], Y.Y. Lee[b], I.B. Logashenko[a,c], R. McNabb[j], W. Meng[b], J. Mi[b], J.P. Miller[a], W.M. Morse[b], D. Nikas[b], C.J.G. Onderwater[h], Yu. F. Orlov[d], C. Ozben[b], J. Paley[a], Q. Peng[a], C. Polly[h], J. Pretz[l], R. Prigl[b], G. zu Putlitz[g], T. Qian[j], S.I. Redin[c,l,*], O. Rind[a], B.L. Roberts[a], N.M. Ryskulov[c], S. Sedykh[h], Y.K. Semertzidis[b], P. Shagin[j], Yu.M. Shatunov[c], E.P. Sichtermann[l], E.P. Solodov[c], M. Sossong[h], A. Steinmetz[l], L. Sulak[a], C. Timmermans[j], A. Trofimov[a], D. Urner[h], P. von Walter[g], D. Warburton[b], D. Winn[e], A. Yamamoto[i], D. Zimmerman[j]

[a]*Department of Physics, Boston University, Boston, MA 02215, USA*
[b]*Brookhaven National Laboratory, Upton, NY 11973, USA*
[c]*Budker Institute of Nuclear Physics, Novosibirsk, Russian Federation*
[d]*Newman Laboratory, Cornell University, Ithaca, NY 14853, USA*
[e]*Fairfield University, Fairfield, CT 06430, USA*
[f]*Kernfysisch Versneller Instituut, Rijksuniversiteit Groningen, KVI, Groningen NL 9747 AA, The Netherlands*
[g]*Physikalisches Institut der Universität Heidelberg, 69120 Heidelberg, Germany*
[h]*Department of Physics, University of Illinois at Urbana-Champaign, Urbana, IL 55455, USA*
[i]*KEK, High Energy Accelerator Research Organization, Tsukuba, Ibaraki 305-0801, Japan*
[j]*Department of Physics, University of Minnesota, Minneapolis, MN 55455, USA*
[k]*Tokyo Institute of Technology, Tokyo, Japan*
[l]*Physics Department, Yale University, New Haven, CT 06511, USA*

Received 8 January 2007; received in revised form 26 April 2007; accepted 12 June 2007
Available online 26 June 2007

Abstract

In the muon ($g-2$) experiment at Brookhaven National Laboratory, the spin precession frequency ω_a is obtained from a standard χ^2 minimization fit applied to the time distribution of decay electrons. The unusually high accuracy (~0.5 ppm) of the experiment puts stringent requirements on the quality of the fit and the level of understanding of the statistical properties of the fitted parameters. We discuss the properties of the fits and their implications on the derived value for ω_a, including estimates of the effect of an imperfect fit

*Corresponding author. Budker Institute of Nuclear Physics, Novosibirsk, Russian Federation. Tel.: +7 383 330 4984; fax: +7 383 330 7163.
E-mail address: redin@inp.nsk.su (S.I. Redin).
✠Deceased.

0168-9002/$ - see front matter
doi:10.1016/j.nima.2007.06.023

function, methods of including additional external information to reduce the error, the effects of splitting the data into many smaller subsets of data, applying different weighting methods to the data using energy information, and various tests of data suitability.

PACS: 02.50.Cw; 02.50.Tt

Keywords: χ^2 minimization fit; Statistical errors; Correlations; Systematic errors; Set–subset variance; Bias of fit parameters

1. Introduction

In the muon $(g-2)$ experiment E821 at Brookhaven National Laboratory [1–6] two quantities must be measured precisely in order to obtain the muon anomalous magnetic moment, or "anomaly", $a_\mu = \frac{g-2}{2}$. One of these is the average magnetic field seen by the muons in the storage volume, measured in terms of the spin precession frequency, ω_p, of a free proton at rest. The other is the average muon spin precession frequency, ω_a. The methods used to extract ω_a from the data are the major subjects discussed in this article.

The average proton spin precession frequency is derived from the frequencies observed in an array of NMR probes mounted on a trolley car, which is moved through the muon storage region every two or three days. The resulting map of the magnetic field in the storage region is weighted according to the distribution of stored muons to obtain the average field. A separate array of NMR probes fixed in locations above and below the storage region are used to monitor changes in the field between trolley mappings.

In the approximations that the focusing electric field can be neglected and $\vec{\beta}\cdot\vec{B}=0$, a_μ is related to the measured quantities B and ω_a by $\omega_a = \frac{eB}{m}a_\mu$ (see Refs. [1–6]). This expression depends on an independent knowledge of the muon mass m and on the constants involved in the conversion of ω_p to B. a_μ can also be calculated through the expression

$$a_\mu = \frac{a_\mu}{(1+a_\mu)-a_\mu} = \frac{\omega_a}{\omega_L-\omega_a} = \frac{\omega_a/\omega_p}{\omega_L/\omega_p-\omega_a/\omega_p} = \frac{\Re}{\lambda-\Re} \quad (1)$$

where ω_L is the Larmor spin precession rate of the muon. We have used Eq. (1) to determine a_μ because $\Re = \omega_a/\omega_p$ is evaluated directly from the quantities measured, and $\lambda = \mu_\mu/\mu_p = 3.18334539(10)$, the ratio of the muon to the proton magnetic moment, is exceptionally well-measured elsewhere [7,8].

The muon $(g-2)$ experiment collected data sets in each of the years 1997–2001. It is important to note that to prevent bias, the ω_a and ω_p analyses for each data set were completely independent. Intermediate results for ω_a and ω_p were always presented with concealed offsets. The offsets were revealed only after all analyses were declared complete, and at that point, no further changes could be made to the values of ω_a and ω_p.

The muon spin precession frequency ω_a appears as a modulation in the measured number of muon decay electrons above some pre-selected energy threshold, E_{thr}, as a function of time, a function well-described by the five-parameter fitting function

$$G(t) = N_\circ e^{-t/\tau}[1 + A\cos(\omega_a t + \phi)]. \quad (2)$$

This functional form applies to the time distribution of decay electrons from one detector for one muon injection (one "spill"), and also to the sum of these spectra, by virtue of the equality

$$\sum_i N_{\circ i} e^{-t_i/\tau}[1 + A_i\cos(\omega_a t_i) + \phi_i] = N_\circ e^{-t/\tau}[1 + A\cos(\omega_a t + \phi)] \quad (3)$$

where the subscript i denotes the variables pertaining to one spill. The t_i are muon decay times measured in a given detector relative to the approximate muon injection time. The typical $E_{thr} \approx 1.8\,\text{GeV}$ produced $A_i \approx 0.35$. The muon injection polarization (helicity $+1$ and -1 for the μ^- and μ^+, respectively) together with the sign of the electron decay asymmetry (-1 and $+1$, respectively) leads to $\phi_i \approx \pi$. The phases ϕ_i are affected to a small extent by, for example, small uncertainties in the muon injection time and by detector-to-detector variations in signal delays. Note that variation in the parameters A_i, ϕ_i, N_{0i} from spill-to-spill *do not affect the values of either* τ *or* ω_a. Small variations in A_i, for example because the detector gains vary from one injection to the next, are of little consequence. Similarly, large variation in ϕ_i from one spill to another do not lead to a systematic shift in the value of ω_a. However, excessive variation in ϕ dilute the magnitude of A in the summed spectrum and thereby increase the statistical uncertainty in ω_a.

Unlike the case of ω_p, where systematic errors are dominant, the determination of ω_a was statistics-limited. Indeed, while extracting fit parameters from a histogram is a familiar technique in high-energy physics, the relative accuracy of our result ($\sigma_{a_\mu}/a_\mu \sim 5\times10^{-7}$) required that the statistical uncertainties on the fit parameters and the correlations among them be well-understood. Naturally, the murky boundary between possible systematic errors and statistical fluctuations must also be examined. In our experiment, the effects of small deviations from the ideal spectral form, generally produced by a combination of beam motion and limited detector acceptance, were carefully studied.

We include discussions of the following topics:

- statistical errors and correlations of the fit parameters;
- reducing the uncertainty on the fit parameters by incorporating additional (external) knowledge of linear combination[s] of those parameters;
- estimates of systematic shifts of fit parameters due to neglected backgrounds;
- comparisons of fit parameters obtained from fitting the full set of data with those obtained from fitting some subset of data (set–subset relations);
- derivation of the bias on the fit parameters resulting from the χ^2 technique when the errors are based on the observed number of events, or on the expected number of events, and comparison with the (unbiased) maximum likelihood technique;
- systematic biases of the fit parameters arising from a χ^2 minimization procedure (the bias due to the method) and bias reduction techniques.

We also discuss briefly

- alternative and less-common data weighting methods, for example weighting according to the asymmetry of the decay electrons in order to optimize the statistical uncertainty;
- the experiment-specific ratio method, in which we regroup data in a special way in order to simplify the functional form of the distribution and in order to reduce the dependence of the fitted value of ω_a on slowly varying multiplicative corrections to the five-parameter function;
- the "folding" method of monitoring for the presence of a background process when its period is known;
- the Kolmogorov–Smirnov histogram compatibility test which we use for data selection and for detector gain corrections.

A detailed uncertainty budget of the muon $(g-2)$ experiment is presented in Ref. [6] and references therein.

In the following text, we derive general equations first and then apply them to our specific needs. Many of those general equations and their derivations may be found in standard textbooks or elsewhere. Nonetheless we choose to present them for completeness and as a framework for other derivations.

Note that for simplicity, we omit the subscript "a" at the muon $(g-2)$ frequency ω.

2. Statistical properties of χ^2 minimization fit

In a typical analysis, one attempts to fit the data with some parameterized function, which is expected to describe the underlying distribution of the data with sufficient precision. The optimal values for these parameters are obtained by minimizing the difference between the data and the function. Some well-established ways to quantify this difference are the squared differences, likelihood function, and χ^2, the last of which was used, almost exclusively, in our experiment.

In this section we will consider the basic properties of the χ^2 minimization technique; some of the limitations, biases, and approximations of this technique will be discussed in greater detail in Section 7.

Suppose we have some (binned) distribution, e.g. a time spectrum, which can be described by a function of L parameters $f(x_1, x_2, \ldots, x_L; t) \equiv f(\boldsymbol{x}; t)$. Then, by definition

$$\chi^2(\boldsymbol{x}) \equiv \sum_n \frac{(f(\boldsymbol{x}; t_n) - \mathscr{N}_n)^2}{\sigma_n^2} \tag{4}$$

where t_n and $\mathscr{N}_n$ are the position of the center of the nth channel of the histogram and its content (number of counts), respectively, and σ_n is the statistical error associated with the nth channel. In our application, we expect the distribution of counts in a given bin to follow Poisson statistics, which becomes nearly Gaussian for large numbers. We require that the number of counts in each channel be sufficiently large that the Gaussian distribution is a good approximation. For our analysis we use $\sigma_n^2 = f(\boldsymbol{x}; t_n)$, which provides a somewhat better fit quality[1] than the more traditional choice $\sigma_n^2 = \mathscr{N}_n$. Thus

$$\chi^2(\boldsymbol{x}) = \sum_n \frac{(f(\boldsymbol{x}; t_n) - \mathscr{N}_n)^2}{f(\boldsymbol{x}; t_n)}. \tag{5}$$

2.1. Statistical fluctuations of $\mathscr{N}_n$

We define $\boldsymbol{x}_\circ \equiv (x_{\circ 1}, x_{\circ 2}, \ldots, x_{\circ L})$ to be the set of (unknown) "true" parameter values for our histogram. Correspondingly, $f(\boldsymbol{x}_\circ; t_n)$ is the "true" value for the number of counts $\mathscr{N}_n$ in the nth channel. For an ensemble of similar histograms (measurements), $\mathscr{N}_n$ follows a Poisson distribution with a mean value equal to $f(\boldsymbol{x}_\circ; t_n)$:

$$\langle \mathscr{N}_n \rangle = f(\boldsymbol{x}_\circ; t_n) \tag{6}$$

where $\langle \ldots \rangle$ means ensemble average. For a given histogram, $\mathscr{N}_n - f(\boldsymbol{x}_\circ; t_n)$ is the statistical fluctuation in the number of counts in the nth channel. In this paper we use equations for ensemble averages of several of the lower powers of $\mathscr{N}_n - f(\boldsymbol{x}_\circ; t_n)$, which follow from equations for the central moments of the Poisson distribution:

$$\langle \mathscr{N}_n - f(\boldsymbol{x}_\circ; t_n) \rangle = 0 \tag{7}$$

$$\langle (\mathscr{N}_n - f(\boldsymbol{x}_\circ; t_n))^2 \rangle = f(\boldsymbol{x}_\circ; t_n) \tag{8}$$

$$\langle (\mathscr{N}_n - f(\boldsymbol{x}_\circ; t_n))^3 \rangle = f(\boldsymbol{x}_\circ; t_n) \tag{9}$$

$$\langle (\mathscr{N}_n - f(\boldsymbol{x}_\circ; t_n))^4 \rangle = f(\boldsymbol{x}_\circ; t_n) + 3(f(\boldsymbol{x}_\circ; t_n))^2 \tag{10}$$

and also the equation

$$\langle (\mathscr{N}_n - f(\boldsymbol{x}_\circ; t_n))(\mathscr{N}_m - f(\boldsymbol{x}_\circ; t_m)) \rangle = 0 \quad \text{for } n \neq m \tag{11}$$

[1] We discuss this choice in Section 7.

which follows from the statistical independence of the numbers of counts in different channels.

2.2. Statistical fluctuations of the fit parameters

As a result of statistical fluctuations in the number of counts in individual histogram channels, minimization of χ^2 in Eq. (5) gives a vector of "optimal" fit parameters $\boldsymbol{x}$, shifted with respect to the "true" value $\boldsymbol{x}_\circ$ by $\Delta\boldsymbol{x} = \boldsymbol{x} - \boldsymbol{x}_\circ$. The components of the vector $\Delta\boldsymbol{x}$ are functions of the fluctuations in $\mathcal{N}_n - f(\boldsymbol{x}_\circ; t_n)$ and, as such, can be derived from the χ^2 minimization requirement. Assuming $f(\boldsymbol{x}_\circ; t_n) \gg 1$, we have

$$0 = \frac{1}{2}\frac{\partial\chi^2}{\partial x_i} \approx \sum_n \frac{f - \mathcal{N}_n}{f} f'_i \approx \sum_n \frac{f_\circ + \sum_j f'_j \Delta x_j - \mathcal{N}_n}{f_\circ} f'_i$$
$$= \sum_j \Delta x_j \sum_n \frac{f'_i f'_j}{f_\circ} - \sum_n \frac{f'_i}{f_\circ}(\mathcal{N}_n - f_\circ) = 0 \tag{12}$$

and hence

$$\Delta x_i = \sum_j (\mathscr{A}^{-1})_{ij} \sum_n \frac{f'_j}{f_\circ}(\mathcal{N}_n - f_\circ) \tag{13}$$

where

$$\mathscr{A}_{ij} = \sum_n \frac{f'_i f'_j}{f_\circ} \tag{14}$$

are the elements of the symmetric matrix $\mathscr{A}$, and the matrix $\mathscr{A}^{-1}$ is the inverse of $\mathscr{A}$. Here for simplicity we introduce the following notations: $f \equiv f(\boldsymbol{x}; t)$, $f_\circ \equiv f(\boldsymbol{x}_\circ; t)$ and $f'_i \equiv (\partial f/\partial x_i)_{\boldsymbol{x}=\boldsymbol{x}_\circ}$ and using the Taylor expansion $f(\boldsymbol{x}, t) = f(\boldsymbol{x}_\circ + \Delta\boldsymbol{x}, t) \approx f_\circ + \sum_j f'_j \Delta x_j$. Also, in Eqs. (12)–(14) and thereafter, expressions like $\sum_n \frac{f'_i f'_j}{f_\circ}$ (where $\sum_n$ means summation over histogram channels) should be understood as $\sum_n f'_i(\boldsymbol{x}, t_n) f'_j(\boldsymbol{x}, t_n)/f'_i(\boldsymbol{x}, t_n)$ and may be written as $\sum_n (\frac{f'_i f'_j}{f_\circ})_n$. Eq. (13), one of the basic equations in statistics, relates the statistical fluctuations of the fit parameters to the statistical fluctuations of the number of counts in individual histogram channels.

If the bin width b of the fitted histogram is significantly shorter than the typical structure of the underlying distribution, the summation in Eq. (14) can be replaced by an integral: $\sum_n \approx \int \mathrm{d}t/b$. Furthermore, b can be eliminated in favor of the total number of events in the histogram, N, through the normalization relation

$$N = \sum_n \mathcal{N}_n \approx \sum_n f_\circ(t_n) \approx \int f_\circ(t)\frac{\mathrm{d}t}{b} \tag{15}$$

yielding

$$\mathscr{A}_{ij} = \frac{N}{\int f_\circ \mathrm{d}t} \int \frac{f'_i f'_j}{f_\circ}\,\mathrm{d}t. \tag{16}$$

In many cases, replacing sums by integrals allows one to obtain *analytical* expressions for the important statistical quantities and simplifies their analysis.

