Gravity Probe B data analysis: III. Estimation tools and analysis results

J W Conklin, M I Heifetz, T Holmes, M Al-Meshari, B W Parkinson, A S Silbergleit, C W F Everitt, A Al-Jaadan, G M Keiser, B Muhlfelder, V G Solomonik and H Al-Jabreen

1 MAE Department, University of Florida, Gainesville, FL 32611, USA
2 HEPL, Stanford University, Stanford, CA 94305-4085, USA
3 King Abdulaziz City for Science and Technology, Riyadh, Saudi Arabia

E-mail: francis1@stanford.edu.

Received 31 January 2015, revised 18 April 2015
Accepted for publication 3 June 2015
Published 17 November 2015

Abstract
This paper provides detailed descriptions of the numerical estimation algorithms used to fit physics-based models to the data from the Gravity Probe B spacecraft, as well as the scientific results of the experiment, and the statistical and systematic uncertainties. The first paper in this series of three data analysis papers derives the mathematical expressions for the signals to be analyzed, and the second paper deals with science data acquisition and their preparation for the relativistic drift rate estimation. The data from each of the four gyroscopes are partitioned into six segments, each spanning several weeks to several months. These segments are first analyzed individually to check the validity of the mathematical models and the accuracy of the estimation routine by examining the consistency of the relativistic drift rate estimates from each of these 24 gyro-segments. Then, the drift rate estimates and uncertainties are calculated for each individual gyroscope and for the four gyroscopes combined. These results are presented and compared with each other and with the prediction of general relativity.

Keywords: general relativity, GP-B, data analysis
1. Introduction

This is the final paper describing the Gravity Probe B (GP-B) data analysis. The GP-B space experiment was set to measure the drift rate of a free gyroscope orbiting the Earth. Its goal was to check general relativity (GR) predictions for the gyro precession found by Schiff in 1962 [1]. Its principle was very simple: put a star-tracking telescope and a high precision gyroscope on orbit, keep the telescope pointing at the guide star (GS) as an inertial reference, and measure the rate of change of the angle between the line of sight to the GS and the gyro spin (see figure 1 in [2] or [3]).

However, the implementation of this simple idea proved to be tremendously difficult: the effects to measure were extremely small, and the number of disturbing factors, which had to be reduced by many orders of magnitude to ensure accurate measurement, was huge. Let us just mention that the quality of the flight gyroscope had to be enhanced \(10^7\) times as compared to the existing ones, and the quality of the tracking telescope \(10^3\) times. It is only the combination of cryogenics with space, supported by the development of numerous special break-through technologies, that allowed us eventually to bring the experiment to a success. The whole experiment, including the science instrument and its development, the attitude and translation control (ATC) system of the spacecraft (S/C), and on-orbit operations, is described in [4–6].

Nevertheless, before the launch detailed plans for analysis of GP-B data were laid out, tested and refined (see [7] and [3], section 4.6). Several unpleasant on-orbit discoveries changed these plans drastically; the three major were the changing gyro polhode period, significantly larger than expected patch effect torques on the gyroscopes, and data segmentation due to S/C anomalies. These new circumstances required, on the one hand, a number of theoretical results to model the disturbances, and, on the other, the most careful data treatment combined with modern estimation techniques (signal filtering), to extract the relativity estimates from the available signals with satisfactory accuracy.

This paper aims at a thorough description of the GP-B data analysis, which was explained only very briefly in [2]. The first two relevant papers, [3] and [8], abbreviated DA I and DA II, contain theoretical derivations of all the signal models needed for the analysis (DA I), and describe preparatory treatment of data to analyze (DA II). Here we...
discuss estimation algorithms and their program realization; in particular, the main filtering tool called 2-second filter (2sF), or two-second estimator. We also present the obtained relativity estimates with their errors, and discuss the results.

Section 2 reflects on the analysis of high frequency (HF) SQUID signal, which by itself has nothing to do with the estimation of relativistic drift of GP-B gyroscopes, but provides vital information for it, particularly, the complete history of the gyro polhode motion and an a priori estimate of the time-varying component of the gyroscope readout scale factor.

The rest of the paper is concerned with the science analysis, that is, with the analysis of the low frequency (LF) SQUID signal. The main estimation tool and the accompanying programs for the representation and evaluation of the analysis results are described in section 3. In section 4 we explain how the actual analysis was carried out by means of parallel computation using parallel processor clusters at Stanford. Section 5 deals with the estimation results obtained from the analysis of each of the four gyro signals, as well as with combining them into final four-gyro result of the experiment. The treatment of systematic errors is also discussed there, and the total experimental error, corrected for over/under-dispersion, is given. The way to calculate this correction of the joint four-gyro estimate error is described in the appendix. Conclusions and discussion are found in section 6.

Let us recall some convenient terms and notations introduced in DA II. There are several steps in GP-B data handling on the ground prior to the main data analysis. The data after each such step forms, in our terms, the corresponding level numbered successively, from Level 0 raw telemetry data to Level 3 data ready for the relativistic drift estimation analysis. For brevity, we usually write LN, (N = 0, 1, 2, 3), instead of Level N; Level N data are kept in the Level N database denoted LN DB.

2. Trapped flux mapping (TFM): polhode phase and angle, and gyroscope scale factor

HF SQUID data were used to estimate the history of GP-B rotors orientation with respect to the pick-up loop, and to produce an estimate of the SQUID scale factor variations caused by trapped flux (TF).

The estimation process was broken up into three parts called levels (see figure 1) allowing us to estimate groups of parameters somewhat independently. Still, as seen in figure 1, Levels A–C are nested structures rather than sequential steps. In brief, after finding the polhode phase and angle, the rotor asymmetry parameter (details below) and the rough TF distribution at Level A, we determine the spin phase at Level B, and then, more accurately, the TF magnetic potential at Level C. The last result, in its turn, allowed us to return to Level B for a better estimate of gyro precession, and so on (see more in sections 2.4 and 2.5).

The estimation procedures at each level are rather well described in [9]. We thus frequently refer to it, concentrating on some significant details missing there, and on a better description and interpretation of the TFM results.

All fits have been done to the HF SQUID signal data in the form of complex harmonics, \( \tilde{H}_n(t) \), of spin frequency that were available at \( \sim 10^5 \) epochs spanning roughly 1 year. We described the data preparation, i.e., harmonics extraction from HF SQUID snapshots, in section 5.4 of paper [8]. The maximum number of harmonics extracted was 51; however, most of the higher harmonics were not used in the fitting process. Moreover, only odd harmonics participated in TFM, because the even ones do not enter the expression for the TF, derived in the appendix A of [3].
2.1. Level A: polhode phase \( \phi_p \)

Level A processing provided estimates for the symmetric polhode phase, \( \phi_p(t, Q = 0) \equiv \phi_{0i}(t) \), and the gyroscope asymmetry parameter, \( Q \). The latter is defined as a ratio of difference between the middle and the smallest moments of inertia to the largest one, see equation (16) of [3]. In section 4 of appendix A of [3] we defined \( \phi_{0i}(t) \) as a polhode phase of a symmetric gyroscope whose polhode period dependence on time, or, equivalently, the history of energy dissipation, is the same as the one for the given asymmetric rotor, in our case, one of the four GP-B gyroscopes. From \( \phi_{0i}(t) \) we then computed the true polhode phase, \( \phi_p(t) \), and angle, \( \gamma_p(t) \), using Euler’s exact solution to his equations of motion for a rotor with the known value of \( Q \).

On the other hand, by integrating the varying polhode rate \( \Omega_p(t) \) we obtained the following model for the symmetric polhode phase ([9], equation (11)):

\[
\phi_p(t) = \phi_{0i} + \Omega_{pa}(t - t_{ref}) - \sum_{m=1}^{M} D_m \frac{\tau_{dis}}{m} e^{-m(t-t_{ref})/\tau_{dis}}.
\]

Here \( \phi_{0i} \) is the symmetric polhode phase at the reference time, \( t_{ref} \), \( D_m \) are constant coefficients, and \( \Omega_{pa} = 2\pi/T_{pa} \) is the asymptotic polhode rate that the polhode frequency approaches as the polhode motion damps out, and the polhode period becomes constant (denoted \( T_{pa} \)). Finally, \( \tau_{dis} \) is the characteristic time of polhode damping determined for each GP-B rotor prior to TFM ([9], table 2; see also [10]); its range was between 1 and 2.5 months depending on the gyroscope.

So, \( 2 + M \) parameters were estimated for the symmetric polhode phase model: \( \Omega_{pa} \), \( \phi_{0i} \), \( D_1 \), \( D_2 \), …, \( D_M \); in our fits \( M \) was never larger than 7. In practice it was more convenient to replace \( \phi_{0i} \) with an equivalent parameter \( l_0 \), the time at which the phase vanished, \( \phi_p(t_0) = 0 \).

The theoretical model for the harmonics \( \tilde{H}_s(t) \) was derived in appendix A of [3]; it is given by equations (A.18), (A.19). It expresses the harmonics through the three Euler angles \( \phi_p(t) \), \( \gamma_p(t) \), and \( \phi(t) \), and coefficients \( A_{lm} \) coming from the spherical harmonics (or multipole) expansion of the scalar magnetic TF potential. The first two Euler angles specify the instant direction of the gyro spin axis in the rotor body (see equation (17) in [3]), and spin phase \( \phi(t) \) describes rotation about it.

In theory, the set of coefficients \( A_{lm} \) is infinite; in practice, we had to choose the maximum number, \( l_{max} \) of the harmonics in our fits. So \( l_{max} \) is the harmonics truncation parameter; it turned out to be one of the key parameters in the overall TFM analysis. Particularly, the maximum number, \( n_{max} \) of odd harmonics \( \tilde{H}_s \) used in the data fits was always taken to be the same as \( l_{max} \) to ensure observability of the coefficients \( A_{lm} \). Indeed, the complex amplitude \( H_s(t) \) of the harmonics \( \tilde{H}_s(t) \) contains all \( A_{lm} \) for which \( l \geq n \). Thus we needed \( l \geq n \); however, taking \( l > n \) resulted in an ill-conditioned fit (see section 5.3.2); so we always set \( n_{max} = l_{max} \).

Truncating the number of harmonics produces a systematic error in the estimation. To assess the significance of this error source each fit was carried out several times, every time with a different \( l_{max} \). This was always done for the nonlinear estimation of the initial polhode phase \( \phi_{0i} \). Typically, for Level A fits we took several values of \( l_{max} \) from the range \( 3 \leq l_{max} \leq 9 \).

We used more harmonics (larger \( l_{max} \)) earlier in the mission, when larger polhode provided better \( t_0 \) observability. We also used more harmonics for gyroscopes 1, 2 than for 3, 4, because the (random) initial conditions for the latter turned out closer to the minimum energy
equilibrium state in which spin axis is permanently aligned with the maximum inertia axis [9]. As a result, the polhode paths around this axis were tighter for rotors 3 and 4, and their total polhode period changes from the initial to the asymptotic value were smaller. The actual value of \( t_0 \) in the fits was found as the average of individual estimates with different \( t_{\text{max}} \) values.

2.1.1. Use of prior information \((t_0, \phi_{\text{si}}, \omega_{\text{si}}, d\omega_t)\). Simultaneous estimation of \( \phi_\text{si}(t), \gamma_\text{si}(t), \phi_\text{fi}(t), \) and \( A_{\text{lm}} \) would require a rather precarious and computationally intensive, if not overwhelming, nonlinear fit. To avoid it at Level A, we took advantage of existing prior information. Namely, we used (a) initial estimates of the polhode rate, \( \Omega_{p}(t) \), which were determined during the science phase of the mission (see [10] for details), and (b) the initial estimates of the spin rate, \( \omega_{\text{si}}(t) \), which came out of nonlinear fits that had produced the harmonics \( \tilde{H}_n \) (see [8], section 5.4).

To do this, we first of all introduced data batching. The HF SQUID snapshots typically occurred in sequences of \( \sim 1000 \) with \( \sim 40 \) s gaps between the snapshots. Each such stretch spanned 12–24 h, with roughly 1–2 d in between. So the \( \tilde{H}_n \) data were conveniently grouped into \( \sim 125 \) of these batches over about one year of science data collection; the fits were carried out one batch at a time.

The estimates of the polhode rate \( \Omega_{p}(t) \) obtained before the TFM analysis were accurate to \( \sim 100 \) nHz [10]. So we could integrate this rate over of a single batch \((\sim 1 \) d\) accruing a maximum polhode phase error of \( 2\pi \times (86400 \) s) \times (100 \) nHz) \( \sim 0.05 \) rad. This accuracy was sufficient for the Level A fitting process. Recall that the constant of integration in the expression (1), i.e., the initial polhode phase \( \phi_{\text{fi}}(0) \), was unknown. So we had to include it in the Level A set of fit parameters (as mentioned, it was actually replaced with the equivalent parameter \( t_0 \)).