2.3. Statistical errors and correlations for χ^2 minimization fit

Eq. (13) can be used to obtain the statistical uncertainties of the fit parameters as well as the correlations between any pair. The error matrix (also known as the covariance matrix) is defined as the ensemble average $\langle \Delta x_i\, \Delta x_j \rangle$:

$$\langle \Delta x_i\, \Delta x_j \rangle = \sum_{ab} (\mathscr{A}^{-1})_{ia} (\mathscr{A}^{-1})_{jb} \sum_{nm} \left(\frac{f'_a}{f_\circ}\right)_n \left(\frac{f'_b}{f_\circ}\right)_m$$
$$\times \langle (\mathcal{N}_n - f_\circ)(\mathcal{N}_m - f_\circ) \rangle$$
$$= \sum_{ab} (\mathscr{A}^{-1})_{ia} (\mathscr{A}^{-1})_{jb} \mathscr{A}_{ab} = (\mathscr{A}^{-1})_{ij}. \tag{17}$$

The error matrix is nothing but the inverse of the matrix $\mathscr{A}$. The specific case $i = j$ of Eq. (17) gives the equation for the statistical errors of the fit parameters:

$$\sigma_i^2 \equiv \langle (\Delta x_i)^2 \rangle = (\mathscr{A}^{-1})_{ii}. \tag{18}$$

We note here that for the case of a one parameter fit, $\sigma_x^2 = \mathscr{A}^{-1}$, and Eq. (13) can be written as

$$\Delta x = \sigma_x^2 \sum_n \frac{f'}{f_\circ}(\mathcal{N}_n - f_\circ). \tag{19}$$

The value of χ^2 obtained from a histogram fit is also a function of statistical fluctuations:

$$\chi^2 = \sum_n \frac{(f - \mathcal{N}_n)^2}{f} \approx \sum_n \frac{(f_\circ - \mathcal{N}_n + \sum_i f'_i \Delta x_i)^2}{f_\circ}$$
$$= \sum_n \frac{(\mathcal{N}_n - f_\circ)^2}{f_\circ} - 2\sum_i \Delta x_i \sum_n \frac{f'_i}{f_\circ}(\mathcal{N}_n - f_\circ)$$
$$+ \sum_{ij} \Delta x_i\, \Delta x_j \sum_n \frac{f'_i f'_j}{f_\circ}$$
$$= \sum_n \frac{(\mathcal{N}_n - f_\circ)^2}{f_\circ} - \sum_{ij} \mathscr{A}_{ij} \Delta x_i\, \Delta x_j. \tag{20}$$

For an ensemble average of χ^2 we have

$$\langle \chi^2 \rangle = \sum_n \frac{\langle (\mathcal{N}_n - f_\circ)^2 \rangle}{f_\circ} - \sum_{ij} \mathscr{A}_{ij} \langle \Delta x_i\, \Delta x_j \rangle$$
$$= \sum_n \frac{f_\circ}{f_\circ} - \sum_{ij} \mathscr{A}_{ij} (\mathscr{A}^{-1})_{ji}$$
$$= \sum_{n=1}^{N_{\mathrm{ch}}} 1 - \sum_{i=1}^{L} 1 = N_{\mathrm{ch}} - L. \tag{21}$$

Thus, we obtain a familiar result: $\langle \chi^2 \rangle$ is equal to the number of channels of the histogram, N_{ch}, minus the number of fit parameters L, i.e. equal to the number of degrees of freedom of the fit.

The mean square deviation of χ^2 from its mean value is, by definition, the variance of χ^2:

$$\sigma_{\chi^2}^2 \equiv \langle (\chi^2 - \langle \chi^2 \rangle)^2 \rangle = \langle (\chi^2)^2 \rangle - 2\langle \chi^2 \rangle \langle \chi^2 \rangle + \langle \chi^2 \rangle^2$$
$$= \langle (\chi^2)^2 \rangle - \langle \chi^2 \rangle^2. \tag{22}$$

While $\langle\chi^2\rangle^2$ is trivially obtained as Eq. (21) squared, evaluation of $\langle(\chi^2)^2\rangle$ is more sophisticated and tedious. It is given in the Appendix. The result for $\sigma^2_{\chi^2}$ is

$$\begin{aligned}\sigma^2_{\chi^2} &= \langle(\chi^2)^2\rangle - \langle\chi^2\rangle^2 \\ &= (N^2_{\rm ch} - 2N_{\rm ch}L + L^2 + 2N_{\rm ch} - 2L) - (N_{\rm ch} - L)^2 \\ &= 2N_{\rm ch} - 2L. \end{aligned} \tag{23}$$

Eqs. (21) and (23) are used to check the quality (goodness) of the fit.

3. Statistical errors and correlations for the $(g-2)$ five-parameter fit

As mentioned above, the function $G(t) = N_\circ e^{-t/\tau}[1 + A\cos(\omega t + \phi)]$ with the fit parameters $(x_1, x_2, x_3, x_4, x_5) = (N_\circ, \tau, A, \omega, \phi)$ is of central importance to the muon $(g-2)$ experiment as it describes well the time distribution of the decay electrons. Evaluation of the elements of matrix $\mathscr{A}$, as well as other important quantities to be discussed later, involves integrations of some combinations of the exponential ($e^{-t/\tau}$), harmonic ($\sin(\omega t + \phi)$ or $\cos(\omega t + \phi)$) and polynomial (t, t^2) functions. In some cases such integrations can be performed analytically. In others, successive approximations in the small dimensionless parameter $\varepsilon = (\tau\omega)^{-1} = 0.011$ may be employed. Some integrations involve $1 + A\cos(\omega t + \phi)$ in the denominator. In order to obtain analytical expressions in such cases we use expansions in $\frac{1}{2}A^2 = 0.08$.

To the leading order in ε and $\frac{1}{2}A^2$, the matrix $\mathscr{A}$ for the fit function $G(t)$ is

$$\mathscr{A} = \begin{bmatrix} \frac{N}{N_\circ^2} & \frac{N}{N_\circ\tau}\left(\frac{t_s}{\tau}+1\right) & 0 & 0 & 0 \\ \frac{N}{N_\circ\tau}\left(\frac{t_s}{\tau}+1\right) & \frac{N}{\tau^2}\left[\left(\frac{t_s}{\tau}+1\right)^2+1\right] & 0 & 0 & 0 \\ 0 & 0 & \frac{N}{2} & 0 & 0 \\ 0 & 0 & 0 & \frac{NA^2\tau^2}{2}\left[\left(\frac{t_s}{\tau}+1\right)^2+1\right] & \frac{NA^2\tau}{2}\left(\frac{t_s}{\tau}+1\right) \\ 0 & 0 & 0 & \frac{NA^2\tau}{2}\left(\frac{t_s}{\tau}+1\right) & \frac{NA^2}{2} \end{bmatrix} \tag{24}$$

where t_s is the start time of the fit. The end time of the fit, $t_{\max}$, is not that important provided it is at least several muon lifetimes, because of the exponentially decreasing number of counts. It is taken to be $+\infty$ for simplicity.

Using Eqs. (18) and (24) one readily finds the statistical errors for the five parameters:

$$\sigma_{N_\circ} = \frac{N_\circ}{\sqrt{N}}\sqrt{(t_s/\tau+1)^2+1} \tag{25}$$

$$\sigma_\tau = \frac{\tau}{\sqrt{N}} \tag{26}$$

$$\sigma_A = \frac{\sqrt{2}}{\sqrt{N}} \tag{27}$$

$$\sigma_\omega = \frac{\sqrt{2}}{\tau A\sqrt{N}} \tag{28}$$

$$\sigma_\phi = \frac{\sqrt{2}}{A\sqrt{N}}\sqrt{(t_s/\tau+1)^2+1}. \tag{29}$$

$\mathscr{A}$ in Eq. (24), as well as its inverse $\mathscr{A}^{-1}$, are diagonal-block (or step) matrices, which allows one to split the problem into three independent ones with lower rank matrices, thus simplifying calculations substantially.

Design parameters were chosen to minimize systematic errors and the statistical errors described in Eqs. (25)–(29). It follows from Eq. (28) that the relative error σ_ω/ω scales as $(\tau\omega)^{-1}$ and hence improves (decreases) for increasing muon energy E_μ (larger $\tau = \tau_\circ\gamma = \tau_\circ E_\mu/mc^2$) and larger magnetic field B (larger $\omega = (g-2)eB/2m$). Indeed, the increasing statistical accuracy of the three muon $(g-2)$ experiments performed at CERN could be traced, in part, to a steady increase of the muon momentum.

In the most recent CERN experiment and in the current BNL experiment, the muon momentum is fixed at $P_\mu = 3.094\,\text{GeV}/c$, implying a γ factor of 29.30 and a time dilated muon lifetime of 64.38 μs. At this *magic momentum*, the focusing electric field has no effect on the muon spin precession. ω is determined by the magnetic field alone, which is made as uniform as possible. The reduction in systematic errors allowed by use of the magic momentum is so great that further increase in muon energy is pointless.

The weak focusing storage ring ($n \approx 0.135$) and its 9 cm aperture diameter produce a range in stored muon momenta of about ±0.2%. For a storage ring of radius 7.112 m and magnetic field $B = 1.45\,\text{T}$, the cyclotron period is about 149.2 ns, and the spin precession period, $T = 2\pi/\omega$, is roughly 4.365 μs. The choice of the magnetic field strength is a trade-off between the competing demands of statistical and systematic accuracy: while a higher field increases ω, the magnetic field should not be so large as to drive the iron pole tips deep into saturation.

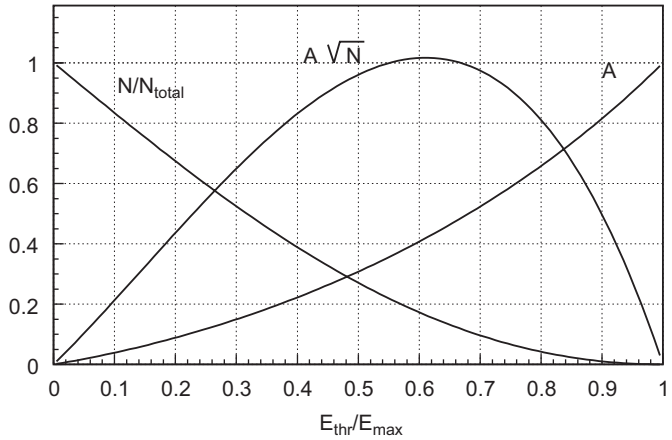

Fig. 1. Relative number of events N/N_{total}, asymmetry A and the combination $A\sqrt{N}$ (arbitrary scale) as functions of relative energy threshold.

The decay electrons, which are collected from a few tens of microseconds to about 600 μs after muon injection into the ring, range in energy from 0 to 3.097 GeV. Because the relative uncertainty in ω (see Eq. (28)) scales as $(A\sqrt{N})^{-1}$, the number of registered decay electrons and their average weak-asymmetry depends on the energy threshold E_{thr} used in the experiment. Thus both $(A\sqrt{N})^{-1}$ and σ_ω/ω also depend on E_{thr}. Fig. 1 shows N and A and the combination $A\sqrt{N}$ as functions of $E_{\text{thr}}/E_{\max}$, where $E_{\max} \approx 3.1\,\text{GeV}$. The maximum of $A\sqrt{N}$, and hence the minimum of σ_ω/ω, is reached at about $E_{\text{thr}}/E_{\max} = 0.6$ or $E_{\text{thr}} = 1.8\,\text{GeV}$.

And, of course, $\sigma_\omega \propto 1/\sqrt{N}$ implies a better statistical accuracy for more muon decays. From Eq. (28) we see that for the typical case of $A = 0.4$, and with $(\tau\omega)^{-1} = 0.011$, we need $N = 6 \times 10^9$ muons to achieve a relative precision of 0.5 ppm on ω.

3.1. Correlations

The non-vanishing off-diagonal elements $\mathscr{A}_{12} = \mathscr{A}_{21} = \frac{N}{N_\circ \tau}(\frac{t_s}{\tau} + 1)$ and $\mathscr{A}_{45} = \mathscr{A}_{54} = \frac{NA^2\tau}{2}(\frac{t_s}{\tau} + 1)$ of the matrix $\mathscr{A}$ in Eq. (24) indicate that there are only two pairs of non-vanishing correlations among the parameters of the five-parameter fit: one correlation is between $N_\circ$ and τ and the other is between ω and ϕ. These matrix elements (and hence correlations) vanish when $t_s = -\tau$, i.e. when $t = 0$ is chosen to be one muon lifetime after the time of the beginning of the fit. Matrix $\mathscr{A}$ becomes diagonal:

$$\mathscr{A} = \begin{pmatrix} \frac{N}{N_\circ^2} & 0 & 0 & 0 & 0 \\ 0 & \frac{N}{\tau^2} & 0 & 0 & 0 \\ 0 & 0 & \frac{N}{2} & 0 & 0 \\ 0 & 0 & 0 & \frac{NA^2\tau^2}{2} & 0 \\ 0 & 0 & 0 & 0 & \frac{NA^2}{2} \end{pmatrix}. \tag{30}$$

It is important to realize that the choice of $t_s = -\tau$ does not improve the statistical errors on these parameters, a point discussed in more detail below, but by choosing $t_s = -\tau$, the *center of gravity* time frame, we do simplify some of the calculations. In some instances, this choice of t_s improves the stability of χ^2 fitting algorithms, whose performance can be hampered by correlations between parameters.

3.2. Improvement of σ_ω through knowledge of the phase

Since the frequency and phase are highly correlated, in general, an independent determination of the phase could be used to reduce the statistical error σ_ω. As we will see, the size of the improvement depends on the time at which the phase is known. For example, if the phase is known only at a time τ after the start of the fit, then there will be no improvement in the error of ω. While independent knowledge of the phase was insufficient to reduce σ_ω in our experiment, we include this discussion for others.

In the muon $(g-2)$ experiment, polarized muons from the forward decay of in-flight pions are injected into the muon storage ring. At the injection point, the spins of negative muons are roughly parallel to their momenta. (For positive muons, the spins and momenta are nearly anti-parallel.) The highest energy electrons from the three-body decay of the muon are the ones which carry the most information about the spin direction, and therefore are the ones of most importance for the $(g-2)$ experiment. The preferred directions of high-energy decay electrons are anti-parallel and parallel to the μ^- and μ^+ spin directions, respectively. Therefore the phases of the oscillation in the number of electrons versus time at the time of injection are the same for both μ^- and μ^+.

Consider a simplified model of injection, taken to be at $t = 0$, where the muon polarization, and hence the phase ϕ in Eq. (2), is known precisely. In that case, ϕ is a constant and can be excluded from the set of fit parameters. To calculate σ_ω for such a case, one has to eliminate the fifth row and fifth column of $\mathscr{A}$ in Eq. (24), which then results in

$$\sigma_\omega = \frac{\sqrt{2}}{\tau A\sqrt{N}} \frac{1}{\sqrt{(t_s/\tau + 1)^2 + 1}}. \tag{31}$$

This is less than σ_ω in Eq. (28) by a factor of $\sqrt{(t_s/\tau + 1)^2 + 1}$. Since data taking always starts later than injection, $t_s \geqslant 0$ and $\sqrt{(t_s/\tau + 1)^2 + 1} \geqslant \sqrt{2}$. Thus the accuracy of ω can be improved by better than a factor $\sqrt{2}$ if the phase (i.e. average polarization of muons) is known precisely at the moment of injection. The amount of improvement depends on the time at which ϕ is known. In the next section we discuss the more realistic case when there is independent knowledge of the $(g-2)$ phase at an *arbitrary* time but with *limited* precision σ_F.

4. Improvement of statistical errors from independent knowledge of a linear combination of fit parameters

Suppose we can measure independently some linear combination of fit parameters, $F = \sum_i C_i x_i$, with a limited precision σ_F. The result of such a measurement is denoted as F_*. In order to improve the statistical errors of the fit parameters x_i through additional knowledge of the combination F, we may add an extra term $(F - F_*)^2/\sigma_F^2$ to the equation for χ^2:

$$\chi^2 = \sum_n \frac{(f - \mathcal{N}_n)^2}{f_\circ} + \frac{(F - F_*)^2}{\sigma_F^2}. \tag{32}$$

We denote $F_\circ \equiv \sum_i C_i x_{\circ i}$ to be the "true" value for F, and introduce $\Delta F_* \equiv F_* - F_\circ$. By definition, the ensemble average of $(\Delta F_*)^2$ is equal to the variance of F: $\langle(\Delta F_*)^2\rangle \equiv \sigma_F^2$. It is convenient to rewrite $(F - F_*)$ in Eq. (32) as

$$\begin{aligned} F - F_* &= (F - F_\circ) - (F_* - F_\circ) = \sum_i C_i(x_i - x_{\circ i}) - \Delta F_* \\ &= \sum_i C_i \Delta x_i - \Delta F_*. \end{aligned} \tag{33}$$

Then Eq. (32) reads

$$\chi^2 = \sum_n \frac{(f - \mathcal{N}_n)^2}{f_\circ} + \frac{1}{\sigma_F^2}\left(\sum_j C_j \Delta x_j - \Delta F_*\right)^2 \tag{34}$$

and for the χ^2 minimization condition we have

$$\begin{aligned} 0 = \frac{1}{2}\frac{\partial \chi^2}{\partial x_i} &= \sum_j \Delta x_j \sum_n \frac{f_i' f_j'}{f_\circ} - \sum_n \frac{f_i'}{f_\circ}(\mathcal{N}_n - f_\circ) \\ &\quad + \frac{1}{\sigma_F^2} C_i \sum_j C_j \Delta x_j - \frac{1}{\sigma_F^2} C_i \Delta F_* \\ &= \sum_j \Delta x_j (\mathscr{A}_c)_{ij} - \sum_n \frac{f_i'}{f_\circ}(\mathcal{N}_n - f_\circ) - \frac{C_i}{\sigma_F^2}\Delta F_* = 0 \end{aligned} \tag{35}$$

where we introduce the symmetric matrix $\mathscr{A}_c$ with matrix elements $(\mathscr{A}_c)_{ij}$:

$$(\mathscr{A}_c)_{ij} = \mathscr{A}_{ij} + \frac{C_i C_j}{\sigma_F^2}. \tag{36}$$

The solution for Eq. (35) is

$$\begin{aligned} \Delta x_i = &\sum_j (\mathscr{A}_c^{-1})_{ij} \sum_n \frac{f_j'}{f_\circ}(\mathcal{N}_n - f_\circ) \\ &+ \sum_j (\mathscr{A}_c^{-1})_{ij} \frac{C_j}{\sigma_F^2}\Delta F_*. \end{aligned} \tag{37}$$

Since statistical fluctuations $\mathcal{N}_n - f_\circ$ and ΔF_* are not correlated, the error matrix $\langle \Delta x_i \Delta x_j\rangle$ for this case is

$$\begin{aligned} \langle \Delta x_i \Delta x_j\rangle = &\sum_{ab} (\mathscr{A}_c^{-1})_{ia}(\mathscr{A}_c^{-1})_{jb} \\ &\times \left\langle \sum_n \frac{f_a'}{f_\circ}(\mathcal{N}_n - f_\circ) \sum_m \frac{f_b'}{f_\circ}(\mathcal{N}_m - f_\circ) \right\rangle \\ &+ \sum_{ab} (\mathscr{A}_c^{-1})_{ia}(\mathscr{A}_c^{-1})_{jb} \frac{C_a C_b}{\sigma_F^4} \langle(\Delta F_*)^2\rangle \\ = &\sum_{ab} (\mathscr{A}_c^{-1})_{ia}(\mathscr{A}_c^{-1})_{jb} \mathscr{A}_{ab} \\ &+ \sum_{ab} (\mathscr{A}_c^{-1})_{ia}(\mathscr{A}_c^{-1})_{jb} \frac{C_a C_b}{\sigma_F^2} \\ = &\sum_{ab} (\mathscr{A}_c^{-1})_{ia}(\mathscr{A}_c^{-1})_{jb} (\mathscr{A}_c)_{ab} \\ = &(\mathscr{A}_c^{-1})_{ij} \end{aligned} \tag{38}$$

which is same as in Eq. (17) with matrix $\mathscr{A}$ replaced by matrix $\mathscr{A}_c$. For the case where several linear combinations of fit parameters $F_k = \sum_i C_{ik} x_i$ $(k = 1, \ldots, K)$ are known with limited precisions σ_{F_k}, matrix $\mathscr{A}_c$ becomes

$$(\mathscr{A}_c)_{ij} = \mathscr{A}_{ij} + \sum_{k=1}^{K} \frac{C_{ik} C_{jk}}{\sigma_{F_k}^2}. \tag{39}$$

Eq. (38) can be used to study how we might reduce σ_ω. We assume that the $(g-2)$ phase is known to be ϕ_* at some time t_* with finite precision σ_F, independent of the fit. We restrict our study to two parameters only, ω and ϕ (as x_1 and x_2), since the other parameters are statistically independent of these two. In this case, the linear combination F is simply ϕ_*: $F = \phi_*$. In terms of the fit parameters ω and ϕ it is

$$F = \phi_* \equiv (\omega t + \phi)_{t=t_*} = \omega t_* + \phi \tag{40}$$

thus $C_1 = t_*$ and $C_2 = 1$. The matrix $(\mathscr{A}_c)_{ij}$ can be written as

$$\begin{aligned} (\mathscr{A}_c)_{ij} &= \mathscr{A}_{ij} + \sigma_F^{-2} C_i C_j \\ &= \begin{pmatrix} \sigma_{\omega\circ}^{-2}\left[\left(\frac{t_s}{\tau}+1\right)^2 + 1\right] + \sigma_F^{-2} t_*^2 & \sigma_{\omega\circ}^{-2}\frac{1}{\tau}\left(\frac{t_s}{\tau}+1\right) + \sigma_F^{-2} t_* \\ \sigma_{\omega\circ}^{-2}\frac{1}{\tau}\left(\frac{t_s}{\tau}+1\right) + \sigma_F^{-2} t_* & \sigma_{\omega\circ}^{-2}\frac{1}{\tau^2} + \sigma_F^{-2} \end{pmatrix} \end{aligned} \tag{41}$$

which gives

$$\sigma_\omega^{-2} = (\mathscr{A}_c^{-1})_{22} = \sigma_{\omega\circ}^{-2}\left(1 + \frac{\sigma_F^{-2}}{\sigma_{\phi\circ}^{-2} + \sigma_F^{-2}} \frac{(t_s + \tau - t_*)^2}{\tau^2}\right). \tag{42}$$

Here we introduce $\sigma_{\omega\circ} = \frac{\sqrt{2}}{\tau A\sqrt{N}}$, which is the statistical error of ω from the five-parameter fit alone (Eq. (28)), and also $\sigma_{\phi\circ} \equiv \frac{\sqrt{2}}{A\sqrt{N}}$, which is the statistical error of the phase in the

center of gravity time frame from the five-parameter fit alone (Eq. (29) for $t_s = -\tau$).

It follows from Eq. (42) that the figure of merit for σ_F, the precision required to improve σ_ω by an independent measurement of the phase, is $\sigma_{\phi\circ}$. Since the goal of the $(g-2)$ experiment is to measure the muon spin $(g-2)$ precession frequency to a relative precision of $\sim 10^{-6}$,

$$\sigma_{\phi\circ} = \sigma_{\omega\circ} \cdot \tau = \frac{\sigma_{\omega\circ}}{\omega} \cdot \omega\tau \sim 10^{-6} \times 100 = 10^{-4} \tag{43}$$

i.e. 0.1 mrad. Note that it takes only $(10^{-4}/2\pi) \times 4.365\,\mu s = 69$ ps for the muon spin to precess by 0.1 mrad, which corresponds to 2.1 cm in the muon (or decay electron) motion. That same level of precision is required for a whole series of measurements: the angle of the polarization of the muons injected into the storage ring, the beam injection time and the position of the beam center at a given time, the average trajectory length or travel time of the decay electrons from the point of decay to the detector, the acceptance of the detectors as a function of spin angle, etc. Because each of these measurements is very challenging, if not impossible, the project was not pursued. Still, Eq. (42) may be useful wherever the five-parameter function $G(t)$ from Eq. (2) can be applied to fit an experimental distribution.

5. Systematic shift of fit parameters due to neglected backgrounds

Suppose we have some low-level background $h(t)$ admixed into the data, which are otherwise described by a multi-parameter function $f(\boldsymbol{x}; t)$. The background might be small enough to evade observation "by eye" or even to spoil χ^2 considerably. Fitting the histogram with the function $f(\boldsymbol{x}; t)$ alone will give parameter values x_i, shifted with respect to the "true" values $x_{\circ i}$ by some $\delta x_i = x_i - x_{\circ i}$. These systematic shifts δx_i are functions of $h(t)$ and as such can be found from the minimization requirement:

$$\begin{aligned}\frac{1}{2}\frac{\partial\chi^2}{\partial x_i} &\approx \sum_n \frac{f - \mathcal{N}_n}{f} f_i' \\ &\approx \sum_j \delta x_j \sum_n \frac{f_j' f_i'}{f} - \sum_n \frac{h}{f} f_i' \\ &= \sum_j \mathcal{A}_{ij} \delta x_i - \sum_n \frac{h}{f} f_i' = 0\end{aligned} \tag{44}$$

and hence

$$\delta x_i = \sum_j (\mathcal{A}^{-1})_{ij} \times \sum_n \frac{h}{f} f_j' \tag{45}$$

with the same matrix $\mathcal{A}$ as in Section 2. Here we use $f(\boldsymbol{x}, t_n) \approx f(\boldsymbol{x}_\circ, t_n) + \sum_j f_j' \delta x_j$ and also $\mathcal{N}_n \approx f(\boldsymbol{x}_\circ, t_n) + h(t_n)$, which follows from $\langle \mathcal{N}_n \rangle = f(\boldsymbol{x}_\circ, t_n) + h(t_n)$. As before, we may replace the summation with an integration, resulting in

$$\begin{aligned}\delta x_i &= \sum_j (\mathcal{A}^{-1})_{ij} \times \frac{1}{b} \int \frac{h(t)}{f} f_j' \, \mathrm{d}t \\ &= \sum_j (\mathcal{A}^{-1})_{ij} \times \frac{N}{\int f \, \mathrm{d}t} \int \frac{h(t)}{f} f_j' \, \mathrm{d}t.\end{aligned} \tag{46}$$

For the special case of a one parameter fit

$$\delta x = \sigma_x^2 \sum_n \frac{h}{f} f'$$

or

$$\delta x = \sigma_x^2 \frac{N}{\int f \, \mathrm{d}t} \int \frac{h(t)}{f} f' \, \mathrm{d}t. \tag{47}$$

In the center of gravity time frame, Eqs. (47) and (30) give the following systematic shifts for the five $(g-2)$ parameters:

$$\delta N_\circ = \frac{1}{e\tau} \int_{-\tau}^{\infty} h(t) \, \mathrm{d}t \tag{48}$$

$$\delta\tau = \frac{1}{eN_\circ \tau} \int_{-\tau}^{\infty} t h(t) \, \mathrm{d}t \tag{49}$$

$$\delta A = \frac{2}{eN_\circ \tau} \int_{-\tau}^{\infty} \frac{h(t) \cos(\omega t + \phi)}{1 + A\cos(\omega t + \phi)} \, \mathrm{d}t \tag{50}$$

$$\delta\omega = -\frac{2}{eN_\circ A\tau^3} \int_{-\tau}^{\infty} \frac{t h(t) \sin(\omega t + \phi)}{1 + A\cos(\omega t + \phi)} \, \mathrm{d}t \tag{51}$$

$$\delta\phi = -\frac{2}{eN_\circ A\tau} \int_{-\tau}^{\infty} \frac{h(t) \sin(\omega t + \phi)}{1 + A\cos(\omega t + \phi)} \, \mathrm{d}t. \tag{52}$$

5.1. Effect of coherent betatron oscillations

As an example of Eq. (51), we consider the systematic shift of the $(g-2)$ frequency caused by coherent radial oscillations of the muon beam. At the beginning of each data taking cycle, muons are injected through a narrow aperture (the "inflector") into the ring. The focus, formed at the injection point, is re-formed at regular intervals in the horizontal and vertical direction. In addition, the magnetic kick required to push muons onto stored orbits was smaller than ideal, resulting in radial oscillations of the beam centroid and an average radius which is about 2.5 mm larger than that corresponding to the magic momentum. The oscillations associated with the motion of the centroid and the vertical and horizontal envelopes are collectively referred to as coherent betatron oscillations (CBO).

CBO modify the decay electron time spectrum through their time-dependent effect on the acceptance of the detector system, which is determined by the position and momentum of the decaying muon. ω is particularly sensitive to the radial motion of the muon beam whose frequency, as seen from a given detector station, is by the

chance very nearly equal to 2ω. If not properly accounted for, this small acceptance oscillation can introduce a systematic shift into the value of ω obtained from a χ^2 fit.

As an example, consider the case of neglecting just one effect of the radial beam motion, the modulation of the average time delay between the emergence of the electron from muon decay and its registration in the detector station. This modulation is equivalent to replacement of time t in the five-parameter function by the CBO modulated time: $t \to t - g(t)$, with CBO function $g(t)$:

$$g(t) = A_{\rm cbo}\cos(\omega_{\rm cbo}t + \phi_{\rm cbo}) \tag{53}$$

where $\omega_{\rm cbo}$ and $\phi_{\rm cbo}$ are the CBO frequency and phase, respectively. Thus we obtain the modified five-parameter function

$$f(t) = N_\circ e^{-(t-g(t))/\tau}(1 + A\cos[\omega(t - g(t)) + \phi]). \tag{54}$$

For a small amplitude of modulation $A_{\rm cbo}$, $f(t)$ can be written as

$$f(t) \approx N_\circ e^{-t/\tau}[1 + A\cos(\omega t + \phi)] + h(t) \tag{55}$$

where $h(t)$ is a small additive background,

$$h(t) = N_\circ e^{-t/\tau} g(t)\omega[\varepsilon + \varepsilon A\cos(\omega t + \phi) + A\sin(\omega t + \phi)] \tag{56}$$

with $\varepsilon = (\omega\tau)^{-1} = 0.011$. Substituting $h(t)$ from Eq. (56) into Eq. (51), to leading order ε we have

$$\frac{\delta\omega}{\omega} \approx -\frac{2A_{\rm cbo}}{e\tau^3}\int_{-\tau}^{\infty} t e^{-t/\tau}\cos(\omega_{\rm cbo}t + \phi_{\rm cbo}) \times \frac{\sin^2(\omega t + \phi)}{1 + A\cos(\omega t + \phi)}\,{\rm d}t. \tag{57}$$

The fraction in the integral can be decomposed into a Fourier series[2]:

$$\frac{\sin^2(\omega t + \phi)}{1 + A\cos(\omega t + \phi)} = \frac{a}{A}\left(1 - a\cos(\omega t + \phi) + \sum_{n=2}^{\infty}[(-a)^n - (-a)^{n-2}]\cos n(\omega t + \phi)\right) \tag{58}$$

where $a \equiv \frac{1-\sqrt{1-A^2}}{A}$, and for the typical value $A \approx 0.4$, $a \approx 0.2$. Since $\omega_{\rm cbo} \approx 2\omega$, we would expect the term $\frac{a}{A}(a^2 - 1)\cos(2\omega t + 2\phi) \approx -\frac{1}{2}\cos(2\omega t + 2\phi)$ to give the leading, "resonance", contribution. Thus, dropping "non-resonance" terms, we arrive at

$$\frac{\delta\omega}{\omega} \approx \frac{A_{\rm cbo}}{2e\tau^3}\int_{-\tau}^{\infty} t e^{-t/\tau}\cos[(\omega_{\rm cbo} - 2\omega)t + \phi_{\rm cbo} - 2\phi]\,{\rm d}t = -\frac{A_{\rm cbo}}{2\tau}\frac{\xi}{\xi^2+1}\left[\frac{2\xi}{\xi^2+1}\cos\psi - \frac{\xi^2-1}{\xi^2+1}\sin\psi\right] \tag{59}$$

where we introduce $\xi \equiv (\omega_{\rm cbo} - 2\omega)\tau$ and $\psi \equiv -(\omega_{\rm cbo} - 2\omega)\tau + \phi_{\rm cbo} - 2\phi$ for simplicity. Furthermore, since $(\frac{2\xi}{\xi^2+1})^2 + (\frac{\xi^2-1}{\xi^2+1})^2 = 1$, we may introduce an angle β such that $\sin\beta = \frac{2\xi}{\xi^2+1}$ and $\cos\beta = \frac{\xi^2-1}{\xi^2+1}$. Then Eq. (59) can be written as

$$\frac{\delta\omega}{\omega} = -\frac{A_{\rm cbo}}{2\tau}\frac{\xi}{\xi^2+1}\sin(\beta - \psi)$$

$$\text{and hence } \left|\frac{\delta\omega}{\omega}\right| \leqslant \frac{A_{\rm cbo}}{2\tau}\frac{|\xi|}{\xi^2+1}. \tag{60}$$

[2]For such decomposition we use Eq. (417.4) from Ref. [9].