The next thing we needed for fitting to the measured harmonics was the spin phase, \( \phi_\text{fi}(t) \). The existing estimates of its rate, \( \omega_{\text{si}}(t) \), however, were accurate only to \( 100 \mu \)Hz [10]. Therefore integrating this rate could not produce an accurate enough result for the spin phase. Indeed, over one day period, a typical data length in a single fit, the error in the spin phase due to integration could be up to \( 2\pi \times (86400 \) s) \times (100 \) \muHz\) \( \sim 50 \) rad. And again, the constant of integration, \( \phi_{\text{si}}(0) \), was not known. Thus, in addition to \( t_0 \) at Level A we had to also simultaneously estimate three spin phase parameters for each batch of data: the initial spin phase \( \phi_{\text{si}}(0) \), the initial spin rate \( \omega_{\text{si}}(0) \), and the spin-down rate \( d\omega_t \). The estimation process used for this was identical to the Level B estimation discussed in section 2.2.

Let us assume for a moment that the asymmetry parameter \( Q \) is known (the next section explains how it was actually determined). Using also the estimated values of \( t_0, \phi_{\text{si}}, \omega_{\text{si}}, \) and \( d\omega_t \) we can then compute the three Euler angles \( \phi_\text{fi}(t), \gamma_\text{fi}(t), \) and \( \phi_\text{fi}(t) \) over every single batch of data. With all this information a least-squares (LSQ) fit for the coefficients \( A_{\text{lm}} \) becomes possible. Essentially, it is the same Level C linear fit, but for a single data batch; so we discuss its details in section 2.3.

With the estimates of \( A_{\text{lm}} \) and all other states from Level A we reconstructed harmonics amplitudes \( H_n(t) \). We then computed the cost function \( c \), the ratio of the post-fit residuals rms to the rms of the measured harmonics (section 2.3.1, equation (9)). The cost function value provided us with a quick assessment of the overall fit quality: for instance, if \( c = 0.1 \), then the fit was roughly a 10% fit.

The nonlinear search for \( t_0 \) exploited a standard Nelder–Mead simplex method (see [10] for details) implemented in the MATLAB function \texttt{fminsearch.m}. In this and all other nonlinear estimations we used the same cost function \( c \).
2.1.2. Asymmetry parameter $Q$. We needed $Q$ to find the true (‘asymmetric’) Euler angles, after the symmetric polhode, $\phi_0(t)$, and spin, $\phi_f(t)$, phases were known (basing on the found values of $t_0$, $\Omega(t)$ for the first of them, and on $\omega_{vi}$, $d\omega_i$, $d\omega_f$ for the second).

Estimating $Q$ proved to be the most computationally intensive part of TFM, because it was sensitive to all fit parameters, even though $Q$ does not depend on them from a physical standpoint. So all the parameters had to be estimated simultaneously with $Q$; as a result, only three separate batches of data were used to determine $Q$ for each gyroscope.

For this, we chose relatively long batches (>1500 data points) early in the mission, when the polhode variations were large. Instead of an additional nested nonlinear search for $Q$, the Level A code was run many times on a single batch with incrementing $Q$ from 0 to 1 in 0.02 steps. For each $Q$, several values of $l_{\text{max}}$ were used. After all the runs we chose the value of $Q$ that corresponded to the minimum of the cost function (out of $\sim$12 estimates, one for each batch fit with a particular value of $l_{\text{max}}$). The best-fit $Q$ was taken then as the average of these estimates, and the error in it was defined as their standard deviation.

The Level A results for $Q$ are given in Table 1 for all four gyroscopes (first published as Table 3 of our paper [9]). The estimation accuracy of $Q$, about 20%, might appear not very high. However, it turned out sufficient for the needed $\sim$1% estimation of LF SQUID scale factor variations caused by TF, due to a low sensitivity of the signal to the rotor asymmetry parameter.

2.1.3. Polhode phase smoothing. A key Level A result was 50–100 estimates of the times $t_0$ when the polhode phase was zero; the estimates spanned much of the whole science phase of the GP-B mission. The actual number of $t_0$ estimates depended on the gyroscope, with 102 estimates for gyroscope 2 and 50–60 estimates for other gyroscopes: a larger polhode means greater observability of $t_0$. Each estimate of $t_0$ represents one batch of data used at Levels B and C. Table 2 shows the exact number of $t_0$ estimates and the period they belong to.

Each $t_0$ had some estimation error inducing the error in the polhode phase. To improve the polhode phase accuracy, we fitted all its estimates to one single model. This ‘smoothed’ the estimates reducing the overall error. It also produced a polhode phase history continuous and self-consistent through the entire mission, that is, available at any moment of time, not

| Gyroscope | $Q$ | Dates used (2004) | Max. harmonic fitted |
|-----------|-----|------------------|----------------------|
| 1         | 0.303 ± 0.069 | 4, 20 October, 19 November | 1, 3, 5, 7 |
| 2         | 0.143 ± 0.029 | 4, 20 October, 19 November | 1, 3, 5, 7 |
| 3         | 0.127 ± 0.072 | 8, 13, 20 October | 1, 3, 5 |
| 4         | 0.190 ± 0.048 | 4, 13, 20 October | 3, 5, 7 |

| Gyroscope | rms of residuals | Dates used | # of batches | $M$ |
|-----------|------------------|------------|--------------|-----|
| 1         | 0.043 rad (2.5°) | October 2004–February 2005 | 58 | 7 |
| 2         | 0.058 rad (3.3°) | September 2004–August 2005 | 102 | 7 |
| 3         | 0.056 rad (3.2°) | September 2004–January 2005 | 55 | 4 |
| 4         | 0.092 rad (5.3°) | September 2004–January 2005 | 53 | 4 |

Class. Quantum Grav. 32 (2015) 224020 J W Conklin et al
just at the snapshot times. Importantly, this history was free of $\pi$ ambiguities, as we explain below.

To model the symmetric polhode phase $\phi_0$ we used its expression (1). The unknown parameters $\phi_0$, $\Omega_0$, and $D_0$ all appear linearly in this model requiring a simple LSQ fit. The polhode phase estimates that we fit to were constructed from the estimates of the initial conditions for each batch of data with some specific $\phi_0$ by means of integrating the estimated polhode rate [10].

In this way we built an estimate of the symmetric polhode phase at the time of every snapshot, since each $\phi_0$ was found from the $\dot{H}_n(t)$ data existing at snapshot times. To get a self-consistent $\phi_0(t)$ we used the estimated polhode rate, accurate to $\sim 600$ nrad s$^{-1}$, for resolving the $\pi$ ambiguities between batches (typically not more that 4 d apart). We had to resolve precisely $\pi$ ambiguities, or half-cycle slips, and not $2\pi$ ambiguities, because the period of polhode modulation of $\phi_0(t)$ was a half of the polhode period. Therefore, we could not recognize the difference between $\phi_p = 0$ and $\phi_p = \pi$; luckily, we did not care, since we were

Figure 2. (a) Level A batch-by-batch polhode phase and LSQ fit of the exponential model for gyroscope 2. (b) Post-fit residuals and standard errors in Level A parameters $\phi_0$. 

Class. Quantum Grav. 32 (2015) 224020 J W Conklin et al
Figure 2 illustrates the fit of the exponential polhode phase model (1) to the estimated polhode phase for gyroscope 2. In figure 2(a) the estimated polhode phase for gyroscope 2 is plotted on top of it is the best-fit result of the exponential polhode phase model. In figure 2(b) we plot the residuals of the fit, along with the standard deviation of the estimated $t_0$ transformed into units of radians using $\Omega_p (t)$ as a conversion factor. We see that the errors in $t_0$ are typically larger than the residuals, indicating that (a) the exponential model describes the data accurately, and (b) the systematic error due to truncating the number of harmonics to $l_{\text{max}}$ may be overestimated. The polhode phase smoothing fits for other gyroscopes are qualitatively similar, except that the length of data used was roughly half as long as the stretch for gyroscope 2.

Table 2 demonstrates the quality of smoothing fits for all gyroscopes. rms of the residuals were all within 0.05 – 0.1 rad. The fit stretch was 5–12 months depending on the gyroscope. During this period, the polhode phase has cycled through $\sim 10^4$ rad, so one can think of the polhode phase smoothing fits as being accurate to 1 part in $10^5$. Also reported in table 2 is the number, $M$, of coefficients $D_m$ in model (1) used for each gyroscope; the corresponding total number of parameters was $2 + M$. The results of the polhode phase smoothing were later used to refine our estimates of the spin phase, as described in the next section. The estimated polhode frequency and polhode angle resulting from Level A, for all four gyroscopes, are shown in figures 3 and 4.
2.2. Level B: spin phase ($\phi_s$)

The spin phase determination at Level B is described in section 5.2 of paper [9]. The detailed account of its core processing, i.e., precise estimation of the spin speed and spin-down rate, is found in [10], so here we touch upon it only briefly.

Using the best (smoothed) polhode phase from Level A for each gyroscope, we had to estimate three parameters, $\phi_s$, $\omega$, and $d\omega$, defined in the previous section, for each of the ~125 data batches. The Level B processing code was the same as the one nested inside the Level A code; it processed a single batch at a time.

The nonlinear search routine for the three spin phase parameters was a modified version of the Nelder–Mead simplex method. The initial frequency, $\omega$, and the linear decay rate, $d\omega$, were estimated separately from the initial spin phase, $\phi_s$, to simplify the nonlinear estimation process. The code iterated between the first and the second searches; a total of four iterations were shown to be a good compromise between accuracy and computation time.

The search for $(\omega, d\omega)$ used a version of the MATLAB function `fminsearch.m` modified by Kozaczuk [11]. It optimized the simplex method when only two parameters of strongly different magnitudes were involved: indeed, the initial frequency, $\omega$, was on the order of 60–80 Hz, while the decay rate, $d\omega$, was just $\sim 10^{-10}$ Hz s$^{-1}$. The nonlinear search for $\phi_s$ used the standard Nelder–Mead method, but with the bounds $-\pi \leq \phi_s < \pi$.

We used two assessments of the estimation errors. One was the difference between the estimates of each of the three parameters determined with different $l_{\max}$, the other was the magnitude of a discontinuity in the spin phase at the boundaries between data batches. Both
Figure 5. Total angular velocity, $\omega_c$, of gyroscopes from Level B estimation.

Figure 6. Gyroscope spin-down rates, $d\omega/dt$, from Level B estimation, with standard deviations.
approaches not only account for formal (statistical) errors, but also for the level of systematic inaccuracies, to a certain extent. In general, the errors in $\omega_{\parallel}$, $\omega_{\perp}$, and $\psi_{\parallel}$ were $\sim 1$ nHz, $\sim 1$ pHz s$^{-1}$ and $\sim 30$ mrad respectively, and the total error in $\phi(t)$ was $\sim 50$ mrad. Spin angular velocities and spin-down rates resulting from Level B, for all four gyroscopes, are shown in figures 5 and 6.

2.3. Level C: potential of magnetic TF ($A_{lm}$)

Levels A and B results led us to computing the histories of the three Euler angles $\phi_p(t)$, $\gamma_p(t)$, and $\psi_p(t)$ for the entire science phase of the GP-B flight. Using them, we estimated coefficients $A_{lm}$ in the spherical harmonics expansion of the TF. The complex harmonics amplitudes $H_n(t)$ depend linearly on these coefficients, so we used a linear LSQ fit; here we discuss some significant details of this estimation.

2.3.1. $A_{lm}$ symmetry and linear LSQ model in the matrix form. When fitting for the coefficients $A_{lm}$, we had to maintain the condition $A_{l-m} = (-1)^m A_{lm}^{*}$, that keeps the magnetic potential real ([3], appendix A, equation A5). This condition has also the advantage of reducing the total number of parameters to estimate. We introduced the real and imaginary parts of $A_{lm}$,

$$A_{lm} = a_{lm} + i b_{lm}; \quad a_{l,-m} = (-1)^m a_{lm}, \quad b_{l,-m} = (-1)^{m+1} b_{lm}. \quad (2)$$

and estimated them, instead of the complex number $A_{lm}$; $A_{00}$ is real, so $b_{00} = 0$.