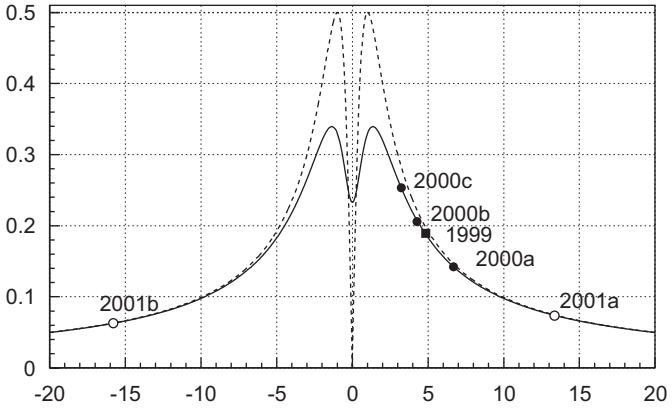

Fig. 2. Relative amplitude of the CBO effect. Horizontal axis: the deviation of the horizontal CBO frequency from two times the spin precession frequency, $\xi = (\omega_{\rm cbo} - 2\omega)\tau$. Vertical axis: the induced shift in ω, in arbitrary units. Dashed line: infinite CBO lifetime. Solid line: typical observed lifetime of 100 μs for the CBO.

The function $|\xi|/(\xi^2+1)$ is shown in Fig. 2 by a dashed line. It represents, in arbitrary units, the relative effect on ω for the case of an infinite CBO lifetime, $\tau_{\rm cbo}$. The effect in the real experiment, where $\tau_{\rm cbo} \sim 100\,\mu$s, is given by the solid line.

6. Set–subset relations for the χ^2 fit

6.1. Set–subset variance for the χ^2 fit parameters

In the following the correlation of errors between the value $\boldsymbol{x}_2$ of parameters fitted on a subset of data and the corresponding value $\boldsymbol{x}_1$ obtained when fitting a subset of the same data is studied.

The variance of the difference $(\boldsymbol{x}_1 - \boldsymbol{x}_2)$ between the subset result and the full set result will be called "set–subset variance" for parameters $\boldsymbol{x}$.

In order to search for perturbations to the five-parameter function, perturbations which are often largest at early times, we performed a series of χ^2 fits with successively later start times, holding fixed the stop time. Since perturbations would manifest themselves as variations in the fitted values of ω or other parameters as a function of the start time, it was important to establish the expected statistical fluctuations in parameters derived from the series of fits. The fits with later start times necessarily involve a subset of the data associated with fits with earlier start times; given the data overlap, there will be a significant correlation in the values of the parameters derived from fits to these spectra.

In this section we denote $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ to be the vectors of fit parameters obtained from χ^2 minimization for the full histogram Ω_1 and for some subset Ω_2 of channels in the histogram, respectively. Denote $\boldsymbol{x}_\circ$ to be the vector of "true" values of fit parameters, common for both Ω_1 and Ω_2. Then

$$x_{1i} - x_{\circ i} \equiv \Delta x_{1i} = \sum_j (\mathscr{A}_1^{-1})_{ij} \sum_{n\in\Omega_1} \frac{f'_j}{f_\circ}(\mathscr{N}_n - f_\circ)$$

$$\text{where } \mathscr{A}_{1ij} = \sum_{n\in\Omega_1} \frac{f'_i f'_j}{f_\circ} \tag{61}$$

$$x_{2i} - x_{\circ i} \equiv \Delta x_{2i} = \sum_j (\mathscr{A}_2^{-1})_{ij} \sum_{m\in\Omega_2} \frac{f'_j}{f_\circ}(\mathscr{N}_m - f_\circ)$$

$$\text{where } \mathscr{A}_{2ij} = \sum_{m\in\Omega_2} \frac{f'_i f'_j}{f_\circ}. \tag{62}$$

Calculate ensemble averages:

$$\langle x_{1i} - x_{2i}\rangle = \langle \Delta x_{1i} - \Delta x_{2i}\rangle = \langle \Delta x_{1i}\rangle - \langle \Delta x_{2i}\rangle = 0 \tag{63}$$

$$\begin{aligned}\langle (x_{1i} - x_{2i})^2\rangle &= \langle (\Delta x_{1i} - \Delta x_{2i})^2\rangle \\ &= \langle \Delta x_{1i}^2\rangle - 2\langle \Delta x_{1i}\,\Delta x_{2i}\rangle + \langle \Delta x_{2i}^2\rangle \\ &= \sigma_{1i}^2 - 2\langle \Delta x_{1i}\,\Delta x_{2i}\rangle + \sigma_{2i}^2.\end{aligned} \tag{64}$$

Using Eqs. (13) and (11), calculate the correlation term in Eq. (64):

$$\begin{aligned}\langle \Delta x_{1i}\,\Delta x_{2i}\rangle &= \sum_{jk} (\mathscr{A}_1^{-1})_{ij} (\mathscr{A}_2^{-1})_{ik} \sum_{n\in\Omega_1}\sum_{m\in\Omega_2} \left(\frac{f'_j}{f_\circ}\right)_n \\ &\quad \times \left(\frac{f'_k}{f_\circ}\right)_m \langle (\mathscr{N}_n - f_\circ)(\mathscr{N}_m - f_\circ)\rangle \\ &= \sum_{jk} (\mathscr{A}_1^{-1})_{ij} (\mathscr{A}_2^{-1})_{ik} \sum_{m\in\Omega_2} \frac{f'_j f'_k}{f_\circ^2} f_\circ \\ &= \sum_{jk} (\mathscr{A}_1^{-1})_{ij} (\mathscr{A}_2^{-1})_{ik} \mathscr{A}_{2jk} = (\mathscr{A}_1^{-1})_{ii} = \sigma_{1i}^2.\end{aligned} \tag{65}$$

Thus the set–subset variance for the fit parameters is

$$\langle (x_{1i} - x_{2i})^2\rangle = \sigma_{1i}^2 - 2\sigma_{1i}^2 + \sigma_{2i}^2 = \sigma_{2i}^2 - \sigma_{1i}^2. \tag{66}$$

N.B.: this equation is valid for any fit function, regardless of the number of parameters and possible correlations among them.

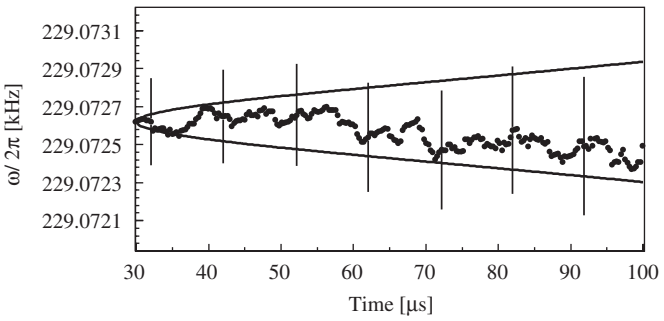

Fig. 3. $(g-2)$ frequency ω as a function of fit start time. The solid lines denote the one σ statistical deviation given by Eq. (67).

Fig. 3 (cf. Fig. 4a from Ref. [3]) shows the $(g-2)$ frequency ω as a function of fit start time. Statistical errors in ω are represented by vertical bars. The envelope is centered at the initial value of the frequency, ω_1 (obtained from a fit of the full set of histogram channels), and indicates the one standard deviation band allowed for a "random walk" of ω, as calculated using Eq. (66):

$$\langle (\omega_1 - \omega_2)^2\rangle = \sigma_{2\omega}^2 - \sigma_{1\omega}^2. \tag{67}$$

Here, $\sigma_{2\omega}$ and ω_2 refer to the subset of data at the later start time. The variations in ω are consistent with statistical fluctuations.

6.2. Set–subset variance for the χ^2 value

The value of χ^2 as a function of fit start time has also been employed for systematic studies. The set–subset variance equations derived in Section 6.1 cannot be applied directly for this case since χ^2 is not a fit parameter, and we have to start from scratch.

As before, we shall use subscripts 1 and 2 for the full set and for a subset of channels of a histogram (Ω_1 and Ω_2, respectively) and also subscript 3 for the subset complementary to Ω_2: $\Omega_3 + \Omega_2 = \Omega_1$. Using Eq. (20):

$$\begin{aligned}\chi_1^2 - \chi_2^2 &= \sum_{n\in\Omega_1} \frac{(\mathscr{N}_n - f_\circ)^2}{f_\circ} - \sum_{ij} \mathscr{A}_{1ij} \Delta x_{1i}\,\Delta x_{1j} \\ &\quad - \sum_{n\in\Omega_2} \frac{(\mathscr{N}_n - f_\circ)^2}{f_\circ} + \sum_{ij} \mathscr{A}_{2ij} \Delta x_{2i}\,\Delta x_{2j} \\ &= \sum_{n\in\Omega_3} \frac{(\mathscr{N}_n - f_\circ)^2}{f_\circ} - \sum_{ij} \mathscr{A}_{1ij} \Delta x_{1i}\,\Delta x_{1j} \\ &\quad + \sum_{ij} \mathscr{A}_{2ij} \Delta x_{2i}\,\Delta x_{2j}.\end{aligned} \tag{68}$$

The mean value of $\chi_1^2 - \chi_2^2$, from the ensemble average of Eq. (68), is given by

$$\begin{aligned}\langle \chi_1^2 - \chi_2^2\rangle &= (N_{\rm ch} - N_{\rm ch2}) - \sum_{ij} \mathscr{A}_{1ij} \langle \Delta x_{1i}\,\Delta x_{1j}\rangle \\ &\quad + \sum_{ij} \mathscr{A}_{2ij} \langle \Delta x_{2i}\,\Delta x_{2j}\rangle \\ &= N_{\rm ch} - N_{\rm ch2} - L + L = N_{\rm ch} - N_{\rm ch2} = N_{\rm ch3}\end{aligned} \tag{69}$$

where $N_{\rm ch3}$ is the number of channels in the subset Ω_3, i.e. the total number of histogram channels $N_{\rm ch}$ minus the number of channels in the subset Ω_2, $N_{\rm ch2}$. The set–subset variance of χ^2 is defined as the mean square deviation of $\chi_1^2 - \chi_2^2$ from its mean value:

$$\sigma_{\chi_{12}^2}^2 \equiv \langle (\chi_1^2 - \chi_2^2 - \langle \chi_1^2 - \chi_2^2\rangle)^2\rangle = \langle (\chi_1^2 - \chi_2^2)^2\rangle - N_{\rm ch3}^2. \tag{70}$$

Evaluate

$$\begin{aligned}\langle(\chi_1^2-\chi_2^2)^2\rangle &= \sum_{n\in\Omega_3}\sum_{m\in\Omega_3}\frac{\langle(\mathscr{N}_n-f_\circ)^2(\mathscr{N}_m-f_\circ)^2\rangle}{(f_\circ)_n(f_\circ)_m}\\ &+\sum_{ijpq}\mathscr{A}_{1ij}\mathscr{A}_{1pq}\langle\Delta x_{1i}\,\Delta x_{1j}\,\Delta x_{1p}\,\Delta x_{1q}\rangle\\ &+\sum_{ijpq}\mathscr{A}_{2ij}\mathscr{A}_{2pq}\langle\Delta x_{2i}\,\Delta x_{2j}\,\Delta x_{2p}\,\Delta x_{2q}\rangle\\ &+2\sum_{ij}\mathscr{A}_{2ij}\sum_{n\in\Omega_3}\frac{1}{f_\circ}\langle(\mathscr{N}_n-f_\circ)^2\Delta x_{2i}\,\Delta x_{2j}\rangle\\ &-2\sum_{ij}\mathscr{A}_{1ij}\sum_{n\in\Omega_3}\frac{1}{f_\circ}\langle(\mathscr{N}_n-f_\circ)^2\Delta x_{1i}\,\Delta x_{1j}\rangle\\ &-2\sum_{ijpq}\mathscr{A}_{1ij}\mathscr{A}_{2pq}\langle\Delta x_{1i}\,\Delta x_{1j}\,\Delta x_{2p}\,\Delta x_{2q}\rangle. \quad (71)\end{aligned}$$

From the six terms in Eq. (71),

- the first term is equal to $N_{\mathrm{ch3}}^2+2N_{\mathrm{ch3}}$ (see Eq. (148) in Appendix);
- the second and third terms are equal to L^2+2L, see Eq. (150);
- in the fourth term, $\sum_n$ is the sum of the histogram channels $n\in\Omega_3$, whereas sums including the terms Δx_{2i} and Δx_{2j} (see Eq. (13)) are over histogram channels $n\in\Omega_2$. Therefore these two categories of sums are statistically independent and the fourth term becomes

$$\begin{aligned}&2\sum_{n\in\Omega_3}\frac{1}{f_\circ}\langle(\mathscr{N}_n-f_\circ)^2\rangle\sum_{ij}\mathscr{A}_{2ij}\langle\Delta x_{2i}\,\Delta x_{2j}\rangle\\ &\quad=2\sum_{n\in\Omega_3}\sum_{ij}\mathscr{A}_{2ij}(\mathscr{A}_2^{-1})_{ij}=2N_{\mathrm{ch3}}L \quad (72)\end{aligned}$$

- the fifth term is

$$\begin{aligned}&\sum_{ij}\mathscr{A}_{1ij}\sum_{n\in\Omega_3}\frac{1}{f_\circ}\langle(\mathscr{N}_n-f_\circ)^2\Delta x_{1i}\,\Delta x_{1j}\rangle\\ &\quad=\sum_{ij}\mathscr{A}_{1ij}\sum_{n\in\Omega_3}\frac{1}{f_\circ}\langle(\mathscr{N}_n-f_\circ)^2\rangle\langle\Delta x_{1i}\,\Delta x_{1j}\rangle\\ &\qquad+\sum_{ijpq}\mathscr{A}_{1ij}(\mathscr{A}_1^{-1})_{ip}(\mathscr{A}_1^{-1})_{jq}\sum_{n\in\Omega_3}\frac{f'_pf'_q}{f_\circ^3}\\ &\qquad\times[\langle(\mathscr{N}_n-f_\circ)^4\rangle-\langle(\mathscr{N}_n-f_\circ)^2\rangle^2]\\ &\quad=N_{\mathrm{ch3}}L+\sum_{pq}(\mathscr{A}_1^{-1})_{pq}\sum_{n\in\Omega_3}\frac{f'_pf'_q}{f_\circ}\left(2+\frac{1}{f_\circ}\right)\\ &\quad\approx N_{\mathrm{ch3}}L+2\sum_{ij}(\mathscr{A}_1^{-1})_{ij}\mathscr{A}_{3ij}. \quad (73)\end{aligned}$$

Here in the last step we neglect the term $1/f_\circ\ll1$ and introduce

$$\begin{aligned}\mathscr{A}_{3ij}&\equiv\sum_{n\in\Omega_3}\frac{f'_if'_j}{f_\circ}=\left(\sum_{n\in\Omega_1}\frac{f'_if'_j}{f_\circ}\right)-\left(\sum_{n\in\Omega_2}\frac{f'_if'_j}{f_\circ}\right)\\ &=\mathscr{A}_{1ij}-\mathscr{A}_{2ij}; \quad (74)\end{aligned}$$

- in the sixth term, we first separate out those parts of sums in Δx_{1i} and Δx_{1j}, which include only channels $n\in\Omega_3$ and hence are statistically independent of Δx_{2i} and Δx_{2j}. The remaining parts of Δx_{1i} and Δx_{1j} are denoted as $\Delta' x_{1i}$ and $\Delta' x_{1j}$. Thus we have

$$\begin{aligned}&\sum_{ijpq}\mathscr{A}_{1ij}\mathscr{A}_{2pq}\langle\Delta x_{1i}\,\Delta x_{1j}\,\Delta x_{2p}\,\Delta x_{2q}\rangle\\ &\quad=\sum_{ijpqab}\mathscr{A}_{1ij}\mathscr{A}_{2pq}(\mathscr{A}_1^{-1})_{ia}(\mathscr{A}_1^{-1})_{jb}\\ &\qquad\times\left(\sum_{n\in\Omega_3}\frac{f'_af'_b}{f_\circ^2}\langle(\mathscr{N}_n-f_\circ)^2\rangle\right)\langle\Delta x_{2p}\,\Delta x_{2q}\rangle\\ &\qquad+\sum_{ijpq}\mathscr{A}_{1ij}\mathscr{A}_{2pq}\langle\Delta' x_{1i}\Delta' x_{1j}\Delta x_{2p}\,\Delta x_{2q}\rangle. \quad (75)\end{aligned}$$

The first term in the right-hand side of Eq. (75) is

$$\begin{aligned}&\sum_{pq}\mathscr{A}_{2pq}(\mathscr{A}_2^{-1})_{pq}\times\sum_{ab}(\mathscr{A}_1^{-1})_{ab}\sum_{n\in\Omega_3}\frac{f'_af'_b}{f_\circ}\\ &\quad=L\times\sum_{ij}(\mathscr{A}_1^{-1})_{ij}\mathscr{A}_{3ij}. \quad (76)\end{aligned}$$

The second term in the right-hand side of Eq. (75) can be found in a similar way as in Eq. (150), with the result

$$(L+2)\times\sum_{ij}(\mathscr{A}_1^{-1})_{ij}\mathscr{A}_{2ij} \quad (77)$$

which coincides with that in Eq. (150) if $\mathscr{A}_1=\mathscr{A}_2$.