We substituted expression (2) into the general model of the harmonics amplitude $H_n$ given by equation (A.19) of appendix A in [3]. For $n$ odd, summing in $l$ only up to its truncation value $l_{max}$ we obtained, after elementary transformations:

$$H_n = \frac{\Phi_0}{2} \sum_{l=0}^{l_{max}} \left( \frac{f_0}{b} \right)^l \left\{ a_{l0} d_{l0}^l(\gamma_p) \right. \left[ \cos mf_p + i \sin mf_p \right] \right.$$

$$\left. + (-1)^m d_{l-m}^l(\gamma_p) \left[ \cos mf_p - i \sin mf_p \right] \right)$$

$$\left. + i b_{l0} \left( d_{l0}^l(\gamma_p) \left[ \cos mf_p + i \sin mf_p \right] \right) \right.$$

$$\left. - (-1)^m d_{l-m}^l(\gamma_p) \left[ \cos mf_p - i \sin mf_p \right] \right\},$$

where $\Phi_0$ is the magnetic flux quantum, $b$ is the pick-up loop radius

$$I_l = \frac{(-1)^l(2j - 1) !!}{2^{l+2}(j + 1)!(\sqrt{2})^{l+2}}, \quad l = 2j + 1,$$

and $d_{lm}(\alpha)$ is the spherical harmonics rotation matrix defined in [12] (see formulas (4.13), (4.14) there). The real and imaginary parts of this expression for $H_n$ are:
\[
\text{Re } H_n = \frac{\Phi_0}{2} \sum_{l \text{ odd}}^{l_{\text{max}}} \left( \frac{r_0}{b} \right)^l d_{\text{in}}(\pi/2) I_l \left\{ a_{l0} d_{\text{in}}'(\tau_p) + \sum_{m=1}^{l_{\text{max}}} \left[ \begin{array}{c} a_{lm} \left( d_{\text{in}}' \left( \tau_p \right) + \left( -1 \right)^m d_{-m}^l \left( \tau_p \right) \right) \cos m\phi_p \\ b_{lm} \left( d_{\text{in}}' \left( \tau_p \right) + \left( -1 \right)^m d_{-m}^l \left( \tau_p \right) \right) \sin m\phi_p \end{array} \right] \right\}, \\
\text{Im } H_n = \frac{\Phi_0}{2} \sum_{l \text{ odd}}^{l_{\text{max}}} \left( \frac{r_0}{b} \right)^l d_{\text{in}}(\pi/2) I_l \left\{ a_{l0} d_{\text{in}}'(\tau_p) + \sum_{m=1}^{l_{\text{max}}} \left[ \begin{array}{c} a_{lm} \left( d_{\text{in}}' \left( \tau_p \right) - \left( -1 \right)^m d_{-m}^l \left( \tau_p \right) \right) \sin m\phi_p \\ b_{lm} \left( d_{\text{in}}' \left( \tau_p \right) - \left( -1 \right)^m d_{-m}^l \left( \tau_p \right) \right) \cos m\phi_p \end{array} \right] \right\}. 
\]

(3)

(4)

Of course, these expressions are linear in \( a_{lm} \) and \( b_{lm} \), since \( H_n \) is linear in \( A_{lm} \).

We recast this model in the usual linear matrix form:

\[
y = A x + w. 
\]

(5)

Here \( y \) represents the measurement, \( x \) are the unknown parameters (states) to be determined, \( A \) is the linear operator, a matrix in the finite-dimensional case, that relates \( x \) and \( y \), and \( w \) is the sensor noise corrupting the measurement \( y \). All terms in equation (5) depend on time except for the state vector \( x \).

We form the measurement matrix, \( y \), by a column of real and imaginary parts:

\[
y = \begin{bmatrix} \text{Re } H_1(t), \text{Im } H_1(t), \text{Re } H_2(t), \text{Im } H_2(t), \ldots, \text{Re } H_{l_{\text{max}}}(t), \text{Im } H_{l_{\text{max}}}(t) \end{bmatrix}^T. 
\]

(6)

With the measurements held at times \( t_1, t_2, \ldots, t_N \), every real and imaginary part of \( H_n(t) \) appearing in equation (6) is itself a vector of the length \( N \). So the measurement matrix \( y \) has the dimensions \( N (l_{\text{max}} + 1) \times 1 \).

The column state vector \( x \) is comprised of the real and imaginary parts of \( A_{lm} \):

\[
x = \begin{bmatrix} a_{10}, a_{11}, b_{11}, a_{30}, a_{31}, b_{31}, a_{32}, b_{32}, \ldots, a_{l_{\text{max}},l_{\text{max}}}, b_{l_{\text{max}},l_{\text{max}}} \end{bmatrix}^T, \\
\quad l = \text{odd};
\]

(7)

the length of this vector is

\[
\sum_{l=0}^{l_{\text{max}}} (2l + 1) = 1 + \frac{3l_{\text{max}}}{2} + \frac{l_{\text{max}}^2}{2}.
\]

(8)

Matrix \( A \) mapping linearly the states \( x \) to the measurement \( y \) is real-valued and has the dimension \( N (l_{\text{max}} + 1) \times 1 \). It is defined by equations (3) and (4), and is block upper-diagonal, as \( H_n \) contains \( A_{lm} \) with \( l \geq n \) only. The matrix \( A \) in its entirety is found in section 4.1.6 of [13], in slightly different notations.

The measurement noise, \( w \), is again of the dimension \( N (l_{\text{max}} + 1) \times 1 \). We assume that \( w \) is a random vector with the zero mean and covariance \( \Sigma_w \); we also assume \( \Sigma_w = \sigma_w^2 I \).
where \( I \) is the identity matrix. This means that the measurements were uncorrelated, and they all had the same covariance \( \sigma_w^2 \).

Let \( \hat{x} \) be the estimated (best-fit) state vector, and \( \mathbf{r} = \mathbf{y} - \mathbf{A} \hat{x} \) be the vector of post-fit residuals. We choose a cost function \( c \) as the two-norm (square root of the sum of squares) of the residuals normalized by the two-norm of the measurement,

\[
c = \frac{\| \mathbf{y} - \mathbf{A} \hat{x} \|_2}{\| \mathbf{y} \|_2}.
\]

The solution \( \hat{x} \) to the standard LSQ problem is

\[
\hat{x} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{y} ;
\]

it also minimizes the cost function \( c \), since \( 1/\| \mathbf{y} \|_2 \) is estimate-independent.

For a general noise covariance matrix \( \Sigma_w \) we should solve the weighted LSQ problem with the weighting matrix \( \mathbf{W} = \Sigma_w^{-1} \). However, in our case \( \Sigma_w = \sigma_w^2 \mathbf{I} \) the weighted and standard non-weighted LSQ problems are identical. As soon as the estimate \( \hat{x} \) is obtained, it is straightforward to reconstruct the coefficients \( A_{lm} \).

2.3.2. A priori distribution of coefficients \( A_{lm} \). For relatively low values of the harmonics truncation parameter, \( l_{\text{max}} \) say \( l_{\text{max}} \lesssim 9 \), the linear model \( \mathbf{A} \) was well-conditioned, and the unweighted LSQ fit provided believable results for \( A_{lm} \). However, for \( l_{\text{max}} \gtrsim 9 \) matrix \( \mathbf{A} \) became ill-conditioned, resulting in the estimates of \( A_{lm} \) that were hardly realistic. Indeed, we had reason to expect that all parameters \( \Phi_0 A_{lm} \) were on the order of 1 volt, but, as \( l_{\text{max}} \) was increased and \( \mathbf{A} \) became ill-conditioned, some of the \( A_{lm} \) estimates grew many orders of magnitude above this level.

This happened because the contributions from certain subsets of the coefficients \( A_{lm} \) are quite similar; for instance, \( A_{15,3} \) and \( A_{13,3} \) both contribute to \( H_{13} \) at the 3rd harmonics of the polhode frequency. In addition, the highest harmonic fitted, \( H_{l_{\text{max}}} \), depends only on \( A_{l_{\text{max}}} \). For large \( l_{\text{max}} \), a relatively large number of parameters \( (2l_{\text{max}} + 1) \) were fitted to a relatively small signal, since \( H_n \) decreased with \( n \) increasing. For example, \( A_{15,3} \) appeared ‘for the first time’ only in the 15th harmonics; \( H_{15} \) was more than ten times smaller than \( H_1 \), and it was fit with 31 parameters, including \( A_{15,3} \). Naturally, the latter was not constrained strongly.

A common practice in such cases is to incorporate additional information about the fit parameters, improving thus their estimation. Since the distribution of fluxons and antifluxons on the surface of the rotors was practically random, it was natural to assume that the coefficients \( A_{lm} \) followed a random distribution with zero mean. In other words, we expected that the TF was not biased in any particular direction on the rotor surface with respect to the body-fixed frame.

To test this hypothesis we examined the \( A_{lm} \) estimates for gyroscope 2 making no prior assumptions about their distribution. The fits to the gyroscope 2 data were relatively well behaved, even for large \( l_{\text{max}} \), because its polhode motion remained active over the entire science period.

Figure 7 shows a histogram of real and imaginary parts of \( A_{lm} = a_{lm} + i b_{lm} \) for gyroscope 2 with \( l_{\text{max}} = 17 \). Clearly, \( a_{lm} \) and \( b_{lm} \) follow some ‘bell-shape’ distribution centered at zero. Also plotted is a Gaussian distribution with zero mean and a standard deviation of 0.87 V, which is the actual standard deviation of coefficients \( \Phi_0 A_{lm} \) for gyroscope 2. Thus it was natural to choose a Gaussian distribution, because it is qualitatively similar to the measured one, and because of its wide use in estimation problems. One can also refer to the central limit theorem of probability theory making the Gaussian to be the best candidate for a distribution.
So we assumed that $a_{lm}$ and $b_{lm}$ were normally distributed, $a_{lm}, b_{lm} \sim \mathcal{N}(0, \sigma_i)$. We also made a quite natural assumption that the coefficients $A_{lm}$ were uncorrelated, hence the covariance matrix $\Sigma$ of the prior distribution of $x$ was diagonal, $\Sigma = \sigma_r^2 I$.

While the a priori distribution taken for $A_{lm}$ was well-founded, we wanted to ensure that the prior information was not weighted too strongly, i.e., did not overly constrain the $A_{lm}$ estimates. If $\sigma_r$ is chosen to be very small, then the a priori information is highly influential in the fit, and all of the $A_{lm}$ estimates tend to zero, resulting in a relatively large cost function $c$. On the other hand, if $\sigma_r$ is very large, then the distribution has but a little influence, and the ill-conditioned model remains an issue. Therefore, we chose $\sigma_r$ so that the problem became well-conditioned, and at the same time the post-fit residuals were not greatly affected. This was achieved by taking $\sigma_r$ as, roughly, the smallest value that did not increase the standard deviation of the post-fit residuals by more than 1%.

2.4. Polhode phase refinement. An iterative approach

The most fundamental parameter determined by TFM was the polhode phase $\phi_p(t)$. Its Level A estimates were typically accurate to 50–100 mrad during the stretches used in the fits; for gyroscopes 1, 3 and 4 it was just about a half of the mission. For the remainder of the periods the needed polhode phase was supplied by integrating the estimated polhode rate, usually with a larger error to be reduced.

A useful technique for determining the error in the polhode phase involves the absolute value, $|H_1(t)|$, of the first harmonics of the fundamental frequency $o_s \pm o_r - \Omega_n$. Using $|H_1(t)|$ had two advantages: (a) $H_1$ was the largest of all HF signals; (b) even more importantly, $|H_n(t)|$ was insensitive to errors in the spin phase (see [8], section 4). In fact, $|H_1(t)|$ only depended on the polhode phase and angle, and on $A_{lm}$; as the polhode angle was simply computed from other parameters, $|H_1(t)|$ really depended on $\phi_p(t)$ and $A_{lm}$ only.

Upon completing Levels A–C, we had estimates of all HF SQUID signal model parameters. We therefore were able to reconstruct the best-fit $|H_1(t)|$ and compare the result with its measured value. Figure 8 shows the measured $|H_1(t)|$ and its reconstruction for gyroscope 1 at two different times. In plot (a) there is a phase lag between the measured and
reconstructed $|H_1|$, while in plot (b) the two curves are in phase. From both plots it is also clear that the two curves agree quite well. The profile of $|H_1(t)|$ depends on $A_{lm}$, while the overall phase depends on $\phi_p(t)$. So we could say with some confidence that $A_{lm}$ were estimated accurately, and the dominant error source in the $H_n(t)$ fits was the polhode phase.

Using the measured and reconstructed $|H_1(t)|$ we found the polhode phase error. Since, by the expressions (3) and (4), $|H_1(t)|$ only contains the polhode harmonics, we fit a linear combination of those

$$|H_1(t)| = \sum_{m=0}^{l_{\text{max}}} K_m \cos m\phi_p(t) + L_m \sin m\phi_p(t)$$

(11)

to both the measured and reconstructed $|H_1(t)|$. For every data batch, we determined coefficients $L_m$ and $K_m$ for both the measured and computed $|H_1(t)|$. The error in the polhode phase during a batch, $\delta\phi_p$, was then calculated as a difference in the phases of the first polhode harmonics resulting from the two fits:

---

Figure 8. Measured and reconstructed polhode phase for gyroscope 1 in (a) September 2004, showing a phase lag, and in (b) October 2004, the two profiles are in-phase.
We then updated the exponential polhode phase model (1) by adding $\delta \phi_p$, and did the Level A re-fit. Since the polhode phase refinement was based on a phase difference, it was actually a correction to the polhode frequency, $\Omega_p(t)$; the initial polhode phase parameter $\phi_{pl}$ was not refined properly. So the refinement process could introduce a small systematic error that biased the updated value of $\phi_{pl}$. To ensure this did not occur, we then re-estimated $\phi_{pl}$ using the Level A determination of the times $t_0$. This did not alter the polhode frequency refinement, but guaranteed the absence of systematic error in the form of a shift in the absolute polhode phase.