Finally, summing up all the terms listed above and using Eq. (74), we obtain an equation for the set–subset variance χ^2:

$$\begin{aligned}\sigma^2_{\chi^2_{12}}&=\langle(\chi_1^2-\chi_2^2-\langle\chi_1^2-\chi_2^2\rangle)^2\rangle\\ &=\langle(\chi_1^2-\chi_2^2)^2\rangle-\langle\chi_1^2-\chi_2^2\rangle^2\\ &=N_{\mathrm{ch3}}^2+2N_{\mathrm{ch3}}-N_{\mathrm{ch3}}^2=2N_{\mathrm{ch3}}\\ &=2N_{\mathrm{ch}}-2N_{\mathrm{ch2}}. \quad (78)\end{aligned}$$

6.3. Set–subset variance for the $(g-2)$ frequency: ω versus energy threshold

For systematic studies, we also compare the results of parameter optimization for the full set of data and for subsets corresponding to separate energy ranges. In general, the phases and asymmetries for the two sets are different, as shown in Section 1. The time distribution for such energy-range subsets has the same starting time t_{s}, ending time $t_{\max}$ and total number of histogram channels N_{ch} as the full set, but fewer counts in the channels: $\mathscr{N}_{n2}\leqslant\mathscr{N}_{n1}$. As before, we use subscripts 1 and 2 for the full set and subset, respectively, and subscript 3 for the complementary subset, such that

$$\mathscr{N}_{n1}=\mathscr{N}_{n2}+\mathscr{N}_{n3}. \quad (79)$$

We expect the time distribution for an energy-range subset of our data to follow the same functional form as the distribution for the full set (i.e. $G(t)$ from Eq. (2)), with some parameters having different "true" values for the full set and subset. We shall refer to the parameters which depend on the energy range as *set–subset variant* parameters, and those having the same "true" values, as *set–subset invariant*. For the muon $(g-2)$ data analysis, fit parameters $N_\circ$, A and ϕ are *set–subset variant* while τ and ω are *set–subset invariant* parameters. We note that the total number of events (decay electrons) and the statistical errors are *set–subset variant* quantities.

We wish to derive a general equation for the set–subset variance of a given *set–subset invariant* fit parameter, where the full data set has been divided into energy-range subsets. For that we consider a multi-parameter χ^2 fit in which one of the fit parameters, say x_i, is *set–subset invariant* and all others are, in general, *set–subset variant*. We denote the function to fit the full set of data as f_1 and that to fit the subset as f_2, meaning

$$\langle \mathscr{N}_{n1} \rangle = f(\boldsymbol{x}_\circ; t_n)_1 \equiv f_{\circ 1}$$

and

$$\langle \mathscr{N}_{n2} \rangle = f(\boldsymbol{x}_\circ; t_n)_2 \equiv f_{\circ 2} \tag{80}$$

in the same fashion as in Section 2. We also introduce function $f_3 \equiv f_1 - f_2$, which is easily shown to fit the complementary subset:

$$\begin{aligned} f(\boldsymbol{x}_\circ; t_n)_3 &= f(\boldsymbol{x}_\circ; t_n)_1 - f(\boldsymbol{x}_\circ; t_n)_2 = \langle \mathscr{N}_{n1} \rangle - \langle \mathscr{N}_{n2} \rangle \\ &= \langle \mathscr{N}_{n1} - \mathscr{N}_{n2} \rangle = \langle \mathscr{N}_{n3} \rangle. \end{aligned} \tag{81}$$

Since parameter x_i is *set–subset invariant* by assumption, the "true" value $x_{\circ i}$ is the same for the full set (x_{1i}) and for the subset (x_{2i}). Therefore, using Eq. (13) we have

$$\begin{aligned} x_{1i} - x_{2i} &= (x_{1i} - x_{\circ i}) - (x_{2i} - x_{\circ i}) = \Delta x_{1i} - \Delta x_{2i} \\ &= \sum_j (\mathscr{A}_1^{-1})_{ij} \sum_n \left(\frac{f'_j}{f_\circ} \right)_1 (\mathscr{N}_n - f_\circ)_1 \\ &\quad - \sum_j (\mathscr{A}_2^{-1})_{ij} \sum_n \left(\frac{f'_j}{f_\circ} \right)_2 (\mathscr{N}_n - f_\circ)_2 \\ &= \sum_j \sum_n \left[(\mathscr{A}_1^{-1})_{ij} \left(\frac{f'_j}{f_\circ} \right)_1 - (\mathscr{A}_2^{-1})_{ij} \left(\frac{f'_j}{f_\circ} \right)_2 \right] \\ &\quad \times (\mathscr{N}_n - f_\circ)_2 - \sum_j \sum_n (\mathscr{A}_1^{-1})_{ij} \left(\frac{f'_j}{f_\circ} \right)_1 (\mathscr{N}_n - f_\circ)_3 \end{aligned} \tag{82}$$

and $\langle x_{1i} - x_{2i} \rangle = 0$ (provided $\langle \mathscr{N}_{n2} - f_{\circ 2} \rangle = \langle \mathscr{N}_{n3} - f_{\circ 3} \rangle = 0$) as expected. Thus the energy-range set–subset variance of x_i is simply $\langle (x_{1i} - x_{2i})^2 \rangle$, which is found by averaging Eq. (82) squared:

$$\begin{aligned} \langle (x_{1i} - x_{2i})^2 \rangle &= \sum_{jk} f_{\circ 2} \Bigg[(\mathscr{A}_1^{-1})_{ij} \left(\frac{f'_j}{f_\circ} \right)_1 (\mathscr{A}_1^{-1})_{ik} \left(\frac{f'_k}{f_\circ} \right)_1 \\ &\quad - (\mathscr{A}_1^{-1})_{ij} \left(\frac{f'_j}{f_\circ} \right)_1 (\mathscr{A}_2^{-1})_{ik} \left(\frac{f'_k}{f_\circ} \right)_2 \\ &\quad - (\mathscr{A}_2^{-1})_{ij} \left(\frac{f'_j}{f_\circ} \right)_2 (\mathscr{A}_1^{-1})_{ik} \left(\frac{f'_k}{f_\circ} \right)_1 \\ &\quad + (\mathscr{A}_2^{-1})_{ij} \left(\frac{f'_j}{f_\circ} \right)_2 (\mathscr{A}_2^{-1})_{ik} \left(\frac{f'_k}{f_\circ} \right)_2 \Bigg] \\ &\quad + \sum_{jk} f_{\circ 3} (\mathscr{A}_1^{-1})_{ij} \left(\frac{f'_j}{f_\circ} \right)_1 (\mathscr{A}_1^{-1})_{ik} \left(\frac{f'_k}{f_\circ} \right)_1 \\ &= \sigma_{1i}^2 + \sigma_{2i}^2 - \sum_{jk} (\mathscr{A}_1^{-1})_{ij} (\mathscr{A}_2^{-1})_{ik} \\ &\quad \times \sum_n \frac{f'_{j1} f'_{k2} + f'_{j2} f'_{k1}}{f_{\circ 1}}. \end{aligned} \tag{83}$$

For the case of a one-parameter fit, Eq. (83) reduces to

$$\langle (x_1 - x_2)^2 \rangle = \sigma_1^2 + \sigma_2^2 - 2\sigma_1^2 \sigma_2^2 \sum_n \frac{f'_1 f'_2}{f_{\circ 1}}. \tag{84}$$

For the muon $(g-2)$ data analysis, we choose $t_s = -\tau$ and use the simpler Eq. (84) to find the energy-range set–subset variance of ω. First, we calculate the only unknown term in Eq. (84), $\sum_n \frac{f'_1 f'_2}{f_{\circ 1}} = \sum_n \frac{\partial G_1}{\partial \omega} \frac{\partial G_2}{\partial \omega} \frac{1}{G_1}$:

$$\begin{aligned} &\sum_n \frac{\partial G_1}{\partial \omega} \frac{\partial G_2}{\partial \omega} \frac{1}{G_1} \\ &\quad \approx \int_{-\tau}^{\infty} \frac{N_{\circ 2} A_1 A_2 t^2 \mathrm{e}^{-t/\tau} \sin(\omega t + \phi_1) \sin(\omega t + \phi_2)}{1 + A_1 \cos(\omega t + \phi_1)} \frac{\mathrm{d}t}{b} \\ &\quad \approx \frac{N_{\circ 2} A_1 A_2 \cos(\phi_1 - \phi_2)}{2b} \tau^3 \mathrm{e} \\ &\quad = \frac{A_1}{A_2} \cos(\phi_1 - \phi_2) \sigma_{2\omega}^{-2}. \end{aligned} \tag{85}$$

Here we use the equation for $1/b$:

$$\frac{1}{b} = \frac{N_2}{\int G_2 \, \mathrm{d}t} = \frac{N_2}{N_{\circ 2} \tau e} \tag{86}$$

to cancel $N_{\circ 2}$, and also replace $N_2 A_2^2 \tau^2 / 2$ by $\sigma_{2\omega}^{-2}$. Then for the energy-range set–subset variance of ω we have

$$\langle (\omega_1 - \omega_2)^2 \rangle = \sigma_{2\omega}^2 - \sigma_{1\omega}^2 \left(2 \frac{A_1}{A_2} \cos(\phi_1 - \phi_2) - 1 \right). \tag{87}$$

We note that if $A_1 = A_2$ and $\phi_1 = \phi_2$, then $\langle (\omega_1 - \omega_2)^2 \rangle = \sigma_{2\omega}^2 - \sigma_{1\omega}^2$ as in Eq. (67) for this set–subset variance.

7. Bias of the fit parameters in a χ^2 minimization fit

In Sections 2.2 and 2.3 we saw that the values of the fit parameters obtained in a minimization of $\chi^2 = \sum_n \frac{(f - \mathscr{N}_n)^2}{f}$

are statistically shifted with respect to the "true" values $x_{\circ i}$ by some $\Delta x_i : x_i = x_{\circ i} + \Delta x_i$, and that these shifts give rise to the statistical errors of the fit parameters $\sigma_i^2 \equiv \langle(\Delta x_i)^2\rangle$ and correlations $\langle \Delta x_i \Delta x_j\rangle$.In this section we investigate another important quantity, namely the systematic bias of the fit parameters, which is nothing more than an ensemble average of the shift: $\langle \Delta x_i\rangle$. In particular, it will be shown that if fits are made to data which are divided into a sufficient number of equivalent subsets, then the overall error will increase significantly.

In order to determine the bias, we rewrite Eq. (12) for the χ^2 minimization requirement in more detail:

$$
\begin{aligned}
0 = \frac{1}{2}\frac{\partial\chi^2}{\partial x_i} &= \sum_n \frac{f_i'}{f}\left(f - \mathcal{N}_n - \frac{1}{2}\frac{(f-\mathcal{N}_n)^2}{f}\right) \\
&= \sum_n \left[\frac{f_i' + \sum_j f_{ij}''\Delta x_j + \cdots}{f_\circ + \sum_j f_j'\Delta x_j + \cdots}\right. \\
&\quad \times \left(f_\circ - \mathcal{N}_n + \sum_j f_j'\Delta x_j + \frac{1}{2}\sum_{jk} f_{jk}''\Delta x_j\,\Delta x_k\right. \\
&\quad \left.\left. + \cdots - \frac{1}{2}\frac{(f_\circ - \mathcal{N}_n + \sum_j f_j'\Delta x_j + \cdots)^2}{f_\circ + \cdots}\right)\right] \qquad (88)
\end{aligned}
$$

and search for a solution in the form of successive approximations $\Delta x_i = \Delta x_i^\circ + \Delta x_i^1 + \cdots$ with the leading approximation being

$$
\Delta x_i^\circ = \sum_j (\mathscr{A}^{-1})_{ij} \sum_n \frac{f_j'}{f_\circ}(\mathcal{N}_n - f_\circ) \qquad (89)
$$

as in Eq. (13). Since $\langle \mathcal{N}_n - f_\circ\rangle = 0$, the ensemble average of Δx_i° vanishes. For the next-to-leading term Δx_i^1, which is of second order in $(\mathcal{N}_n - f(\boldsymbol{x}_\circ; t_n))/\mathcal{N}_n$, we have the equation

$$
\begin{aligned}
&\sum_n \frac{f_i'}{f_\circ}\left[\sum_p f_p'\Delta x_p^1 + \frac{1}{2}\sum_j\sum_k f_{jk}''\Delta x_j^\circ\,\Delta x_k^\circ\right. \\
&\quad - \frac{1}{2}\frac{(f_\circ - \mathcal{N}_n + \sum_j f_j'\Delta x_j^\circ)^2}{f_\circ} \\
&\quad + \left(-\frac{\sum_j f_j'\Delta x_j^\circ}{f_\circ} + \frac{\sum_j f_{ij}''\Delta x_j^\circ}{f_i'}\right) \\
&\quad \left.\times\left((f_\circ - \mathcal{N}_n) + \sum_k f_k'\Delta x_k^\circ\right)\right] = 0 \qquad (90)
\end{aligned}
$$

which has the solution

$$
\begin{aligned}
\Delta x_p^1 = \sum_i (\mathscr{A}^{-1})_{pi}\Bigg[&-\sum_j\sum_n \frac{f_i' f_j'}{f_\circ^2}(\mathcal{N}_n - f_\circ)\Delta x_j^\circ \\
&+ \sum_{jk}\sum_n \frac{f_i' f_j' f_k'}{f_\circ^2}\Delta x_j^\circ\,\Delta x_k^\circ \\
&+ \sum_j\sum_n \frac{f_{ij}''}{f_\circ}(\mathcal{N}_n - f_\circ)\Delta x_j^\circ \\
&- \sum_{jk}\sum_n \frac{f_{ij}'' f_k'}{f_\circ}\Delta x_j^\circ\,\Delta x_k^\circ \\
&- \frac{1}{2}\sum_{jk}\sum_n \frac{f_i' f_{jk}''}{f_\circ}\Delta x_j^\circ\,\Delta x_k^\circ \\
&+ \frac{1}{2}\sum_n \frac{f_i'}{f_\circ^2}(\mathcal{N}_n - f_\circ)^2 \\
&- \sum_j\sum_n \frac{f_i' f_j'}{f_\circ^2}(\mathcal{N}_n - f_\circ)\Delta x_j^\circ \\
&+ \frac{1}{2}\sum_{jk}\sum_n \frac{f_i' f_j' f_k'}{f_\circ^2}\Delta x_j^\circ\,\Delta x_k^\circ\Bigg]. \qquad (91)
\end{aligned}
$$

The systematic bias in the pth fit parameter from a χ^2 minimization is given by the ensemble average of Eq. (91). In the averaging, the first and third terms in square brackets cancel the second and fourth terms, respectively, and the seventh and eighth terms partially cancel each other. The net result is

$$
\begin{aligned}
\langle \Delta x_p\rangle = \langle \Delta x_p^1\rangle = &-\frac{1}{2}\sum_{ijk}(\mathscr{A}^{-1})_{pi}(\mathscr{A}^{-1})_{jk}\sum_n \frac{f_i' f_{jk}''}{f_\circ} \\
&+ \frac{1}{2}\sum_i (\mathscr{A}^{-1})_{pi}\sum_n \frac{f_i'}{f_\circ} - \frac{1}{2}\sum_{ijk}(\mathscr{A}^{-1})_{pi}(\mathscr{A}^{-1})_{jk} \\
&\times \sum_n \frac{f_i' f_j' f_k'}{f_\circ^2}. \qquad (92)
\end{aligned}
$$

For the case of one-parameter fit, Eq. (92) gives

$$
\langle \Delta x\rangle = -\frac{1}{2}\sigma_x^4\sum_n \frac{f' f''}{f_\circ} + \frac{1}{2}\sigma_x^2\sum_n \frac{f'}{f_\circ} - \frac{1}{2}\sigma_x^4\sum_n \frac{f'^3}{f_\circ^2}. \qquad (93)
$$