At this point it was natural to assess the impact of the newly refined polhode phase on the spin phase $\psi(t)$ and coefficients $A_{lm}$. Hence we returned to the Level B processing and re-determined the spin phase using the refined polhode phase; after this we performed the Level C processing to determine again coefficients $A_{lm}$. Polhode phase refinement and the subsequent Levels B and C processing constituted a single iteration of an estimation process that should converge to the optimal solution. We continued these iterations (see figure 1) until the cost function (9) reached its minimum.

Figure 10 shows the cost function $c$ versus the maximum number, $l_{\text{max}}$, of harmonics used in the Level C fit; two final iterations for gyroscope 1 are presented. The figure demonstrates that both increasing $l_{\text{max}}$ and performing successive iterations reduce the cost function, i.e., the post-fit residuals. Recall that increasing $l_{\text{max}}$ increased also $n_{\text{max}} = l_{\text{max}}$ meaning larger amount of data used for fitting. Thus $c$ does not have to decrease when $l_{\text{max}}$ grows. The fact that it actually went down with the growth of $l_{\text{max}}$ indicates that our three Euler angles were accurate.

We wanted not only the cost function to converge to a minimum, but also the estimated parameters converging to their best-fit values. We checked the polhode phase convergence by examining its correction (12). In figure 9 we plotted the polhode phase correction for
It is clear that the error in the polhode phase was decreasing after each iteration.

Figures 10 and 9 give just two indicators of convergence. In fact, the cost function, the polhode phases and the coefficients \( A_{lm} \) all converged to finite values for each of the gyroscopes, demonstrating that our processing technique was robust and provided reliable parameter estimates.

### Table 3. Accuracy achieved in estimation of GP-B rotor dynamics parameters.

| Parameter                  | Absolute error | Relative error |
|----------------------------|----------------|----------------|
| Angular velocity, \( \omega \) | \(-10 \text{ nHz}\) | \(-10^{-10}\) |
| Spin-down rate, \( d\omega/dt \) | \(-1 \text{ pHz s}^{-1}\) | \(-10^{-2}\) |
| Polhode rate, \( \Omega_p \)       | \(-10 \text{ nHz}\) | \(-10^{-4}\) |
| Polhode phase, \( \phi_p \)       | \(-50 \text{ mrad}\) | \(-10^{-6}\) |
| Polhode angle, \( \gamma_p \) | \(-0.1-10 \text{ mrad}\) | \(-10^{-4}-10^{-1}\) |
| Spin phase, \( \phi_s \)       | \(-50 \text{ mrad}\) | \(-10^{-9}\) |
| TM orientation              | \(-70 \text{ mrad}\) | \(-10^{-9}\) |

### Table 4. Trapped magnetic potential mapping (Level C) results.

| Gyro No. | Data used | Relative residuals | Number of harmonics | Relative size of variations | Formal error in \( C_{TF}^{\text{C}} \) relative to \( C_g \) |
|----------|-----------|-------------------|--------------------|-----------------------------|--------------------------------------------------|
| 1        | \(-1 \text{ year}\) | 1.1% | 21 | 0.2%-3.0% | \(1.5 \times 10^{-3} - 7.0 \times 10^{-5}\) |
| 2        | \(-1 \text{ year}\) | 1.5% | 25 | 0.5%-1.5% | \(6.0 \times 10^{-5} - 3.0 \times 10^{-5}\) |
| 3        | \(-1 \text{ year}\) | 2.6% | 21 | 0.01%-1.0% | \(2.0 \times 10^{-4} - 1.6 \times 10^{-4}\) |
| 4        | \(-1 \text{ year}\) | 2.8% | 21 | 0.1%-0.3% | \(8.5 \times 10^{-5} - 6.5 \times 10^{-5}\) |
2.5. TFM results and their comparison with the results of LF SQUID signal analysis (scale factor variations)

Table 3 lists the estimated errors for the parameters specifying the motion of a rotor in the inertial frame. Also shown is the error in the orientation of the rotor with respect to the SQUID pickup loop, which is simply computed as the root-sum-squared of the errors in the three Euler angles. While the overall uncertainty in the rotor orientation was around 70 mrad (4 deg), it is the uncertainty in polhode phase, which was 50 mrad (3 deg), that played the most important role in the overall experiment uncertainty.

The accuracy of the trapped magnetic potential, i.e., of the coefficients $A_{lm}$ (Level C), is given in Table 4. The post-fit residuals are at a few percent level for all gyroscopes, which translates into an error in the predicted TF scale factor of $\sim 10^{-4}$ relative to the total scale factor, $C_g$. The highest uncertainty in the estimated $C_g^{TF}$ was for gyroscope 3, even though the post-fit residuals were smaller than those for gyroscope 4. This is because the polhode motion of gyroscope 3 vanishes rather quickly, resulting in poorer observability of the coefficients $A_{lm}$.

Figure 11 shows polhode variations in the TF contribution to the scale factor for all four gyroscopes over a period of a few hours on 10 November 2004. The time dependent standard deviations in $C_g^{TF}$ are provided in Figure 12.

The total gyroscope readout scale factor, $C_g(t)$, including the polhode modulation, was found independently in the LF SQUID signal analysis. However, the extent to which the LF analysis was capable of distinguishing between the London moment (LM) and LF TF contributions was somewhat uncertain. To verify this, it was useful to compare the HF and LF results on the scale factor variations. We plotted the LF results in Figure 11 along with the HF
ones, and their general agreement is seen to be good. The difference between the two (in the rms sense) is given in figure 13; the two determinations agree to $\sim 10^{-3}$ or better relative to the total scale factor, $C_g$. The best agreement is for gyroscope 3, despite the highest uncertainty in the estimated $C_g^{TF}$. This is not contradictory, because polhode variations in $C_g^{TF}$ relative to $C_g$ for gyroscope 3 were by far smaller than for other gyroscopes, as shown by the dashed lines in the figure.

There were two main products of the HF TFM that were utilized in the LF science analysis, which produced estimates of the relativistic drift rate. They were the polhode phase and angle and estimates of the time-varying contributions of the TF to the gyroscope readout scale factor. Estimates of the polhode phase and angle, with uncertainties listed in table 4, were vital inputs for models of both types of classical torques acting on the gyroscopes and of the scale factor variations. The TFM estimates of the time-varying scale factor, with uncertainties provided in figure 12, were used as a priori estimates and were refined further by the LF analysis.

3. Two-second filter (estimator): algorithm and implementation

The 2sF was the name given to the primary data analysis estimator, which used the LF SQUID and other signals to estimate the relativistic drift rates, as well as all other parameters needed to describe the data, for each of the four gyroscopes.
There are four key requirements for the success of any data analysis, including that of GP-B:

1. Understanding the gyro motion—patch effect torque models.
2. Understanding the readout signal structure—the measurement model.
3. Data preparation—accumulating experience during data preprocessing.
4. Numerical implementation of the filter—estimation techniques.

Each of the four requirements is discussed in detail.

### 3.1. Gyro motion and signal structure

A unique feature of GP-B LF SQUID readout was its dependence on the dynamics of gyro motion in the presence of disturbing patch-effect torques. Their rather complex model derived in [3] included both the misalignment and roll-polhode resonance torques. The analytic solution of the equations of motion found there provided the much needed explicit expression for the inertial gyro orientation at any moment of time (in practice, we used data points separated by two-second intervals, for the reasons explained in section 5 of [8], see also [7]).

One particular complication of the measurement model was the time-dependent misalignment torque coefficient, with its own sophisticated enough sub-model. The complexity of gyro dynamics strongly manifested itself also in its elaborate frequency composition (determining the corresponding time signatures) involving a number of critically important frequencies with widely spread values. Those were the roll, orbital, and annual frequencies, as well as the time-varying polhode frequency and its integer multiples.
As explained in section 5 of [3], the model of gyroscope readout was nonlinear. It included a time varying scale factor with the model derived in section 5.4 of [3], and the time history of the S/C orientation, which was estimated from the science telescope signal whose model is given in section 6 of the same paper (see also [8], section 5). The readout model can be written in the general form as

\[ z(t) = h(x, u(t)) + \nu(t). \]  

Here \( z(t) \) is the LF SQUID data, the function \( h \) is the nonlinear model for \( z(t) \) depending on time, the set, \( x \), of constant parameters to be estimated, also called the state vector, and on the vector of additional signals, \( u(t) \), i.e., on the external inputs to the model. The signals forming the vector \( u(t) \) included the telescope signal, the history of gyro orientation, the S/C roll phase, and the polhode phase and angle of a rotor. The LF SQUID data were contaminated by noise \( \nu(t) \).

### 3.2. Data preparation

Preparing data for the final analysis was crucial for its outcome. We gave a detailed account of this sophisticated process in [8].

### 3.3. The 2sF

The 2sF is a nonlinear least-square Bayesian estimator capable of handling large data sets (~10^7 data points) and estimating large numbers of parameters (~10^2–10^3).

The 2sF incorporated several sub-models for both the SQUID and telescope readouts, particularly those that described the gyro motion resulting from the linear relativistic motion and the one caused by classical torques. The filter was designed to be modular and scalable, so that the number of parameters estimated for each sub-model could be chosen independently (a special procedure of choosing these numbers is given in section 5.1). We described the LF SQUID signal model, including the pointing signal, in sections 5 and 6 of [3]. The measurement equation is also discussed in section 5.3 of [8].

The filter used only the LF SQUID data from the guide star valid (GSV) periods of the orbits, when the telescope was locked onto the GS in a stable manner, and its signal measured the S/C pointing.

#### 3.3.1. Discrete measurement equation

To describe the 2sF we write, first of all, the measurement equation (13) for a discrete set of measurement times \( t_k \), \( k = 1, 2, \ldots, K \):

\[ z_k = h_k(x) + \nu_k; \quad z_k = z(t_k), \quad h_k = h(x, u_k), \quad u_k = u(t_k). \]  

Employing now \( K \)-dimensional vectors

\[ Z = [z_1, z_2, \ldots, z_K]^T, \quad H(x) = [h_1(x), h_2(x), \ldots, h_K(x)]^T, \quad \nu = [\nu_1, \nu_2, \ldots, \nu_K]^T, \]

we can reduce the above equation (14) to the compact vector form

\[ Z = H(x) + \nu, \]  

used in the sequel.
3.3.2. The iterative algorithm. The 2sF is an iterative algorithm based on the linearization of the measurement model (16) about the current state vector estimate, $\hat{x}$ (of the dimension $n$), which is obtained at every step of iteration together with the corresponding covariance matrix, $P$. The linearized equation (16) is:

$$Z = H(\hat{x}) + J(\hat{x})(x - \hat{x}) + \mathcal{O}(x - \hat{x})^2 + \mathcal{N},$$

where the $K \times n$ Jacobian matrix is given by

$$J = \frac{\partial H}{\partial x} \bigg|_{x = \hat{x}}, \quad J_{kl} = \frac{\partial h_k}{\partial x_l} \bigg|_{x = \hat{x}}, \quad k = 1, 2, ..., K, \quad l = 1, 2, ..., n. \quad (18)$$

We now introduce the vector of innovations, $\delta Z$, also known as pre-fit residuals

$$\delta Z = Z - H(\hat{x}), \quad \text{dim}(\delta Z) = K, \quad (19)$$

and the correction vector

$$\delta x = x - \hat{x}. \quad (20)$$

The innovation, which is a standard term in the estimation theory, is the deviation of the calculated model value of the signal from the measured one, at a given iterative step.

Using these notations and working to linear order in $\delta x$, we rewrite equation (17) as the following linear structure:

$$\delta Z = J(\hat{x})\delta x + \mathcal{N}. \quad (21)$$

Since $\delta Z$ and $J(\hat{x})$ are known, equation (21) presents a linear estimation problem. Its unique solution for the estimate, $\delta \hat{x}$, with its computed covariance matrix, $P_0$, leads to the new updated estimate implied by the definition (27):

$$\hat{x}' = \hat{x} + \delta \hat{x}. \quad (22)$$

In our nonlinear estimator, the 2sF, we observed a stable convergence of iterative estimates: after some number of iterations (up to 20–30) the correction vector $\delta \hat{x}$ became small (see details in section 3.3.6) and, therefore, the estimate stabilized, i.e., $\hat{x}'$ became close to $\hat{x}$. This convergence asserted that the covariance matrix $P'$ of the estimate $\hat{x}'$ could be approximated by the covariance matrix $P_0$ of the correction $\delta \hat{x}$.
Because of this, it was natural to calculate the covariance matrix $P'$ of the current estimate $\hat{x}'$ at each step of the iteration process as $P' = P_0$.

In principle, solution (22) could be obtained for all available data points at once by means of the standard least square fit. However, in our case we had nearly $10^7$ data points and several hundred parameters to estimate, for each gyroscope. This means that the full Jacobian would have a very large dimension $\sim 10^7 \times 10^3$ causing thus serious computational difficulties, given the computational resources we had at our disposal. To overcome these difficulties, we used instead the so called batch-sequential information square root linear estimator [14]; we describe this technique below in section 3.3.5.

A flow chart of our iterative algorithm is shown in figure 14.

3.3.3. The Jacobian computation: sigma point technology. One of the key elements in the implementation of the above algorithm is the computationally efficient Jacobian calculation. As often happens in nonlinear estimation problems, the direct linearization combined with the explicit analytical computation of the Jacobian leads to biased estimates of the state vector, and, eventually, to the lack of robustness of the whole numerical procedure. The remedy is provided by the sigma point technology [15], a numerical method replacing analytical computation of $J$, which proved to be effective in a number of nonlinear estimation problems. The essence of this approach is constructing, in the vicinity of the the estimate $\hat{x}$, a family of the so called sigma-points by calculating, along with the component $\hat{x}'_l$, the corresponding quantities $\hat{x}'_{l\pm\sigma}$, $l = 1, 2, \ldots, n$ shifted by the covariance, $\sigma_l$, of the estimate $\hat{x}'_l$.

As a result, the estimate $\hat{x}$ is surrounded by a ‘cloud’ of sigma-points

$$\hat{x}'_\pm = \{ \hat{x}_1', \hat{x}_2', \ldots, \hat{x}_l' \pm \sigma_l, \ldots, \hat{x}_n' \}^T, \quad l = 1, 2, \ldots, n,$$

in the space of states. The elements of the Jacobian are computed numerically in terms of these points as:

$$J_{kl} = \frac{1}{2\sigma_l} \left[ h_k(\hat{x}'_l) - h_k(\hat{x}'_{l\pm}) \right].$$

Two qualitative ideas lie behind this construction that shed some light on its practical efficiency. One is of a geometric nature: the exact Jacobian $J(\hat{x})$ represents a hyper-plane in the state space which is tangent to the surface $H(x) = \text{const} = H(\hat{x})$ at the point $\hat{x}$. Alternatively, the hyper-plane determined by the sigma-Jacobian (24) is a secant one in the vicinity of $\hat{x}$, and it carries more information about the local structure of the surface than the former.

The other idea combines this geometric insight with a statistical motive: the actual position of the secant hyper-plane is specified by the covariances of the estimate $\hat{x}$ characterizing its uncertainty caused by the signal noise.

3.3.4. Initialization of iterations. The estimation process was bayesian, with the a priori state vector estimate and covariance, $\hat{x}^0$ and $P^0$ taken from previous analyses, including TFM and the two-floor estimator, described in detail in section 5.3 of [8] (see also [7]). As explained in this section, it is important in nonlinear estimation problems to start iterations not too far from the global minimum to make the convergence fast and robust. The two-floor estimator, working with preprocessed signals (gyro and telescope ‘slopes’ presented in section 5.2 of [8]), provided a priori estimates for the initial gyroscope orientations and parameters...
associated with the misalignment and roll-polhode resonance torques. The TFM analysis, discussed in section

3.3.5. Significant details of the estimation algorithm implementation. We now describe batch-
sequential information square root linear estimator that we used to solve our problem. The LF
SQUID data set, spanning one year of science data collection sampled at 0.5 Hz, was
subdivided into relatively small batches and processed one batch at a time. We took a single
GSV period of 45-50 min as one data batch; with the 2 s separation, it gives 1350–1500 data
points in a batch. There was one GSV period per orbit, and roughly 3000 orbits of usable data
were accumulated [8].

First, the computed innovations for each orbit were put through a spectral filter to
eliminate noise components outside the frequency range $0.5f_r \leq f \leq 1.5f_r$, where $f_r$ is the roll
frequency.

We mark the data from the $m$th orbit batch by a subscript $m$, $m = 1, 2, ..., M$, where
$M \sim 3,000$ is the total number of batches; $K_m$ stands for the number of data points in the
$m$th batch. The signal noise at the $m$th batch is represented by the $K_m \times K_m$ covariance matrix $R_m$
that was calculated during data preparation (see [8], section 5). The essence of sequential
approach is in the following.

For each batch $m$, $m = 1, 2, ..., M$, two quantities were calculated: the information
matrix $I_m$, and the weighted innovation $W_m$,

\begin{equation}
I_m = J_m^T R_m^{-1} J_m, \quad W_m = J_m^T R_m^{-1} \delta Z_m.
\end{equation}

(25)

According to the given definitions, the size of the information matrix is $n \times n$, and $W_m$ is a
column vector of the length $n$.

Accumulating then all the batches, we computed the cumulative information $n \times n$ matrix, $I$, as a sum of $I_m$, and the cumulative $n \times 1$ vector of weighted innovations, $W^{(Z)}$, as a
sum of $W_m$:

\begin{equation}
I = \sum_m I_m, \quad W^{(Z)} = \sum_m W_m.
\end{equation}

(26)

After the two matrices (26) were computed, the solution $\delta \hat{x}$ of the linear estimation problem
(21), i.e., the correction to the current state vector estimate, was found by the formula:

\begin{equation}
\delta \hat{x} = I^{-1} W^{(Z)}.
\end{equation}

(27)

producing the new state vector estimate, $\hat{x}'$ according to the expression (22). The covariance
matrix of $\delta \hat{x}$ was determined as

\begin{equation}
P_{\delta \hat{x}} = I^{-1}.
\end{equation}

(28)

Expressions (27) and (28) are equivalent to the classical solution of the LSQ problem (21)
formulated for the whole stretch of all available data, i.e., combined in a single batch.

Numerically all these operations were carried out using the library of Bierman algorithms
[14], using 64 bit words for the numerical calculations. Those algorithms exploit the square-
root householder transformations. The use of the square root of the information matrix, $I^{1/2}$,
was needed to ensure numerical stability of the estimation algorithm. It also effectively
doubled the numerical precision of the results. Processing data from the $m$th batch returned
the pair $I_m^{1/2}$ and $W_m$ which were used in the formulas (27) and (28).

This is how each iteration described in section 3.3.2 was accomplished in our data
analysis: the new estimate, $\hat{x}'$, and its covariance matrix, $P'$, were expressed, through $\delta \hat{x}$ and
$P_{\delta \hat{x}}$, by the formulas (22) and (23), respectively.
3.3.6. Convergence of iterations. Convergence of the iterative estimation algorithm was controlled by monitoring the scalar quantity

\[ \Delta = \delta \delta^T I \delta \].

This is the size of the correction vector taking into account the corresponding uncertainty, which provides a valid criterion for convergence since the vast majority of the elements of \( \delta \) tend to be large relative to their uncertainty. The iterations were stopped when \( \Delta \) reduced to some specified small value.

In nonlinear estimation problems it is important to use additional information, if available, to enhance and stabilize the iteration convergence. In our case the additional information about a part of the estimated parameters (not relativistic drift rates!) came from TFM and the algorithm initialization (see section 3.3.4). If the quantity (29) was still large after some iteration, the process was considered to have not yet converged, and the adaptive weighted Levenberg–Marquardt algorithm \cite{16} was applied to reduce \( \Delta \), using this additional information. This approach turned out to be effective allowing for stabilization of our estimation process convergence.

This concludes the description of the 2sF that we used to obtain the relativistic drift rate estimates from the GP-B signal. Our approach allowed us to carry out the necessary numerical calculations; however, with the computational resources available, they would take too long a time. Help came from the parallel computing described next.

4. Parallel computation

The 2sF was computationally intensive, requiring days to complete the analysis for a single gyroscope on conventional personal computers. Since about a thousand single gyroscope analyses were required, conventional computers were deemed unacceptable given programmatic time limitations. Thus, the decision was made to investigate a parallel processing implementation for the 2sF. A speed-up between 5 and 10 times over the serial version was considered sufficient for our purposes.

To implement the parallel solution, a cluster of multiprocessors was made available by the Stanford Aeronautics and Astronautics Department. The cluster consisted of 44 64 bit CPUs running Linux which was identical to the development environment used to develop

Figure 15. Structure of the serial code.
the serial filter. The MatlabMPI library, a Matlab parallel processing library developed by the MIT Lincoln Laboratory, was chosen for the parallel implementation, since it required little modification of the existing serial Matlab code and it provided the necessary reduction in computation time.

The following is an outline of the parallel computing implementation of the 2sF (see also figure 15).

- Initialize.
  - Load filter configuration.
  - Load initial parameters.
- Iterate until \( \Delta x = x_i - x_{i-1} \), the difference between successive estimates of the parameters, is below a predefined value.
  - Loop over gyroscopes (4).
    * Loop over data segments.
    * Loop over time steps.
      - Build \( h(x_i) \), the SQUID model as a function of the current parameter estimates.
      - Build Jacobian, \( J(x_i) = \partial f/\partial x \), with dimension (number of time steps \( \times \) number of parameters).
      - Load data \( Z(t) \), and perform fit \( Z(t) = J(x_i) \Delta x \).
    * Add a priori information (Bayesian estimation).
- Display results.
- Exit.

The most computationally intensive parts of the 2sF were first identified for consideration for parallelization, since they would naturally provide the greatest improvement in computation speed.

4.1. Parallel processing development approach

By studying the detailed profiling of the code execution, we identified two major areas where the computation was most intensive. These areas were:

1. Building the Jacobian matrix \( J(x) \) for each data segment, and
2. loading \( Z(t) \) measurements (see previous sections) and performing the data fit.

Building of the Jacobian matrix was done column by column, since each column represented one parameter of the model. Parallelization was introduced by assigning a subset of parameters to each CPU. Therefore, each CPU built part of the Jacobian matrix. Then all Jacobian components were merged together by a master CPU.

Unlike the construction of the Jacobian matrix, the fitting process for each data segment was performed row by row, because each row represented one particular time stamp. Therefore, parallelization was introduced by assigning a set of time stamps to each CPU. Each CPU performed a LSQ fit to its assigned data set and then returned the resulting parameter estimates and covariance to the master CPU, where the estimates were combined.

4.2. Parallel overhead

The parallel overhead is the amount of time needed to coordinate between processes carried out by the individual CPUs. This overhead does not exist in serial code execution so it is always referred to as the price of parallel processing. In our case we encountered two major areas of overhead related to data communication and coordination between the CPUs. The
first area was observed immediately after each CPU completed its part of the Jacobian: each CPU had to wait for the other CPUs to complete their parts before starting to build its part of the information matrix. The factors that contributed to this overhead were:

(1) Inter-processor communication medium. Processors in the cluster communicated by sending/receiving messages via the hard drive. The hard drive is a relatively slow communication medium in computer architectures. The filter performance degraded when many processors were added to the cluster due to the increasing load on the hard drive. This issue was alleviated by adopting a custom data communication technique, instead of using the MatlabMPI functions. This has reduced communication time by about 60%, as the plots in figure 16 show.

(2) Varying processing loads between processors while building the Jacobian. This is because different parts of the parameter vector required different computation times. This
means that all processors cannot proceed to the next step until the one with the highest load is done. This issue was resolved by developing a custom load balancing algorithm. It takes into account the computation weight of each group of parameters, including all submodels described in [3], and assigns the states to processors accordingly. The cost $P$ of reading from and writing to the disk was also taken into account according to

$$P = \sum_k \text{length}(\text{parameter subset}) \times (\text{computation weight}) + (\text{disk IO weight}),$$  \hspace{1cm} (30)

where $k$ denotes each subset of parameters.

Applying this algorithm produced a uniform load among all processors as seen on figure 17.

(3) Weak memory management for each processor resulted in massive disk swapping that slowed processing and worsened data communication. This was resolved by tightly managing memory allocation to processors to prevent swapping to the disk, or ‘out of memory’ errors. Memory was allocated on demand and released immediately once not needed. We managed to save $\sim 40\%$ of memory per processor, and eliminated ‘out of memory’ errors.

4.3. Further code optimizations

With the help of profiling tools, we were able to find parts of the code that could be optimized to work more efficiently and hence contribute to speeding up the filter operation. Our efforts focused on using best practices in programming rather than changing how the filter works. Examples of these kinds of optimizations include:

- Caching values that are computed repeatedly but yielded the same results every time.
- Caching values that are read from the disk repeatedly.

Returns on optimization efforts yielded a 90% improvement in efficiency. For example, computing $h(x)$, the vector of time signatures (see the analysis model in [3]), went from taking $\sim 1$ h per run to $< 5$ min. The final structure of the parallel code is presented in the diagram, figure 18. In this figure, the new coordinating components are shown, i.e., distributing the Jacobian to all nodes, and sending information matrices to the master node.
4.4. Run scheduler

The computer cluster utilized by the data analysis team was also used by other research groups at Stanford. In addition, roughly 12 high-speed serial computers were also available for data analysis. As a result, a custom scheduler was needed to execute both the serial and parallel data analysis runs efficiently. It was initiated with a batch of options files.