7.1. Bias on the fitted muon $(g-2)$ frequency

Let us estimate the bias on ω in the case that the data are divided into a number of subsets before fits to function $G(t)$ are performed. As before, we set the histogram start time at $-\tau$, which makes all parameters statistically independent, and allows the use of Eq. (93). Replacing sums by integrals, for the three terms in Eq. (93) we have

$$
\begin{aligned}
\langle \Delta\omega\rangle_1 &\approx \frac{1}{2}\left(\frac{\sqrt{2}}{\tau A\sqrt{N}}\right)^4 \times \frac{N}{\int_{-\tau}^{\infty} N_\circ e^{-t/\tau}[1 + A\cos(\omega t + \phi)]\,dt} \\
&\quad \times \int_{-\tau}^{t_{\max}} \frac{N_\circ e^{-t/\tau} A t\sin(\omega t + \phi) \times N_\circ e^{-t/\tau} A t^2\cos(\omega t + \phi)}{N_\circ e^{-t/\tau}[1 + A\cos(\omega t + \phi)]}\,dt \\
&\approx \frac{2}{\tau^4 A^2 N} \times \frac{1}{e\tau} \times \frac{1}{2\omega}[-e^{-t/\tau} t^3\cos 2(\omega t + \phi)]\Big|_{-\tau}^{t_{\max}} \\
&\sim \frac{1}{\omega\tau^2 A^2 N} \qquad (94)
\end{aligned}
$$

$$
\langle \Delta\omega\rangle_2 \approx \frac{1}{2}\left(\frac{\sqrt{2}}{\tau A\sqrt{N}}\right)^2 \frac{1}{b}\int_{-\tau}^{t_{\max}} \frac{N_\circ e^{-t/\tau} A t\sin(\omega t + \phi)}{N_\circ e^{-t/\tau}[1 + A\cos(\omega t + \phi)]}\,dt
$$

$$\approx \frac{1}{\omega^2\tau^2 ANb}[-\omega t\cos(\omega t+\phi)]\Big|_{-\tau}^{t_{\max}}$$
$$\sim \frac{t_{\max}+\tau}{\omega\tau^2 ANb} = \frac{N_{\rm ch}}{\omega\tau^2 AN} \quad (95)$$

where $(t_{\max}+\tau)/b = N_{\rm ch}\sim 3600$ is the typical number of histogram channels;

$$\langle\Delta\omega\rangle_3 \approx \frac{1}{2}\left(\frac{\sqrt{2}}{\tau A\sqrt{N}}\right)^4 \times \frac{N}{\int_{-\tau}^{\infty} N_\circ e^{-t/\tau}[1+A\cos(\omega t+\phi)]\,dt} \times \int_{-\tau}^{t_{\max}} \frac{(N_\circ e^{-t/\tau}At\sin(\omega t+\phi))^3}{(N_\circ e^{-t/\tau}[1+A\cos(\omega t+\phi)])^2}\,dt$$
$$\approx \frac{2}{\tau^4 AN}\times\frac{1}{e\tau}\times\frac{3}{4\omega}[-e^{-t/\tau}t^3\cos(\omega t+\phi)]\Big|_{-\tau}^{t_{\max}}$$
$$\sim \frac{3}{2\omega\tau^2 AN}. \quad (96)$$

Because of the factor $N_{\rm ch}\sim 3600$ in its numerator, $\langle\Delta\omega\rangle_2$, the second term in Eq. (93), is dominant. The ratio of the bias $\langle\Delta\omega\rangle \approx \langle\Delta\omega\rangle_2$ to the statistical uncertainty σ_ω is

$$\frac{\langle\Delta\omega\rangle}{\sigma_\omega} = \frac{N_{\rm ch}}{\omega\tau^2 AN}\times\left(\frac{\sqrt{2}}{\tau A\sqrt{N}}\right)^{-1} = \frac{N_{\rm ch}}{\omega\tau\sqrt{2N}}. \quad (97)$$

For a data set consisting of $N\sim 10^6$ decay electrons, we have $\langle\Delta\omega\rangle/\sigma_\omega = 0.03$, for $N\sim 10^9\langle\Delta\omega\rangle/\sigma_\omega = 10^{-3}$, etc. These are negligible, of course. However, if for technical reasons, for systematic studies, etc., the whole data set has to be divided into $N_{\rm div}$ equal parts and then fit separately, the weighted average being used as a final result, then the bias-to-σ_ω ratio is enhanced by a factor of $N_{\rm div}$. Therefore, in order to keep the bias below the statistical uncertainty, one may divide the entire data set in no more than $\sigma_\omega/\langle\Delta\omega\rangle = \omega\tau\sqrt{2N}/N_{\rm ch}$ equal parts. For the typical data set of 10^9 registered decay electrons, the whole data set may be divided into no more than 1000 equal parts.

7.2. Possible improvement of the χ^2 fit

In this section we show that the dominant (second) term in the expression for the bias on ω (Eqs. (94)–(96)) is typical for an arbitrary fit function, and discuss ways to eliminate it. The three terms in Eqs. (94)–(96) for the bias of fit parameters can be estimated in the general case as

$$\frac{\langle\Delta x\rangle_1}{x}\sim\frac{C_\sigma^4}{N^2}\times\frac{N}{\int f_\circ\,dt}\times\int\frac{f'f''}{f_\circ}\,dt = \frac{1}{N}\times\frac{C_\sigma^4 C_{12}}{C_\circ} \quad (98)$$

$$\frac{\langle\Delta x\rangle_2}{x}\sim\frac{C_\sigma^2}{N}\times\frac{1}{b}\int\frac{f'}{f_\circ}\,dt\sim\frac{1}{N}\times C_\sigma^2 C_1\frac{t_{\max}-t_{\rm s}}{b} = \frac{1}{N}\times C_\sigma^2 C_1\times N_{\rm ch} \quad (99)$$

$$\frac{\langle\Delta x\rangle_3}{x}\sim\frac{C_\sigma^4}{N^2}\times\frac{N}{\int f_\circ\,dt}\times\int\frac{f'^3}{f_\circ}\,dt = \frac{1}{N}\times\frac{C_\sigma^4 C_3}{C_\circ} \quad (100)$$

where $C_\circ, C_\sigma, C_{12}, C_1$ and C_3 are some functions of parameters of the fit function f. The dimensionless combinations of these functions in Eqs. (98)–(100) contain no "amplitude" parameters (like $N_\circ$ for the five-parameter fit function $G(t)$). Instead, because they contain only "shape" parameters (like τ, A, ω and ϕ for $G(t)$), they are *typically* of the same order, as in the case of Eqs. (94)–(96). Thus the $\langle\Delta x\rangle_2$ term, having a large extra factor $N_{\rm ch}$, is *expected* to be dominant for any fit function.

All the same, the bias in the fit parameters can be reduced and the quality of the fit, improved. We rewrite Eqs. (92) and (93) as

$$\langle\Delta x_p\rangle = -\frac{1}{2}\sum_{ijk}(\mathscr{A}^{-1})_{pi}(\mathscr{A}^{-1})_{jk}\sum_n\frac{f'_i f''_{jk}}{f_\circ} + \eta\sum_i(\mathscr{A}^{-1})_{pi}\sum_n\frac{f'_i}{f_\circ} - \eta\sum_{ijk}(\mathscr{A}^{-1})_{pi}(\mathscr{A}^{-1})_{jk}\sum_n\frac{f'_i f'_j f'_k}{f_\circ^2} \quad (101)$$

and

$$\langle\Delta x\rangle = -\frac{1}{2}\sigma^4\sum_n\frac{f'f''}{f_\circ} + \eta\sigma^2\sum_n\frac{f'}{f_\circ} - \eta\sigma^4\sum_n\frac{f'^3}{f_\circ^2} \quad (102)$$

with $\eta = \frac{1}{2}$. We recall that in our definition of χ^2 in Section 2, we use the equation $\sigma_n^2 = f(\boldsymbol{x}, t_n)$ to determine the variance σ_n^2 of the number of counts in the nth channel of the histogram. It can be shown (see for instance Ref. [10]) that for a χ^2 fit with $\sigma_n^2 = \mathscr{N}_n$ the bias of fit parameters is also given by Eqs. (101) and (102), but now with $\eta = -1$. Moreover, the bias for the log-likelihood fit, which may be considered as a minimization of the function $\mathscr{F}$:

$$\mathscr{F} \equiv \sum_n\left(\mathscr{N}_n\ln\frac{\mathscr{N}_n}{f} + f - \mathscr{N}_n\right) \quad (103)$$

is also given by Eqs. (101) and (102), with $\eta = 0$. Thus the second terms in Eqs. (101) and (102), which are dominant in the case of the χ^2 fit, are zero in the case of the log-likelihood fit. In the case where the full data set is divided into many equal subsets, we should expect the parameters determined by the likelihood method to be relatively free from bias. Moreover, the bias of the χ^2 fit with $\sigma_n^2 = f$ is less than that with $\sigma_n^2 = \mathscr{N}_n$ by a factor of -2. The troublesome second term can be eliminated by simply combining the two χ^2 fits in the proportion $-2:1$. In Ref. [10] we present this and several other ways to improve the quality of the χ^2 fit to the level of the log-likelihood fit. Perhaps, the most simple of them is to use a "corrected" χ^2

function, which we define as

$$\chi^2_{\rm corr} \equiv \sum_n \left(\frac{(f-\mathcal{N}_n)^2}{\mathcal{N}_n} - \frac{2}{3}\frac{(f-\mathcal{N}_n)^3}{\mathcal{N}_n^2} \right). \tag{104}$$

The function $\chi^2_{\rm corr}$ is nothing but the first two terms of Taylor's expansion of the function $\mathcal{F}$ from Eq. (103) in a power series of $(f-\mathcal{N}_n)/\mathcal{N}_n$.

N.B.: The improvement of the χ^2 fit, as discussed in this section, is relevant only in special cases, like that discussed in Section 7.1 where the data have to be divided into many parts. Otherwise, the χ^2 fit generally works perfectly well. In particular, χ^2 and log-likelihood fits to our time spectra give virtually identical results, well within statistical errors. Given that the maximum likelihood fit takes considerably more computer time, we use χ^2 minimization for most of our fits.

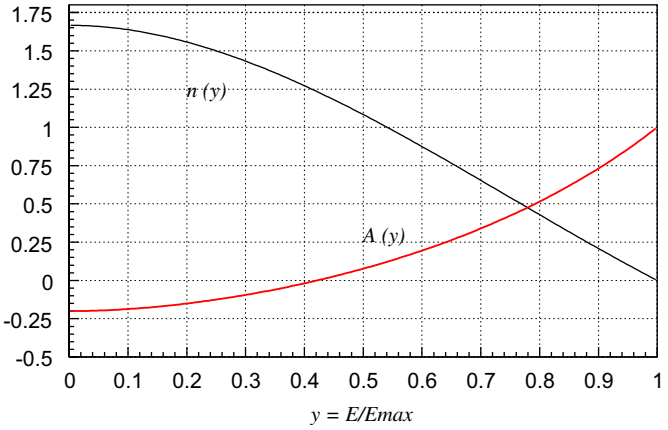

Fig. 4. Differential probability density of decay electrons $n(y)$ and asymmetry $A(y)$ as functions of relative energy $y = E/E_{\max}$.

8. Weighting method

The conventional approach used in data analysis for the Brookhaven experiment was to fit the spectra of the number of events above a given energy threshold versus time. In this section we study the advantage of weighting the entries in the time histogram by some function of the measured energy of decay electrons, in order to reduce the statistical error on the fitted value of ω.

As mentioned above, the electron decay asymmetry, A, is a function of electron energy. If $y = E/E_{\max}$ is the relative energy, then the number of events (or probability density) and the asymmetry as functions of y are given in Eqs. (105) and (106):

$$n(y) = \tfrac{1}{3}(y-1)(4y^2-5y-5)\varepsilon_n(y) \tag{105}$$

$$A(y) = \frac{-8y^2+y+1}{4y^2-5y-5}\varepsilon_A(y) \tag{106}$$

respectively. The effects of detector acceptance (including finite energy resolution) are accounted for by $\varepsilon_n(y)$ and $\varepsilon_A(y)$. The functions $n(y)$ and $A(y)$ are plotted in Fig. 4 for the case of $\varepsilon_n = 1$ and $\varepsilon_A = 1$. We note that in this case

$$\int_0^1 n(y)\,{\rm d}y = 1 \quad \text{and} \quad \int_0^1 n(y)A(y)\,{\rm d}y = 0. \tag{107}$$

In the following we assume that $n(y)$ is normalized as in Eq. (107)

Earlier we showed, that if we include only decay electrons of energy greater than $E_{\rm thr}$ in our analysis, a threshold of roughly 1.8 GeV (which maximizes NA^2) will minimize the statistical error. In practice, the detector system records electrons with y as small as 0.2, so that the lower part of the spectrum has an asymmetry opposite in sign to the high-energy one. Direct inclusion of those decays would decrease the average value of NA^2 and thereby increase the statistical error. We would clearly do better to assign negative weights to the low-energy events. More generally, we can introduce some weight function $p(y)$ and consider a modified probability density function (pdf).

Now the two-dimensional (time t and relative energy y) pdf for $G(t,y)$ is

$$\begin{aligned}\mathcal{P}(t,y) &= \frac{G(t,y)}{\int_{-\tau}^{\infty}\int_0^1 G(t,y)\,{\rm d}y\,{\rm d}t} \\ &\approx \frac{n(y)}{\tau e}{\rm e}^{-t/\tau}(1+A(y)\cos(\omega t+\phi)).\end{aligned} \tag{108}$$

For convenience we consider the binned pdf, $\mathcal{P}_{mq}$, where the first subscript refers to the time bin and the second one to the energy bin:

$$\begin{aligned}\mathcal{P}_{mq} &= \frac{n_q}{\tau e}{\rm e}^{-t_m/\tau}(1+A_q\cos(\omega t_m+\phi))b_y b_t \\ &= \frac{b_y b_t}{\tau e}{\rm e}^{-t_m/\tau}(n_q + n_q A_q\cos(\omega t_m+\phi)).\end{aligned} \tag{109}$$

Here the corresponding bin widths b_t and b_y are explicitly included. Multiplying $\mathcal{P}_{mq}$ by the total number of events N, one gets the expected number of events in the (m,q)th bin. Summing up all such events over a threshold $q_{\rm thr}$ (corresponding to $y_{\rm thr}$) for fixed m, one gets the number of events in the mth channel of the time distribution:

$$\begin{aligned}\mathcal{N}_m &\equiv \sum_{q>q_{\rm thr}} N\mathcal{P}_{mq} \\ &= N\frac{b_t b_y}{\tau e}{\rm e}^{-t_m/\tau}(1+\langle A\rangle_y\cos(\omega t_m+\phi))\end{aligned} \tag{110}$$

where

$$\langle A\rangle_y \equiv \frac{\sum_{q>q_{\rm thr}} n_q A_q}{\sum_{q>q_{\rm thr}} n_q} \tag{111}$$

is the average asymmetry, A, for events with energy above the threshold $y_{\rm thr}$. In the conventional analysis, the distribution in Eq. (110) is fit with the five-parameter function $G(t)$.

Now, for a *weighted* pdf we have

$$\mathscr{P}_{mq}p_q = p_q\frac{b_t b_y}{\tau e}n_q e^{-t_m/\tau}(1+A_q\cos(\omega t_m+\phi)) = \frac{b_t b_y}{\tau e}e^{-t_m/\tau}(n_q p_q + n_q p_q A_q\cos(\omega t_m+\phi)) \quad (112)$$

and the content[3] of the mth time bin is

$$\mathscr{N}^p_m \equiv \sum_{q>q_{\rm thr}} N\mathscr{P}_{mq}p_q \quad (113)$$

$$= N\frac{b_t b_y}{\tau e}e^{-t_m/\tau}(\langle p\rangle_y + \langle pA\rangle_y\cos(\omega t_m+\phi)). \quad (114)$$

This time distribution may be fit with the five-parameter function:

$$f(t) = e^{-t/\tau}(C + A\cos(\omega t+\phi)) \quad (115)$$

where, as compared to function $G(t)$, we remove the overall normalization parameter $N_\circ$ in favor of the offset C because it better handles the situation where $C = \langle p\rangle_y \approx 0$.

Next, we need to evaluate the statistical error for the mth channel of the time distribution given in Eq. (114). We rewrite Eq. (113) to show the statistical fluctuations of all terms in the sum explicitly:

$$\mathscr{N}^p_m = \sum_{q>q_{\rm thr}} N\mathscr{P}_{mq}p_q = \sum_{q>q_{\rm thr}}(N\mathscr{P}_{mq} \pm \sqrt{N\mathscr{P}_{mq}})p_q \quad (116)$$

and obtain the statistical error for $\mathscr{N}^p_m$:

$$\sigma^2_m \equiv \langle\mathscr{N}^{p2}_m\rangle - \langle\mathscr{N}^p_m\rangle^2 = \sum_{q>q_{\rm thr}} N\mathscr{P}_{mq}p_q^2 \quad (117)$$

$$= N\frac{b_t b_y}{\tau e}e^{-t_m/\tau}[\langle p^2\rangle_y + \langle p^2A\rangle_y\cos(\omega t_m+\phi)]. \quad (118)$$

Now we construct the χ^2 function as in Eq. (4) and, following the steps outlined in Section 2.2, derive the equation for the error matrix:

$$\langle\Delta x_i\,\Delta x_j\rangle = (\mathscr{A}^{-1})_{ij} \quad \text{where } \mathscr{A}_{ij} = \sum_m \frac{f'_i f'_j}{\sigma^2_m} \quad (119)$$

which is actually the same as in Eqs. (17) and (14) respectively.[4] Then we replace the sum in Eq. (119) by an integral, diagonalize the matrix $\mathscr{A}$ with the choice of histogram start time $t_s = -\tau$, and calculate the corresponding diagonal matrix element for the fit parameter ω, which is in our case σ^{-2}_ω:

$$\sigma^{-2}_\omega = \frac{N}{\tau e}\int_{-\tau}^{\infty}\frac{t^2 e^{-t/\tau}\langle pA\rangle_y^2\sin^2(\omega t+\phi)}{\langle p^2\rangle_y + \langle p^2A\rangle_y\cos(\omega t+\phi)}\,dt \approx \frac{N\langle pA\rangle_y^2\tau^2}{2\langle p^2\rangle_y}. \quad (120)$$

For a practical implementation of the weighting method, we form a histogram as a function of m (time) by summing $N\mathscr{P}_{mq}$ over some chosen subset of the energy, q, in Eqs. (113) and (117). Then we fit the resulting distribution with $G(t)$, or with the function given in Eq. (115), and obtain fit values and statistical errors for ω and other fit parameters.