From the many existing scheduler algorithms, we chose a first come, first served scheduler for its simplicity. There was no need to prioritize certain runs relative to others. So we assigned a waiting time for the execution of each, and if this time ended without the run having finished, we halted the run and made the processor move to the next run in the queue. The scheduler was used to schedule the more than 1100 individual gyroscope analysis runs that needed to be executed over roughly a 1 year period.

In addition to handling run execution, the scheduler managed version control for the data analysis software. It did this by creating a unique directory for every run containing its options file, code and results. Each of the over 1100 runs was cataloged by date of the run and by revision of the data analysis code.

4.5. Outcomes of parallel processing

Using parallel processing to test the filter proved to be very useful. The first version of the parallel filter was seven times faster than the serial one which reduced typical run time from a week to a single day. Further improvements led to more speedup and the final parallel filter was fourteen times faster, reducing thus the run time to a half of a day. Figure 19 illustrates the speedup gained with each version. The parallel filter was used to run more than 1100 analyses, which helped the team modify the filter promptly.
4.6. Tools for analyzing estimation results

The large number of data analysis runs, options available for each run, and parameters estimated mandated the need for a user interface that managed the runs and transformed estimated parameters into data that could be easily visualized. The result viewer was therefore developed which was a graphical user interface (GUI) that allowed the science team to search for runs, displays results, and generate related plots.

The results viewer facilitated comparison of the results from different runs by plotting them in a single chart. It also allowed the science team to search for certain runs by gyroscope and segment number, which were documented by the version control system. The user could view important variables of the selected run such as the relativist drift rate, polhode modulation coefficients for the readout scale factor, and classical torque coefficients. Additional run information including plots of the convergence rates, and the final relativistic drift rate estimate confidence ellipses were automatically generated and placed in the associated directory for each run. In addition, the results viewer incorporated some functions to show the residual analysis or plot certain reconstructed signals, including the gyroscope orientation time histories or the misalignment torque coefficients as a function of time. Finally, the GUI could save the comparison results between different runs and document it by the date it was produced.

5. Relativistic drift rate estimation and results

5.1. Determination of the number of parameters to be estimated

Before the final analysis of any gyroscope, the number of parameters in the LF SQUID model had to be determined, for each gyro-segment (see below). This is because several sub-models (see [3], sections 4–7) describing patch effect torques and readout scale factor are infinite linear combinations of basic functions, with the coefficients to be estimated; naturally, these series had to be properly truncated for numerical processing. Table 5 lists the various sub-models for which the number of parameters had to be determined.

To find the appropriate number of terms to include, we designed and implemented a special procedure. It was performed many times for each gyroscope, every time increasing the number of parameters in a single sub-model. The initial ‘baseline’ analysis estimated between zero and two parameters per sub-model, producing a baseline relativistic drift rate estimate. Subsequently, we increased the number of parameters for a single sub-model by the minimum amount (typically one or two). We then compared the new relativity estimates with the baseline ones, and if the results were statistically consistent, then we took the number of terms in that particular sub-model equal to the baseline number. If, on the other hand, the results were not statistically consistent, then we took the new, larger, number of sub-model terms as a new baseline, and ran the analysis again with the increased number of parameters.

This procedure continued until the addition of one more parameter to any of the sub-models did not change the result in a statistically significant way. By definition, the (new) relativity estimates in both the North–South and West–East directions were considered statistically significant if they did not differ from the baseline ones by more than one–half of the standard deviation. In fact, the addition of a single parameter beyond the baseline number, for most sub-models would actually change the relativistic drift rates by less than 0.1 of the standard deviation. Figure 20 graphically depicts this procedure. In the illustrative example shown in figure 20, three terms are used as the baseline number and an additional analysis using four terms is used as one parameter sensitivity run.
5.2. Single gyroscope estimation and results

Six out of ten available data segments (see [8], section 2.1) were utilized in the data analysis. The first segment was excluded because of inadequate S/C pointing, and three other segments were too short (∼1 week or less) to be of any serious value. The complete nonlinear data analysis by the 2sF was independently performed for each of the four gyroscopes. First, every signal segment was analyzed, and these analyses produced 24 relativistic drift rate estimates for each gyroscope. This was done to verify mathematical models we used to describe the physics of the experiment, and the estimation techniques in place. These individual gyroscope results were not used to calculate the final experimental result. Instead, the data from all six segments were combined to produce a single estimate for each of the four gyroscopes. The combined six-segment result for each gyroscope resulted in smaller statistical errors compared with the linear combination of the results from each single segment. This is because several elements of the model, including the gyroscope readout scale factor, the misalignment torque coefficient, and the roll-polhode resonance torque, span all six segments and enter the measurement model in a nonlinear way. The combined six-segment analysis therefore requires fewer estimation parameters than the sum of the six individual segment analyses. When ready, the four individual gyroscope results were combined in a linear way to produce the final four-gyro joint result. A combined four-gyro nonlinear analysis would have produce the same result as the linear combination of the four individual results since the only states that were common to all four gyroscopes were the two relativistic drift rate estimates.

For each of the 24 ‘gyro-segment’ analyses we used the procedure described in the previous section to determine the number of parameters needed to model each gyroscope. In general, the number of parameters needed for each gyro-segment was smaller than the corresponding number to model each gyroscope using all six segments of data together.

Relativistic drift rate estimates for all 24 gyro-segments are shown in figures 21–24. Plotted in these figures are the 95% confidence ellipses associated with the relativistic drift rate estimates and their statistical errors for each gyro-segment. Also plotted are the 95% confidence ellipses for each gyroscope using all six segments of data together. Systematic errors for individual gyro-segment estimates were not determined, but would, of course, tend to increase the size of the confidence ellipses. Nevertheless, from the figures it is clear that the
results from each gyro-segment are statistically consistent with one another and with the six-
segment result. This consistency builds confidence in the mathematical models we derived to
describe the gyro dynamics and readout, as well as in the accuracy of the $2sF$ estimation
routine. The same mathematical models and estimation techniques were used for each gyro-

Figure 21. Relativistic drift rate estimates for gyroscope 1 using only data from
individual segments (colored ellipses) and using all segments (black ellipse).

Figure 22. Relativistic drift rate estimates for gyroscope 2 using only data from
individual segments (colored ellipses) and using all segments (black ellipse).
segment, though, as said, the number of parameters in each estimation was generally different.

Once the analysis of each gyro-segment was completed and the mathematical models and estimation techniques were verified, we analyzed all six segments of data together. Again, we

Figure 23. Relativistic drift rate estimates for gyroscope 3 using only data from individual segments (colored ellipses) and using all segments (black ellipse).

Figure 24. Relativistic drift rate estimates for gyroscope 3 using only data from individual segments (colored ellipses) and using all segments (black ellipse).
exploited the above procedure to determine the baseline number of parameters that had to be estimated for a given gyroscope.

The number of parameters and the amount of data used for each gyroscope is given in table 6. The data rate was 0.5 samples/sec, so one day of data corresponded to 43 200 samples of LF SQUID signal, pointing signal, roll phase signal, etc. The pointing signal was based on a model for the vehicle motion using both the telescope signal during GSI and the LF SQUID signals from all four gyroscopes during both GSV and GSI [8]. The number of the estimated states per gyroscope depended essentially on the number of roll-polhode resonances it went through: every one of them required two states to estimate. Only $\sim 25\%$ of the states had a correlation coefficient $0.1 > 0.1$ with the two relativistic drift rate parameters, and the largest correlations ($\sim 0.5$) only existed for a few states. The $\chi^2$ value based on the post-fit residuals was $\lesssim 1$ for all gyroscopes.

Next we performed a parameter sensitivity analysis to make sure that the result of the experiment did not depend on any particular choice of the parameters number. In addition to the baseline relativistic drift rate estimate, nine other estimates were computed; for each them, the number of parameters used in one of the sub-models was increased by the minimum amount. Hence ten drift rate estimates and covariances were obtained for each gyroscope. We

Table 5. Parameter sensitivities evaluated for each gyroscope.

| Run | Description | Relevant equation in [3] |
|-----|-------------|-------------------------|
| 1   | Baseline number of parameters | |
| 2   | $+ 2m$ TF scale factor parameters, polhode harmonics | Equation (45), $C_{\alpha\beta}^{\alpha\beta}$, $m > 0$ |
| 3   | $+ 1$ TF scale factor parameter, low frequency | Equation (45), $C_{\alpha\beta}^{\alpha\beta}$, $m = 0$ |
| 4   | $+ 1$ polhode phase exponential model parameter | See section 4.2 |
| 5   | $+ l + 1$ trapped magnetic potential parameters | Equation (A19), $A_{lm}$ |
| 6   | $+ 2$ electronics scale factor parameters | Equation (48), $d_{\alpha\beta}^{\alpha\beta}$ |
| 7   | $+ 2$ misalignment torque coefficient parameters | Equation (61), $k_{\alpha\beta}^{\alpha\beta}$ |
| 8   | $+ 2$ misalignment torque coeff. parameters, low frequency | See section 5.2 |
| 9   | $+ 1$ telescope scale factor parameter | Equation (80), $C_T$ |
| 10  | Alternate S/C pointing model | See [17] |

Table 6. Number of parameters and days of data used for each gyroscope.

| Source | No. parameters | No. days of data |
|--------|----------------|------------------|
| Gyroscope 1 | 154 | 200.7 |
| Gyroscope 2 | 662 | 219.7 |
| Gyroscope 3 | 139 | 188.8 |
| Gyroscope 4 | 272 | 241.7 |
| All gyroscopes | 1227 | 850.9 |
present details for each of the nine parameter sensitivity runs, including references to the equations describing each sub-model, in table 5. Of the nine sensitivity runs, only one, run 10 in table 5, involved some adjustments to the assumptions used for the model. It studied the sensitivity to the S/C pointing, $\vec{\tau}$, discussed in detail in [17]; see also [3], section 6.1.

We found the final drift rate estimate for each gyroscope as the weighted mean of the ten estimates associated with the parameters sensitivity runs; the runs are given in table 5. Table 7 lists these weighted means in both the North–South and West–East directions. The weighted mean estimate, as well as the ten sensitivity run estimates, is also plotted for all four gyroscopes in figure 25. The covariances associated with the scatter of the ten sensitivity estimates were also computed; the resulting 95% confidence ellipses for each gyroscope are shown figure 25. It is interesting that the drift rate estimates for some gyroscopes were more sensitive to the number of states than for others. In particular, gyroscope 1 was more, and gyroscope 4 was less sensitive to this number.

We calculated the total covariance of the drift rate estimates for each gyroscope as the sum of the statistical, systematic, and parameter sensitivity covariances. The parameter sensitivity covariance was just discussed; it is plotted in figure 25. The statistical covariance was computed by the 2sF in each of its runs by scaling the LF SQUID measurement noise with the sensitivity of the measurement to relativistic drift rate in the usual way. The Jacobian matrix, comprised of the partial derivatives of the measurement with respect to each state,
scaled the measurement noise so that the statistical uncertainty accounts for the correlation between the drift rate estimates and all other parameters. Figure 26 compares the statistical and parameter sensitivity using 95% confidence intervals.

The systematic covariance accounts for the relatively small uncertainties caused by all effects that were not included in the measurement model through either the gyro dynamics or the readout model. These systematic uncertainties are discussed below in section 5.4. They include small contribution of classical torques, particularly those due to the rotor asphericity, and readout errors associated with the analog-to-digital (A/D) converters, S/C roll phase and timing uncertainties, telescope nonlinearities, and SQUID cross-couplings. Figure 27 shows the overall 95% confidence ellipses for each gyroscope which is the sum of the statistical, parameter sensitivity, and systematic uncertainties.

Figure 26. The 95% confidence intervals associated with the parameter sensitivities (dotted ellipses) and the statistical errors (dashed ellipses) for each gyroscope.

Figure 27. The 95% confidence intervals associated with the parameter sensitivities (dotted ellipses), the statistical errors (dashed ellipses), and the total errors (solid ellipses) for each gyroscope. The total error covariance is the sum of the parameter sensitivity, the statistical, and the systematic covariances.
5.3. Combined gyroscope estimation and results

The mean relativistic drift rate estimates for each gyroscope weighted by their statistical covariances were combined linearly to produce the aggregate result of the experiment [2]. The total experimental uncertainty was again the sum of statistical, sensitivity, and systematic errors. We modified the statistical uncertainty to account for the slight over-dispersion in the West–East, and slight under-dispersion in the North–South, directions among the four estimates produced by individual gyroscopes. The standard statistical approach used to handle under- and over-dispersion is described in appendix.

We computed the parameter sensitivity covariance of the combined four-gyroscope estimates as a result of all $10^4$ possible combinations of the individual gyroscope sensitivity
estimates. Figure 28 shows the joint four-gyroscope 95% confidence ellipses associated with the parameter sensitivity uncertainty, the statistical uncertainty, and the total experimental uncertainty. The systematic uncertainty for the combined result is discussed in the next section.

We give numerical values of relativistic drift rate estimates and their one-sigma total uncertainties for each gyroscope, as well as the combined four gyroscope results, in table 7. The North–South and West–East relativistic drift rates predicted by GR are also provided, for comparison. We explained how the GR drift rate predictions were calculated in [3], section 2.2. Figure 29 provides a graphical representation of the results listed in table 7.

5.4. Treatment of systematic errors

Systematic uncertainty arises from effects not included in the measurement model. These effects did not need to be modeled because we established that they were sufficiently small to not appreciably alter the science result. The root sum square impact of unmodeled effects on the relativity results for any one gyroscope is shown to be less than 2.6 mas yr\(^{-1}\), and their combined impact on the four gyroscope result is less than 1.3 mas yr\(^{-1}\). More than 200 individual unmodeled effects were analyzed prior to launch. ‘Error tree’ software allowed us to calculate the experiment uncertainty as a function of 56 parameters such as rotor mass unbalance, rotor suspension imperfections, S/C pointing uncertainty, readout uncertainties and roll rate imperfections.

After launch the effect of systematics on the experiment uncertainty was reassessed. Several new sources were identified, and the impact of all sources was updated based upon flight data.

Systematic errors can be grouped into seven categories as follows:

1. Gyroscope torques.
2. Telescope pointing effects.
3. Gyroscope readout effects.
4. Roll phase and timing effects.
5. GR proper motion uncertainty.
6. Solar geodetic effect uncertainty.
7. GR prediction uncertainty.

Gyro torque and readout contributions to systematic uncertainty vary by gyroscope. The other effects are gyroscope-independent.

5.4.1. Gyroscope torques. A perfect GP-B gyroscope would be torque-free and thus would have no torque-induced drift rate. In practice, imperfections in implementation resulted in torques and drift rate errors, but aside from the modeled patch-effect torques their total impact on the combined four gyroscope experiment uncertainty was less than 0.1 mas yr\(^{-1}\) in both the NS and WE directions.

There are two classes of torques: support-dependent and support-independent. Support dependent torques arise from residual rotor acceleration and the electrostatic support force generated by the gyro suspension system. As its name implies, support-independent torques are independent of the suspension effort. Both types of torques are conveniently split into three groups characterized by the spinning rotor position inside the housing:

1. Rotor, with its small imperfections due to rotor shape and mass unbalance, is centered in the cavity with its spin axis perfectly aligned with the S/C roll axis (nominal position).
The rotor spin axis is perfectly aligned with the S/C roll axis, but the rotor center is not at the center of the housing cavity (miscentered position).

Rotor is centered in the housing cavity, but its spin axis is not aligned with the S/C roll axis (misaligned position).

Note that both the misalignment angle and miscentering were small (the latter as compared to the nominal rotor-to-housing gap), so the torques had only to be calculated to linear order in these parameters, making the above separation of their effects possible (torques in each position are found independently of those in two other positions). Analysis performed prior to launch \[18\] bounded the net effect of all torques to less than 0.1 mas yr\(^{-1}\).

Support independent torques include those due to magnetic \(B \times H\) interaction, differential gas damping, gravity gradient, and electrostatic torques due to position sensing, rotor charge, and eddy currents. The \(B \times H\) torque arises from interaction of the rotors LM with the gyro local magnetic shield. Spatial variation of the residual gas surrounding each gyro causes differential gas damping torque. The torque due to the Earth gravity gradient is proportional to the sine of the angle between the orbit plane and the direction to the GR. To reduce electrostatic torques the position sensing signal was minimized, the rotor potential was kept below 0.015 V and no highly conductive materials were allowed near the gyroscopes. Prior to launch the predicted drift rate due to all support independent torques was less than 0.01 mas yr\(^{-1}\).

The best estimate of torque-induced drift rate depends upon many parameters. The precise values of some of these were not known until after launch. Some of the values achieved on-orbit were better than expected reducing the torque-induced drift rate, while others were worse and thus gave a classical drift rate increase. The on-orbit rotor spin speed, \(f_r\), ranged from 61.8 revolutions per second to 82.1 revolutions per second, or approximately half the pre-launch expectation. With most torques independent of \(f_r\) most torque-induced drift rates, proportional to the ratio of a torque over \(f_r\), were expected to increase. A few torques were proportional to \(f_r\) so that the associated induced drift rates were independent of

### Table 8. Contributions to experiment uncertainty.

| Contribution                  | NS (mas yr\(^{-1}\)) | WE (mas yr\(^{-1}\)) |
|-------------------------------|----------------------|----------------------|
| Statistical (modeled)         | 16.8                 | 5.9                  |
| Systematic (unmodeled)        |                      |                      |
| Parameter sensitivity         | 7.1                  | 4.0                  |
| GR motion                     | 0.1                  | 0.1                  |
| Solar geodetic effect         | 0.3                  | 0.6                  |
| Telescope readout             | 0.5                  | 0.5                  |
| Other readout uncertainties   | <1                   | <1                   |
| Other classical torques       | <0.3                 | <0.4                 |
| Total stat. + Sys.            | 18.3                 | 7.2                  |

### Table 9. Mission-average spin axis misalignment (in arcsec).

| Component | Gyro 1 | Gyro 2 | Gyro 3 | Gyro 4 |
|-----------|--------|--------|--------|--------|
| NS        | -4.85  | 4.37   | 6.58   | 9.23   |
| WE        | -7.29  | -3.68  | -7.31  | -7.27  |

(2) The rotor spin axis is perfectly aligned with the S/C roll axis, but the rotor center is not at the center of the housing cavity (miscentered position).

(3) Rotor is centered in the housing cavity, but its spin axis is not aligned with the S/C roll axis (misaligned position).
Two torques scaled as the square of $f_s$ leading to increased drift rate with $f_s$. In practice, the reduced $f_s$ caused an increased total classical drift rate. In addition to $f_s$, the values for many torque parameters were updated following launch using flight data. Rotor shape terms, measured at odd harmonics of spin frequency, were reduced by a factor of three for gyro four to more than a factor of eight for gyro three. This improvement resulted from conservative assumptions made prior to launch combined with a post-launch rotor mass unbalance that was lower than the pre-flight estimate. The actual rotor mass unbalances ranged from 10.1 to 4.8 nm or 65% to 300% smaller than pre-launch worst-case estimates [19, 20]. Centrifugal bulge was approximately a factor of four smaller than expected due to reduced rotor spin speed. The mission-averaged orbit co-inclination was $10^{-3}$ rad, or 30% lower than the pre-launch estimate. Due to its operational mode, gyroscope 2 had ten times higher support force than expected. The maximum rotor TF was 0.3 nT, or three times better than the 0.9 nT requirement. The gas pressure was less than $10^{-12}$ Pa, or $10^3$ times better than required. The rotor misalignment was larger than the pre-launch expectation. Table 9 gives the NS and EW mission-averaged misalignment for the four gyroscopes.

Table 10 gives the gyroscopic drift rates due to all unmodeled support-dependent and support-independent torques in the North–South and West–East directions (exclusive of patch-effect torques). Torques due to rotor misalignment were larger than both nominal position torques and miscentered position torques for the on-orbit conditions as specified immediately above. The combined four gyro experiment uncertainty due to unmodeled torques was less than 0.2 mas yr$^{-1}$ for all gyroscopes except # 2, whose classical drift rate still did not exceed 0.6 mas yr$^{-1}$.

### 5.4.2. Telescope pointing effects.

Three potential sources of telescope pointing error were: (1) inertially-fixed thermal biases, (2) telescope scale factor, and (3) optics and electronics nonlinearity.

1. **Inertially-fixed thermal biases.** Thermal variation at roll frequency in window 4, the telescope detectors and the telescope electronics [21] caused biases in the telescope data that were indistinguishable from inertially-fixed pointing errors. Using measured temperatures and temperature sensitivities these biases gave equivalent drift rate errors of less than 0.4 mas yr$^{-1}$.

2. **Telescope scale factor.** Telescope pointing had to be subtracted from the gyro signal to give drift rates referenced to the GR. The telescope scale factor was used to convert telescope output signal to S/C pointing, thereby allowing the needed subtraction. Uncertainty in the scale factor caused uncertainty in the subtraction and therefore uncertainty in the estimates of the relativistic drift rates. The scale factor and its uncertainty were determined by comparing signals present in both the gyro and telescope. The telescope scale factor uncertainty was measured to 1%. With a typical inertial mispointing of 20 mas, the error associated with scale factor uncertainty was 0.2 mas yr$^{-1}$. 

| Component | Gyro 1 | Gyro 2 | Gyro 3 | Gyro 4 |
|-----------|--------|--------|--------|--------|
| NS        | 0.18   | 0.58   | 0.12   | 0.13   |
| WE        | 0.18   | 0.58   | 0.12   | 0.13   |
Optics and electronics nonlinearity. The telescope provided pointing information in two directions for S/C ATC. Telescope noise and ATC limitations prevented perfect tracking of the GS. The resulting non-zero telescope signal was proportional to pointing for small angles with a cubic correction term to give the most accurate pointing information. Ground and flight analysis provided the values of coefficients at the linear and cubic terms for pointing angles as large as the 400 mas maximum, accepted for use in the science analysis. Use of the cubic term eliminated much of the error due to telescope nonlinearity, leaving only the 4% uncertainty in the cubic term as error. This error applied to the relativity analysis gave an uncertainty of less than 0.1 mas yr$^{-1}$.

5.4.3. Gyroscope readout effects. Six potential sources of gyroscope readout systematic uncertainty were: (1) coupling to the Earth magnetic field, (2) thermal effects, (3) gyro-to-gyro cross coupling, (4) sensitivity to rotor position in its housing cavity, (5) preferred codes in the A/D converters that processed the science data, and (6) uncertainty due to interference from other GP-B electronics. The total impact of these sources on experiment uncertainty was less than 1.0 mas yr$^{-1}$.

Prior to launch, coupling of the Earth magnetic field to each readout was measured with the rotors delevitated, giving an ac magnetic field attenuation factor greater than the required $2 \times 10^{12}$. On-orbit, this measurement was repeated with the collection of 23 h of SQUID data. The power spectral density of these data showed no observable signal at roll frequency, thus providing a 0.3 mas yr$^{-1}$ upper limit on magnetic coupling.

The impact of thermal effects on the experiment uncertainty was bounded by active SQUID temperature control providing better than 2 μK stability near key science frequencies, and a low SQUID temperature sensitivity of less than 0.3 arcsec K$^{-1}$. An end-to-end on-orbit measurement of temperature sensitivity was performed with the rotors delevitated. The power spectral density of SQUID data acquired in this configuration showed no observable signal at roll frequency, thus providing a 0.4 mas yr$^{-1}$ upper limit on thermal effects.
Readout cross coupling, i.e., signal coupling between adjacent gyroscopes may cause uncertainty. This coupling may arise from magnetic coupling between adjacent gyroscope pickup loops, or from cross coupling inside the two readout electronics boxes (each of them operated two SQUIDs). The fractional amount of magnetic coupling was determined by measuring the size of a rotors unique spin frequency signal as detected in an adjacent readout system and comparing it to the size of the signal in its own readout system. An upper limit for electronic coupling was determined by turning a SQUID on or off and measuring any resulting change to adjacent readouts. In figure 30, the large change in the output voltage when SQUIDs 3 and 4 were turned on caused no observable change in the output of the slowly varying signals from SQUIDs 1 and 2. No detectable change was observed in repeated on-off cycles, providing a 0.2 mas yr\(^{-1}\) limit on drift rate error due to cross coupling.

Readout sensitivity to a rotor location within its housing and the presence of rotor motion at key science frequencies gave rise to some uncertainty. Rotor motion due to gravity gradient at roll frequency modulated by twice orbital frequency allowed measurement of rotor
sensitivity. Figure 31 shows the SQUID signal (top) that resulted from the rotor motion caused by gravity gradient (bottom). The data are Fourier components at the gravity gradient frequency. The step corresponds to a change in drag free configuration. The resulting 2 mas yr\(^{-1}\) frequency-independent readout sensitivity, when applied to the measured motion of the rotor at key frequencies, bounded the corresponding drift rate to 0.1 mas yr\(^{-1}\).

The conversion of a SQUIDs analog output voltage to digital counts was imperfect. Figure 32 below is a histogram of the normalized number of times a count value occurred as a function of count value modulo 256. Perfect conversion would give a nominal 0.5 for the normalized number of occurrences for each count value. The comb-like pattern indicates preferred codes. These preferred counts were observed at specific times in the data acquisition process. Figure 33 shows the normalized number of count occurrences for converter bits 1–9 as a function of time modulo 0.1 s. The impact of these preferred counts on the relativity result was less than 0.1 mas yr\(^{-1}\).