8.1. Weighting with $p(y) = A(y)$ and $p(y) = y$

In our experiment, we use two weighting functions: $p(y) = A(y)$ and the conventional $p(y) = 1$. Fig. 5 shows the figure of merit σ^{-2}_ω (in relative units) as a function of energy threshold for these weighting functions and also for the case $p(y) = y$. Of these, it is evident that the weighting method with $p(y) = A(y)$ gives the best statistical precision for ω. It is, in fact, *the* best weighting method, as shown in the next section. Using the asymmetry weighting with an energy threshold of $y_{\rm thr}\sim 0.3$ reduces the statistical error on ω by 10% or more.

Another virtue of the $p(y) = y$ weighting method is that it can substantially reduce systematic errors caused by *pile-up*, when two decay electrons of fractional energies y_1 and y_2, respectively, are detected simultaneously (i.e. piled-up) and are counted as a single electron of energy $y = y_1 + y_2$. The pile-up contribution to the energy dependent time spectrum can be described by the equation:

$$f_{pu}(t,y) = \left(\frac{1}{2}\int_0^y \mathscr{P}(t,y')\mathscr{P}(t,y-y')\,dy'\right) - \left(\mathscr{P}(t,y)\int_0^1 \mathscr{P}(t,y')\,dy'\right) \quad (121)$$

where the first term is the probability of two events at energies y' and $y-y'$ to pile-up and form an event at energy y. The second term is the probability of an event at energy y to pile-up with any other event and thus no longer contribute to the distribution at an energy y. With the help of the characteristic function of the pdf one may verify that

$$\int_0^2 y f_{pu}(t,y)\,dy = 0 \quad (122)$$

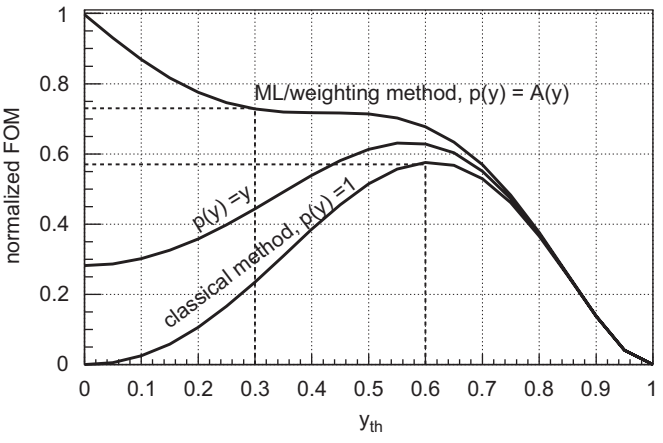

Fig. 5. Figure of merit σ^{-2}_ω as a function of energy threshold $y_{\rm thr} = E_{\rm thr}/E_{\rm max}$ for the case $\varepsilon_n = 1$ and $\varepsilon_A = 1$. The ML weighting refers to the optimal case which follows from a maximum likelihood argument in the text.

[3]Here it is a weighted sum of events, thus not integer in general.

[4]We note that Eq. (119) for $\mathscr{A}_{ij}$ is more general than Eq. (14) because σ^2_m is, in turn, more general than $\sigma^2_m = f_\circ$ in Eq. (14).

and hence the pile-up contribution vanishes (for $y_{\rm thr}=0$) or is greatly reduced using a weight $p(y)=y$. If the acceptance parameters ε_N and ε_A are both equal to 1, all particles would be accepted and the energy-weighted asymmetry would be small. However, the detectors are designed so that the acceptance is low for low-energy electrons and high for high-energy electrons, and in this circumstance, the energy-weighting works reasonably well.

8.2. Cramér–Rao bound for the five-parameter distribution $G(t)$

The theoretical lower limit for the precision of a single or non-correlated fit parameter for a given event distribution (e.g. parameter ω for the time–energy distribution $\mathscr{P}(t,y)$ if the histogram start time is chosen to be $-\tau$) is set by the Cramér–Rao limit [11]:

$$\frac{1}{\sigma_\omega^2}\leqslant\left\langle\left(\frac{\mathrm{d}\ln L}{\mathrm{d}\omega}\right)^2\right\rangle_{\rm pdf}$$
$$\text{or, equivalently,}\frac{1}{\sigma_\omega^2}\leqslant-\left\langle\frac{\mathrm{d}^2\ln L}{\mathrm{d}\omega^2}\right\rangle_{\rm pdf}\tag{123}$$

where $\ln L$ is the (unbinned) log-likelihood function, and averaging over the pdf (or over collected events) is assumed. Thus, the lowest bound corresponds to the maximum likelihood method.

Since the probability of a particular, kth, event is

$$L_k=\frac{n(y_k)}{\tau e}\mathrm{e}^{-t_k/\tau}(1+A(y_k)\cos(\omega t_k+\phi))\tag{124}$$

the total log-likelihood function is

$$\ln L=\sum_{k=1}^{N}\ln L_k=\sum_{k=1}^{N}\ln n(y_k)-N\ln\tau-N-\sum_{k=1}^{N}\frac{t_k}{\tau}+\sum_{k=1}^{N}\ln(1+A(y_k)\cos(\omega t_k+\phi)).\tag{125}$$

Applying Eq. (123) we have

$$\frac{1}{\sigma_\omega^2}\leqslant\left(\frac{\mathrm{d}\ln L}{\mathrm{d}\omega}\right)^2=\sum_{k=1}^{N}t_k^2A^2(y_k)\left(\frac{\sin(\omega t_k+\phi)}{1+A(y_k)\cos(\omega t_k+\phi)}\right)^2$$
$$\approx\frac{1}{2}\sum_{k=1}^{N}t_k^2A^2(y_k)\equiv\frac{N}{2}\langle t^2A^2\rangle_{t,y}=\frac{N}{2}\tau^2\langle A^2\rangle_y\tag{126}$$

where in the last step we use the fact that asymmetry is not time-dependent, and as before, τ^2 is the average value of t^2 in the range $-\tau$ to infinity.

Thus for the figure of merit, σ_ω^{-2}, we obtain

$$\sigma_\omega^{-2}=\tfrac{1}{2}N\tau^2\langle A^2\rangle_y\tag{127}$$

as the Cramér–Rao lower bound for our case. By comparison, the σ_ω^{-2} value obtained in Section 3 for the classical method, $p=1$, may be written as

$$\sigma_\omega^{-2}=\tfrac{1}{2}N\tau^2\langle A\rangle_y^2\tag{128}$$

where $\langle A\rangle_y$ is the decay electron asymmetry averaged over electrons above the energy threshold, $y>y_{\rm thr}$. Since $\langle A\rangle_y^2<\langle A^2\rangle_y$, the classical method is not optimal. The weighting method with weight function $p(y)=A(y)$, on the contrary, gives rise to the combination $\langle A^2\rangle_y$:

$$\sigma_\omega^{-2}=\frac{1}{2}N\tau^2\frac{\langle pA\rangle_y^2}{\langle p^2\rangle_y}=\frac{1}{2}N\tau^2\langle A^2\rangle_y\tag{129}$$

see Eq. (120), and hence reaches the Cramér–Rao limit. The optimal weighting is uniquely given by asymmetry weighting. The result does not depend on the particular y dependence of $n(y)$ or $A(y)$, which are taken from the data.

9. The ratio method—a new method of extracting the anomalous precession frequency

In the ratio method, the full data set is randomly split into four independent subsets of equal size. The corresponding time spectra from the four subsets are combined so that the exponential decay of the muon and the overall normalization are effectively divided out; the resulting spectrum is fit for the amplitude, frequency and phase of the $(g-2)$ oscillation. An advantage of the ratio method is that, like the exponential decay itself, any slowly varying multiplicative distortion of the spectrum will be much reduced.

The three-parameter ratio fitting function $r(t)$ is derived from the five-parameter function $G(t)$. We introduce the functions

$$u_+(t)\propto\tfrac{1}{4}G(t+T/2)=\tfrac{1}{4}N_\circ\mathrm{e}^{-t/\tau}\mathrm{e}^{-T/2\tau}\times[1-A\cos(\omega t+\phi+\tfrac{1}{2}\omega\delta_T)]\tag{130}$$

$$u_-(t)\propto\tfrac{1}{4}G(t-T/2)=\tfrac{1}{4}N_\circ\mathrm{e}^{-t/\tau}\mathrm{e}^{T/2\tau}\times[1-A\cos(\omega t+\phi-\tfrac{1}{2}\omega\delta_T)]\tag{131}$$

$$v_1(t)\propto\tfrac{1}{4}G(t)=\tfrac{1}{4}N_\circ\mathrm{e}^{-t/\tau}[1+A\cos(\omega t+\phi)]\tag{132}$$

$$v_2(t)\propto\tfrac{1}{4}G(t)=\tfrac{1}{4}N_\circ\mathrm{e}^{-t/\tau}[1+A\cos(\omega t+\phi)]\tag{133}$$

which describe the four statistical data sets shifted in time by either one half of a $(g-2)$ period ($u_\pm$), or not shifted ($v_{1,2}$). Here the fixed parameter T is the nominal $(g-2)$ period, known *a priori* (e.g. from previous experiments) to 5–10 ppm accuracy, which is sufficient in this application. The difference from the true value is given by δ_T.

The function $r(t)$ is constructed as

$$r(t)=\frac{v_1(t)+v_2(t)-u_+(t)-u_-(t)}{v_1(t)+v_2(t)+u_+(t)+u_-(t)}.\tag{134}$$

It can be simplified by the use of successive approximations on small parameters $(T/4\tau)^2=288\,\text{ppm}$ and $\omega\delta_T/2=\pi\times10\,\text{ppm}=31.4\,\text{ppm}$. Keeping terms greater than 0.1 ppm, we have

$$r(t)=-A\cos(\omega t+\phi)+\left(\frac{T}{4\tau}\right)^2-\left(\frac{T}{4\tau}\right)^2A^2\cos^2(\omega t+\phi).\tag{135}$$

In fact, once the last term was shown to have a negligible effect on the fitted value for ω, the three-parameter function

$$r(t) = -A\cos(\omega t + \phi) + 2.88 \times 10^{-4} \tag{136}$$

was used instead.

We note that, if the weighting of the counts in the histograms were $u_+ : u_- : v_1 : v_2 = e^{T/2\tau} : e^{-T/2\tau} : 1 : 1$ instead of $1:1:1:1$, then the last two terms in Eq. (135) would be eliminated but, as mentioned, the approximations of that equation are sufficiently good for this case.

The statistical fluctuation of $r(t_n)$ is driven by the statistical fluctuations of $v_{1,2}(t_n)$ and $u_{+-}(t_n)$:

$$\sigma^2_{v_1}(t_n) = v_1(t_n); \quad \sigma^2_{v_2}(t_n) = v_2(t_n); \quad \sigma^2_{u_+}(t_n) = u_+(t_n)$$
$$\sigma^2_{u_-}(t_n) = u_-(t_n) \tag{137}$$

giving

$$\sigma^2_r(t_n) = \frac{1 - r^2(t_n)}{v_1(t_n) + v_2(t_n) + u_+(t_n) + u_-(t_n)}. \tag{138}$$

To a good approximation, $1 - r^2(t_n) \approx 1$ and

$$v_1(t_n) + v_2(t_n) + u_+(t_n) + u_-(t_n) \approx N_\circ e^{-t_n/\tau} \tag{139}$$

therefore $\sigma^2_r(t_n)$ can be written as

$$\sigma^2_r(t_n) = (N_\circ e^{-t_n/\tau})^{-1}. \tag{140}$$

The χ^2 function can be constructed as in Eq. (4) and the equation for the error matrix is

$$\langle \Delta x_i \, \Delta x_j \rangle = (\mathscr{A}^{-1})_{ij}$$
$$\text{where now } \mathscr{A}_{ij} = \sum_n \frac{r'_i r'_j}{\sigma^2(t_n)} = \sum_n N_\circ e^{-t_n/\tau} r'_i r'_j. \tag{141}$$

The equations for the matrix elements in Eq. (141) are very close to the corresponding equations for the fit parameters A, ω and ϕ of the five-parameter fit. Indeed, the matrix $\mathscr{A}$ for the ratio fit method is (to leading order in $\varepsilon = 0.011$ and $\frac{1}{2}A^2 = 0.08$)

$$\mathscr{A} = \begin{bmatrix} \frac{N}{2} & 0 & 0 \\ 0 & \frac{NA^2\tau^2}{2}\left[\left(\frac{t_s}{\tau}+1\right)^2 + 1\right] & \frac{NA^2\tau}{2}\left(\frac{t_s}{\tau}+1\right) \\ 0 & \frac{NA^2\tau}{2}\left(\frac{t_s}{\tau}+1\right) & \frac{NA^2}{2} \end{bmatrix} \tag{142}$$

which coincides with the right bottom corner of the matrix $\mathscr{A}$ in Eq. (24) for the five-parameter fit. Thus the equations for the statistical errors of the fit parameters A, ω and ϕ are given in Eqs. (27)–(29), respectively.

Finally, we note that while slow modulations ($T \gg \frac{2\pi}{\omega}$) largely cancel in the ratio, fast modulations, such as those caused the CBO, are not canceled, and must be dealt within the same way as for the conventional analysis.

10. The folding method

In the *folding method*, a simple but powerful method of checking for the presence of a periodic process (period T') in a time distribution, the distribution is cut into time slices of duration T' which are then re-summed. Mathematically, this is equivalent to the transformation $t \to t'$, where

$$t' = \frac{t}{T'} - \text{integer}\left(\frac{t}{T'}\right). \tag{143}$$

As an illustration, in Fig. 6 we give the positron time distribution from a 1997 "engineering" run [1]. Sharp background peaks with spacing 2.7 μs are clearly seen together with the $(g-2)$ oscillations which have a period of 4.37 μs. The 2.7 μs peaks were produced by a problem, later remedied, in extracting protons from the BNL Alternating Gradient Synchrotron (AGS).

In Fig. 7 the data from the same run are plotted as a function of t'. Note that t' varies from 0 to 1, the latter corresponding to the length of one period, T'. The highest peak at $t' = 0.90$ corresponds to the 2.7 μs structure seen in Fig. 6. Seven smaller peaks are also clearly seen. All together, the eight separate peaks correspond to the eight bunches of protons stored at uniform intervals around the AGS ring.

The ability to resolve background structure in t' is very sensitive to the exact value of the re-summation period T' and becomes sharpest for $T' = 2.694$ μs, the true AGS cyclotron period (the case actually shown in Fig. 7). Because even a small contamination of these AGS *flashlets* can lead to a significant systematic error in ω, this exceptionally sensitive tool was essential to the accuracy of our result.

The folding method is largely complementary to the Fourier analysis method. The Fourier method, which we use extensively for our systematics studies (see Ref. [6]), is usually superior if one is searching for an unknown

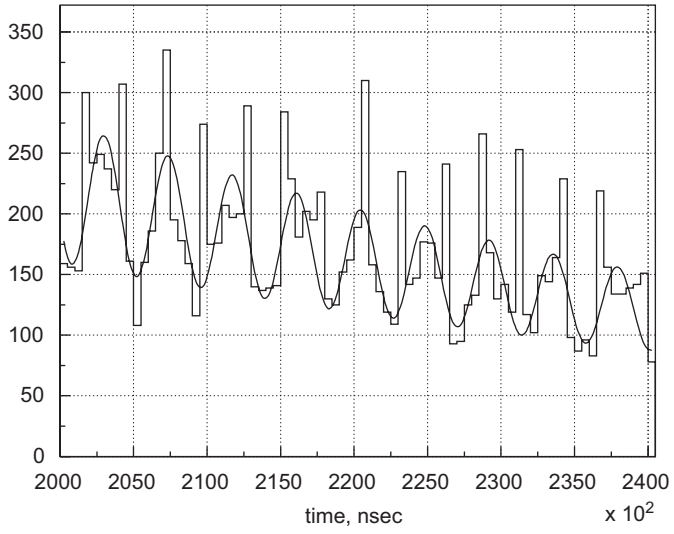

Fig. 6. Fragment of decay positrons time distribution from a 1997 "engineering" run.