Spurious signals in readout systems 1 and 3 were traced to a single non-readout electronics box that was used to monitor system temperatures, Earth magnetic field, and other non-critical items. The spurious signals disappeared when this box was turned off 28 April 2005. After this date, the box was operated only occasionally.

The spurious signal had a chirped time signature. Its frequency varied from 0 Hz to above the readout 5 Hz bandwidth. SQUID data was excised when the frequency of the spurious signal was near roll frequency or near the 5 Hz data sampling rate. Approximately 8% of the data was rejected. The resulting impact on experiment uncertainty was less than 0.5 mas yr\(^{-1}\).

Note that separate from the readout effects described above was the scale factor variation of the readout electronics. This scale factor variation was modeled [3]. The contribution of this modeling to the experiment statistical uncertainty was described in the previous section. 5.4.4. Roll phase and timing effects. An error in roll phase determination is equivalent to a spurious coordinate rotation, resulting in possible mixing of the NS and WE drift rate estimates. Timing errors cause a similar effect.

Redundant star trackers measured the roll phase. These two sensors were mounted on the S/C exterior. Although the sensors were attached to a common graphite ring that was
designed to constrain differential and common mode motion, the sensors orientations were found to depend upon temperature. Analysis of long term sensor orientation bounded the resulting roll phase uncertainty as defined by the variation of the sensors measured roll phase difference. A fit of differential mode distortion to thermal gradients provided ring compression and torsion. Figure 34 shows the ring distortion in terms of equivalent roll phase motion, with sensor A giving a 6 arcsec rms variation over the mission. This variation corresponds to less than 0.1 marcs yr$^{-1}$ error in the relativistic drift rates.

Figure 34. Roll phase variation over the course of the mission.
The phase of the rolling S/C may be compared with the phase of one of the spinning rotors. This comparison is useful because the rotor is not influenced by S/C thermal distortion. The spin phase is determined from the rotor TF signal present in the gyro readout at spin frequency and its harmonics. Figure 35 shows roll phase error as determined from gyroscope 1 spin data (in blue) and from star tracker A data (in red).

![Figure 35. Roll phase uncertainty versus time.](image)

![Figure 36. Timing stability.](image)

The phase of the rolling S/C may be compared with the phase of one of the spinning rotors. This comparison is useful because the rotor is not influenced by S/C thermal distortion. The spin phase is determined from the rotor TF signal present in the gyro readout at spin frequency and its harmonics. Figure 35 shows roll phase error as determined from gyroscope 1 spin data (in blue) and from star tracker A data (in red).
Time information is derived from the on-board GPS receiver which supplies a pulse-per-second signal at coordinated universal time (UTC) integer seconds. Conversion between vehicle time and UTC is performed in ground processing. An offset of 0.044 s is incorporated to account for the latency of the readout system low pass filter. Figure 36 shows 2 ms timing stability, consistent with 0.2 mas yr\(^{-1}\) drift rate uncertainty (see more on timing in paper [22]).

5.4.5. GS proper motion uncertainty. Uncertainty in the proper motion of the guide star IM Pegasi contributed to the experiment uncertainty. Using very-long-baseline-interferometry, the proper motion of IM Pegasi (HR 8703) was studied [23] using 35 epochs between 1997 and 2005. The proper motion in an extragalactic reference frame closely related to the International Celestial Reference Frame 2 (ICRF2) was measured to be \(-20.83 \pm 0.09\) mas yr\(^{-1}\) in right ascension and \(-27.27 \pm 0.09\) mas yr\(^{-1}\) in declination: here the uncertainty is the total standard uncertainty including plausible systematic uncertainty.

5.4.6. Solar geodetic effect uncertainty. The measured gyroscope drift rates had to be adjusted for the 19.2 mas yr\(^{-1}\) solar geodetic effect caused by the motion of the Earth about the Sun. The uncertainty associated with the solar geodetic effect was conservatively based on the uncertainty of previous measurements. This uncertainty has contributions in both the NS and WE directions of 0.3 and 0.6 mas yr\(^{-1}\) respectively. It is worth noting that this potential systematic error is negligible if either the PPN parameter, \(\gamma\), is used to calculate the solar geodetic effect, or if GR is assumed to be correct in the case of the solar geodetic effect.

5.4.7. GR prediction uncertainty. The uncertainties in the predicted drift rates for the GP-B gyroscopes are dominated by the uncertainty in the S/C orbit. S/C position and velocity were determined using on-board GPS [24]. The data were fit in 30 h-long batches. Altitude and velocity differences in successive, overlapping batches define the uncertainty. The position uncertainty is less than 2.5 m and the velocity uncertainty is less than 2.2 mm s\(^{-1}\). These uncertainties are \(\sim 10^6\) smaller than the S/C position and velocity values, and result in negligible experiment uncertainty.

6. Concluding remarks

We have described how the GP-B results were obtained by analyzing the properly prepared LF SQUID signal. We also explained the analysis of HF SQUID signal (TFM) that provided the needed parameters of gyro motion (spin speed, spin-down rate, polhode angle and phase, and an initial estimate of the time-varying gyroscope scale factor). Finally, we discussed the results, as well as their statistical and systematic uncertainties in much detail.

This paper concludes the trilogy on GP-B data analysis; the underlying theory was presented in [3], while paper [8] was devoted to the handling of data prior to the final analysis (data preprocessing and preparation).

The GP-B data reduction effort, that was going on during the flight in 2004–2005, and was completed only by 2011, required a broad set of achievements, from experimental and theoretical physics insights to engineering accomplishments to a high level proficiency in numerical implementation of signal analysis (estimation techniques). It was successful only due to an effective collaboration of diverse specialists—theoretical and experimental physicists, engineers of different specializations, specialists in signal analysis, computer scientists and programmers.
Of course, it is impossible to tell the whole story of our data analysis. We tried, instead, to give a detailed enough, but not overwhelming, picture answering the most significant questions we imagined. The intent is that these papers might be helpful for future analysis of laboratory and space experiments dealing with a complex variety of measured signals.

Acknowledgments

This work was partly supported by a KACST gift to Stanford University. The authors are grateful to all people who contributed to the GP-B data reduction, directly or indirectly.

Appendix. Corrections for over/underdispersion in four-gyroism relativity results

Here is how the error bars of the combined result are found. Let \( i = \{1, 2, 3, 4\} \) designate the gyroscopes, and \( \nu = \{\text{ns, we}\} \) indicate the inertial directions defined in section 3.1 of DA I. We denote \( r_\nu^i \) the estimate of the relativistic drift in the direction \( \nu \) from the \( i \)th gyroscope, such as \( r_{\text{ns}}^1 \) for the NS estimate from the first gyroscope, or \( r_{\text{we}}^3 \) for the WE estimate from the third gyroscope, etc. We also write \( r_\nu \) for the average of the four estimates in the direction \( \nu \).

Each pair of estimates \( r_{\text{ns}}^i, r_{\text{we}}^i \) comes with their 2 \( \times \) 2 covariance matrix \( P^i \),

\[
P^i = \begin{pmatrix} a^i & b^i \\ b^i & c^i \end{pmatrix}, \quad a^i = \left( \sigma_{\text{ns}}^i \right)^2, \quad b^i = \left( \sigma_{\text{we}}^i \right)^2,
\]

whose diagonal elements are the corresponding variances (sigmas squared). The covariance matrix of the combined four-gyro result giving its statistical error is defined in a standard way as

\[
P_{\text{stat}} = \left( \sum_{i=1}^{4} P^i \right)^{-1} = \begin{pmatrix} a & b \\ b & c \end{pmatrix}
\]

Were the four measurements of the relativistic drift rates by different gyroscopes completely independent, this would be the final answer for the statistical error of their combination. However, as explained in section 4, this is not the case, so the covariance matrix should be corrected for under- or over-dispersion in the following way.

The chi-square test statistic for discrepancies between the four estimates of the relativistic drift rates, \( \chi_\nu^2 \), in each of the two directions is

\[
\chi_\nu^2 = \frac{1}{3} \sum_{i=1}^{4} \left( \frac{r_\nu - r_\nu^i}{\sigma_\nu^i} \right)^2
\]

(note that the denominator of the factor in front of the sum is the number of measurements minus unity). The corrected covariance matrix is then obtained from the covariance matrix (A2) by the formula:

\[
P_{\text{corr}} = \begin{pmatrix} a \chi_{\text{ns}}^2 & b \chi_{\text{ns}} \chi_{\text{we}} \\ b \chi_{\text{ns}} \chi_{\text{we}} & c \chi_{\text{we}}^2 \end{pmatrix}
\]

The total error must include also the spread of \( 10^4 \) estimates obtained in 10 sensitivity runs for each gyroscope, which is described by the corresponding covariance matrix \( P_{\text{sens}} \), systematic
errors (or their upper bounds) given by the matrix \( P_{\text{syst}} \) must also be added. Therefore, the total covariance matrix is:

\[
P_{\text{tot}} = P_{\text{corr}} + P_{\text{ens}} + P_{\text{syst}},
\]

three different sigmas represented by the diagonal elements of the three matrices on the right are thus summed up in squares.

References

[1] Schiff L I 1960 Possible new experimental test of general relativity theory Phys. Rev. Lett. 4 215–7
[2] Everitt C W F et al 2011 Gravity Probe B: final results of a space experiment to test general relativity Phys. Rev. Lett. 106 221101–14
[3] Silbergleit A S et al 2015 Gravity Probe B data analysis: I. Coordinate frames and analysis models Class. Quantum Grav. 32 224018
[4] Everitt C W F et al 2009 The Gravity Probe B data analysis Space Sci. Rev. 148 53–70
[5] Keiser G M et al 2009 Gravity Probe B Rev. Nuovo Cimento 32 555–89
[6] Everitt C W F et al 2015 The Gravity Probe B test of general relativity Class. Quantum Grav. 32 224001
[7] Heifetz M et al 2009 The Gravity Probe B data analysis filtering approach Space Sci. Rev. 148 411–28
[8] Silbergleit A S et al 2015 Gravity Probe B data analysis: II. Science data and their handling prior to the final analysis Class. Quantum Grav. 32 224019
[9] Silbergleit A et al 2009 Polhode motion, trapped flux, and the GP-B science data analysis Space Sci. Rev. 148 397–409
[10] Salomon M et al 2011 Nanohertz frequency determination for the Gravity Probe B HF SQUID signal Rev. Sci. Instrum. 82 125110
[11] Kozaczuk J A 2007 Precise determination of the spin speed and spin down rate of Gravity Probe B gyroscopes Physics Honors Thesis Stanford University
[12] Rose M E 1995 Elementary Theory of Angular Momentum (New York: Dover)
[13] Salomon M 2008 Properties of Gravity Probe B gyroscopes obtained from high-frequency SQUID signal PhD Thesis Dept. Aero/Astro Stanford University
[14] Bierman G J 1977 Factorization Methods for Discrete Sequential Estimation (New York: Dover)
[15] van der Merwe R 2004 Sigma-point Kalman filters for probabilistic inference in dynamic state-space models PhD Thesis OGI School of Science & Engineering at Oregon Health & Science University
[16] Levenberg K 1944 A method for the solution of certain nonlinear problems in least squares Q. Appl. Math. 2 164–8
[17] Conklin J W et al 2015 Precision attitude control of the Gravity Probe B satellite Class. Quantum Grav. 32 224015
[18] Silbergleit A S, Everitt C W F, Heifetz M I, Keiser G M and Turneaure J P 2003 Gravity Probe B error budget: preflight status. Gravitational radiation Proc. 27th Spanish Relativity Meeting ed J A Miralles, J A Font and J A Pons (Spain: Universitat d’Alacant, Alicante) pp 161–6
[19] Buchman S, Lipa J A, Keiser G M, Muhlfelder B and Turneaure J P 2015 The Gravity Probe B gyroscope Class. Quantum Grav. 32 224004
[20] Marcelja F, DeBra D B, Keiser G M and Turneaure J P 2015 Precision spheres for the Gravity Probe B experiment Class. Quantum Grav. 32 224007
[21] Wang S et al 2015 The design and performance of the Gravity Probe B telescope Class. Quantum Grav. 32 224008
[22] Li J, Keiser G M, Lockhart J M, Ohshima Y and Shestople P 2015 Timing system design and tests for the Gravity Probe B relativity mission Class. Quantum Grav. 32 224014
[23] Rainer M I et al 2012 VLBI for Gravity Probe B. V. Proper motion and parallax of the guide star, IM Pegas Ari Astrophys. J. Suppl. Ser. 201 5
[24] Shestople P, Ndili A, Hanuschak G and Small H 2015 Gravity Probe B orbit determination Class. Quantum Grav. 32 224017