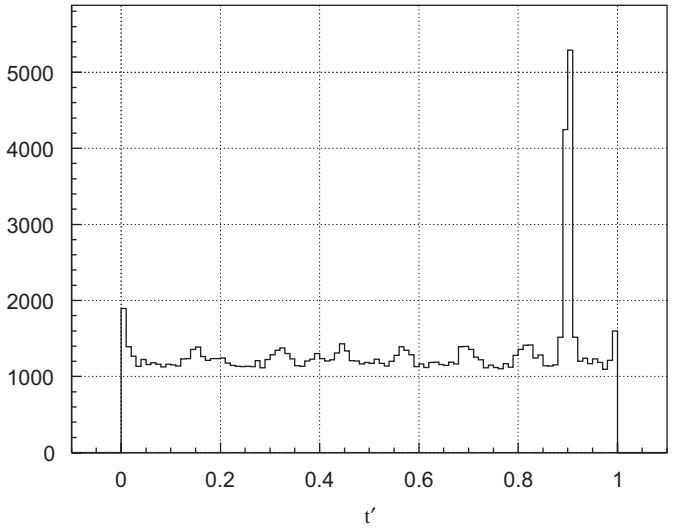

Fig. 7. Distribution of t' for the same run as in Fig. 6.

periodic process. In turn, the folding method has an advantage for the detection of a process with known period, especially if the functional form of the process is far from a pure sine wave. Another important advantage of the folding method is that the distribution of t' is a regular event distribution, $\{\mathcal{N}_n\}$, with statistical fluctuations $\sqrt{\mathcal{N}_n}$ and with no correlation among different channels. In that respect, the Fourier transformation plot is more complicated for analysis.

In our search for CPT/Loretz violating oscillation in a_μ versus sideral time (to be published elsewhere) the limits we get are the same for χ^2 and Fourier analyses and for the folding method.

11. Kolmogorov–Smirnov test

The Kolmogorov–Smirnov compatibility test is a standard statistical procedure that can be used to determine whether or not two data sets, consisting of N_1 and N_2 entries, have the same underlying parent distribution. From each data set, one constructs a normalized cumulative distribution which grows monotonically from 0 to 1. These two cumulative distributions are then compared, point by point. The largest difference between the two, d_K, is used as a measure of the compatibility of the two distributions.

It was shown by Kolmogorov [12] that when comparing two samples from the same parent distribution, d_K follows a distribution which is *independent* of the parent distribution. In the limit of infinite data sets, the distribution of d_K is such that the probability of d_K to exceed some particular $d_{K\circ}$ is given by

$$P(d_K > d_{K\circ}) = 2\sum_{k=1}^{\infty}(-1)^{k-1}\exp\left[-2k^2 d_{K\circ}^2\left(\frac{N_1 N_2}{N_1+N_2}\right)\right]. \tag{144}$$

The sensitivity of this test is not uniform. It tends to be more sensitive near the median value and to be less sensitive at the extreme ends of the cumulative distribution. Several variants that improve the uniformity have been proposed in Refs. [13,14].

The probabilities for small data sets ($N_1, N_2 \leqslant 80$) are commonly available in tables. In general, they can be generated using a Monte Carlo technique which simulates statistical fluctuations of otherwise identical distributions.

In our experiment, we mostly used the Kolmogorov–Smirnov test and its variants for selection or rejection of decay electron/positron data, which come in separate short runs (10^4–10^5 events) and from separate detectors. Runs were rejected when data from a large number of detectors showed low probability ($P < 10^{-2}$). Likewise detectors were rejected when their data had $P < 10^{-2}$ for a sequence of runs. The probability distribution of accepted data was uniform. In addition, an effective detector gain correction procedure based on the Kolmogorov–Smirnov test was developed.

There are several advantages of the Kolmogorov–Smirnov test. The parent distribution does not need to be known, which reduces the possibility that some bias to the derived value of ω could be introduced by the data sample selection. In addition, the test does not involve any fits or iteration procedures, nor does it require the data to be binned.

12. Conclusions

In this paper, we have described some of the basic statistical ideas which have been applied to the analysis of the muon $(g-2)$ data. We have obtained equations for statistical errors, correlations of fit parameters and other properties of a χ^2 fit of the muon $(g-2)$ time spectra. These include

- statistical errors and correlations of fit parameters;
- reducing the uncertainty on fit parameters by incorporating additional knowledge, such as external knowledge of the precession phase;
- systematic shifts of fit parameters due to neglected background;
- comparisons of fit parameters obtained from fitting the full set of data versus some subset of the data (set–subset relations);
- systematic biases of the fit parameters from a χ^2 minimization procedure and bias reduction techniques;
- optimizing the statistical power of the data by weighting individual events by energy or asymmetry;
- the ratio method, in which the muon decay is effectively removed thus simplifying the fit for ω;
- the "folding" method for detecting a periodic background process in a time distribution;
- the Kolmogorov–Smirnov histogram compatibility test, which we used extensively in the screening of data runs.

The results obtained in this study, as well as the more general statistical equations included in this paper, can be applied to fits obtained in many kinds of experiments, but particularly those with periodic signals.

Similar studies may be found in literature. For instance, Refs. [15,16] explicitly deal with the bias of least-squares fits to Poisson distributed data and present solutions like maximum likelihood estimation based on Poisson statistic [15] or other alternatives [16]. Ref. [17] shows that the uncertainty derived from a least squares is rather meaningless if the fitted function does not rigorously apply to the data. The "long-term effects" in Ref. [17] seem related to Section 5 of this paper ("systematic shift due to neglected background").

Acknowledgments

This work is partially supported by the Russian Foundation for Basic Research, Grants 06-02-16156, 06-02-16445 and the U.S. National Science Foundation.

Appendix

Here we calculate the ensemble average of $(\chi^2)^2$, which is

$$\begin{aligned}\langle(\chi^2)^2\rangle &= \sum_n\sum_m \frac{\langle(\mathcal{N}_n-f_\circ)^2(\mathcal{N}_m-f_\circ)^2\rangle}{(f_\circ)_n(f_\circ)_m} \\ &\quad - 2\sum_{ij}\mathscr{A}_{ij}\left\langle \Delta x_i\,\Delta x_j \sum_n \frac{1}{f_\circ}(\mathcal{N}_n-f_\circ)^2\right\rangle \\ &\quad + \sum_{ijpq}\mathscr{A}_{ij}\mathscr{A}_{pq}\langle \Delta x_i\,\Delta x_j\,\Delta x_p\,\Delta x_q\rangle \end{aligned} \tag{145}$$

as it follows from Eq. (20). The average of $(\mathcal{N}_n-f_\circ)^2(\mathcal{N}_m-f_\circ)^2$ in the first term in Eq. (145) is equal to $(f_\circ)_n(f_\circ)_m$ for the case $n\neq m$, while for the case $n=m$ we have $\langle(\mathcal{N}_n-f_\circ)^4\rangle = f_\circ+3f_\circ^2$, see Eq. (10). Thus for the first case we have

$$\begin{aligned}&\sum_{n\neq m}\sum_m \frac{\langle(\mathcal{N}_n-f_\circ)^2(\mathcal{N}_m-f_\circ)^2\rangle}{(f_\circ)_n(f_\circ)_m} \\ &\quad = \left(\sum_n \frac{\langle(\mathcal{N}_n-f_\circ)^2\rangle}{f_\circ}\right)^2 - \sum_n \frac{\langle(\mathcal{N}_n-f_\circ)^2\rangle^2}{f_\circ^2}\end{aligned} \tag{146}$$

and for the second case

$$\sum_{n=m}\sum_m \frac{\langle(\mathcal{N}_n-f_\circ)^2(\mathcal{N}_m-f_\circ)^2\rangle}{(f_\circ)_n(f_\circ)_m} = \sum_n \frac{\langle(\mathcal{N}_n-f_\circ)^4\rangle}{f_\circ^2} \tag{147}$$

and hence

$$\begin{aligned}&\sum_n\sum_m \frac{\langle(\mathcal{N}_n-f_\circ)^2(\mathcal{N}_m-f_\circ)^2\rangle}{(f_\circ)_n(f_\circ)_m} \\ &\quad = \left(\sum_n \frac{\langle(\mathcal{N}_n-f_\circ)^2\rangle}{f_\circ}\right)^2 - \sum_n \frac{\langle(\mathcal{N}_n-f_\circ)^2\rangle^2}{f_\circ^2} \\ &\qquad + \sum_n \frac{\langle(\mathcal{N}_n-f_\circ)^4\rangle}{f_\circ^2} \\ &\quad = \left(\sum_{n=1}^{N_{\rm ch}} 1\right)^2 - \sum_{n=1}^{N_{\rm ch}} 1 + \left(\sum_{n=1}^{N_{\rm ch}}\frac{1}{f_\circ}+\sum_{n=1}^{N_{\rm ch}} 3\right) \\ &\quad = N_{\rm ch}^2 + 2N_{\rm ch}\end{aligned} \tag{148}$$

where in the last step we neglect the term $1/f_\circ \ll 1$.

In the second term in Eq. (145), the combination of $\Delta x_i\,\Delta x_j$ gives a non-vanishing contribution only when in the decomposition

$$\begin{aligned}\Delta x_i\,\Delta x_j = &\sum_{pq}(\mathscr{A}^{-1})_{ip}(\mathscr{A}^{-1})_{jq}\sum_v \frac{f'_p}{f_\circ}(\mathcal{N}_v-f_\circ) \\ &\times \sum_w \frac{f'_q}{f_\circ}(\mathcal{N}_w-f_\circ)\end{aligned}$$

indices v and w are equal. Therefore

$$\begin{aligned}&\sum_{ij}\mathscr{A}_{ij}\left\langle \Delta x_i\,\Delta x_j \sum_n \frac{1}{f_\circ}(\mathcal{N}_n-f_\circ)^2\right\rangle \\ &\quad = \sum_{ijpq}\mathscr{A}_{ij}(\mathscr{A}^{-1})_{ip}(\mathscr{A}^{-1})_{jq}\sum_v\sum_n\left(\frac{f'_pf'_q}{f_\circ^2}\right)_v \\ &\qquad \times\left(\frac{1}{f_\circ}\right)_n\langle(\mathcal{N}_v-f_\circ)^2(\mathcal{N}_n-f_\circ)^2\rangle \\ &\quad = \sum_{ijpq}\mathscr{A}_{ij}(\mathscr{A}^{-1})_{ip}(\mathscr{A}^{-1})_{jq} \\ &\qquad \times\left[\left(\sum_v \frac{f'_pf'_q}{f_\circ^2}\langle(\mathcal{N}_v-f_\circ)^2\rangle\right)\left(\sum_n\frac{\langle(\mathcal{N}_n-f_\circ)^2\rangle}{f_\circ}\right)\right. \\ &\qquad \left. -\sum_n\frac{f'_pf'_q}{f_\circ^3}\langle(\mathcal{N}_n-f_\circ)^2\rangle^2+\sum_n\frac{f'_pf'_q}{f_\circ^3}\langle(\mathcal{N}_n-f_\circ)^4\rangle\right] \\ &\quad = \sum_{ijpq}\mathscr{A}_{ij}(\mathscr{A}^{-1})_{ip}(\mathscr{A}^{-1})_{jq}\times\left[\left(\sum_v\frac{f'_pf'_q}{f_\circ}\right)\left(\sum_{n=1}^{N_{\rm ch}}1\right)\right. \\ &\qquad \left. -\sum_n\frac{f'_pf'_q}{f_\circ}+\sum_n\frac{f'_pf'_q}{f_\circ^2}+3\sum_n\frac{f'_pf'_q}{f_\circ}\right] \\ &\quad \approx \sum_{ijpq}\mathscr{A}_{ij}(\mathscr{A}^{-1})_{ip}(\mathscr{A}^{-1})_{jq}\times(N_{\rm ch}\mathscr{A}_{pq}+2\mathscr{A}_{pq}) \\ &\quad = (N_{\rm ch}+2)L\end{aligned} \tag{149}$$

where in the second-to-last step we neglect the term $\frac{f'_pf'_q}{f_\circ^2}$ as being of relative order $1/f_\circ \ll 1$.

Now evaluate the last term in Eq. (145):

$$\begin{aligned}&\sum_{ijpq}\mathscr{A}_{ij}\mathscr{A}_{pq}\langle \Delta x_i\,\Delta x_j\,\Delta x_p\,\Delta x_q\rangle \\ &\quad = \sum_{ijpq}\mathscr{A}_{ij}\mathscr{A}_{pq}[\langle \Delta x_i\,\Delta x_j\rangle\langle \Delta x_p\,\Delta x_q\rangle + 2\langle \Delta x_i\,\Delta x_p\rangle\langle \Delta x_j\,\Delta x_q\rangle] \\ &\qquad -\sum_{ijpq}\sum_{abcd}\mathscr{A}_{ij}\mathscr{A}_{pq}(\mathscr{A}^{-1})_{ia}(\mathscr{A}^{-1})_{jb}(\mathscr{A}^{-1})_{pc}(\mathscr{A}^{-1})_{qd} \\ &\qquad \times\sum_n\frac{f'_af'_bf'_cf'_d}{f_\circ^4}[\langle(\mathcal{N}_n-f_\circ)^2\rangle^2\end{aligned}$$

$$
\begin{aligned}
&+ 2\langle(\mathscr{N}_n - f_\circ)^2\rangle^2 - \langle(\mathscr{N}_n - f_\circ)^4\rangle] \\
&= \sum_{ijpq} \mathscr{A}_{ij}\mathscr{A}_{pq}(\mathscr{A}^{-1})_{ij}(\mathscr{A}^{-1})_{pq} \\
&\quad + 2\sum_{ijpq} \mathscr{A}_{ij}\mathscr{A}_{pq}(\mathscr{A}^{-1})_{ip}(\mathscr{A}^{-1})_{jq} \\
&\quad + \sum_{abcd}(\mathscr{A}^{-1})_{ab}(\mathscr{A}^{-1})_{cd}\sum_n \frac{f'_a f'_b f'_c f'_d}{f_\circ^4}(-3f_\circ^2 + f_\circ + 3f_\circ^2) \\
&= L^2 + 2L + \sum_{abcd}(\mathscr{A}^{-1})_{ab}(\mathscr{A}^{-1})_{cd}\sum_n \frac{f'_a f'_b f'_c f'_d}{f_\circ^3} \\
&\approx L^2 + 2L. \qquad (150)
\end{aligned}
$$

In the last step we neglect the term $\frac{f'_a f'_b f'_c f'_d}{f_\circ^3}$, which is of order $1/f_\circ \ll 1$. Finally, for $\langle(\chi^2)^2\rangle$ in Eq. (145) we have

$$
\begin{aligned}
\langle(\chi^2)^2\rangle &= (N_{\mathrm{ch}}^2 + 2N_{\mathrm{ch}}) - 2(N_{\mathrm{ch}}L + 2L) + (L^2 + 2L) \\
&= N_{\mathrm{ch}}^2 - 2N_{\mathrm{ch}}L + L^2 + 2N_{\mathrm{ch}} - 2L. \qquad (151)
\end{aligned}
$$

References

[1] R.M. Carey, et al., Phys. Rev. Lett. 82 (1999) 1632.
[2] H.N. Brown, et al., Phys. Rev. D 62 (2000) 091101.
[3] H.N. Brown, et al., Phys. Rev. Lett. 86 (2001) 2227.
[4] G.W. Bennett, et al., Phys. Rev. Lett. 89 (2002) 101804.
[5] G.W. Bennett, et al., Phys. Rev. Lett. 92 (2004) 161802.
[6] G.W. Bennett, et al., Phys. Rev. D 73 (2006) 072003.
[7] D.E. Groom, et al., Eur. Phys. J. C 15 (2000) 1.
[8] W. Liu, et al., Phys. Rev. Lett. 82 (1999) 711.
[9] H.B. Dwight, Tables of Integrals and Other Mathematical Data, Macmillan Publishing Co., New York, 1961.
[10] S.I. Redin, Fit of experimental histograms: bias of fit parameters, preprint BINP 2002-73, Novosibirsk, Russia, 2002 (in Russian, with extended abstract in English).
[11] R.J. Barlow, Statistics: A Guide to the Use of Statistical Methods in the Physics Sciences, Wiley Publishing Company, New York, 1989 (Section 5.1.3).
[12] A.N. Kolmogorov, Foundations of the Theory of Probability, Chelsea Publishing Co., New York, 1950.
[13] T.W. Anderson, D.A. Darling, Ann. Math. Stat. 23 (1952) 192.
[14] N.H. Kuiper, Proc. K. Ned. Acad. Wet. A 63 (1962) 38.
[15] T. Hauschield, M. Jentschel, Nucl. Instr. and Meth. A 457 (2001) 384.
[16] S. Pomme, J. Keightley, Applied modeling and computations in nuclear science, in: T.M. Semkow, S. Pomme, S.M. Jerome, D.J. Strom (Eds.), ACS Symposium Series, vol. 945, American Chemical Society, Washington, DC, 2006, pp. 316–334, ISBN 0-8412-3982-7.
[17] S. Pomme, Applied modeling and computations in nuclear science, in: T.M. Semkow, S. Pomme, S.M. Jerome, D.J. Strom (Eds.), ACS Symposium Series, vol. 945, American Chemical Society, Washington, DC, 2006, pp. 282–292, ISBN 0-8412-3982-7.