A Sub-Millimeter Parallel-Plate Test of Gravity

Charles A. Hagedorn

A dissertation submitted in partial fulfillment of the requirements for the degree of:

Doctor of Philosophy

University of Washington

2015

Reading Committee:

Jens Gundlach, Chair

Eric Adelberger

Blayne Heckel

Ann Nelson

Program Authorized to Offer Degree:

Department of Physics

University of Washington

Abstract

A Sub-Millimeter Parallel-Plate Test of Gravity

Charles A. Hagedorn

Chair of the Supervisory Committee:

Professor Jens H. Gundlach

Department of Physics

Newton's inverse-square law of gravitation is the oldest standing mathematical description of a fundamental interaction. Experimental tests of gravity's distance-dependence define a frontier between our understanding of gravity and many proposed forms of new physics. We performed a parallel-plate test of the gravitational inverse-square law at sub-millimeter distances. Using a tantalum/titanium torsion balance locked in electrostatic feedback, we performed a null search for deviations from uniformity of the gravitational field of a planar tantalum/aluminum mass. Key innovations include an interferometric isolating-foil position monitor, a completely reproducible analysis tool-chain, and a publicly-unblinded analysis. Using the traditional Yukawa parametrization, our results are consistent with Newton's inverse-square law at $\alpha = 1$ over distances from λ =104 μm to 1320 μm with 95% confidence.ou

Listing 1: Version Information

```
commit 8e15b1f02f059a74e5d42aa7714dd6183db683e5
Author: charlie <charlie@dicke.nanoradian.com>
Date:   Thu Jun 18 15:39:58 2015 -0700
    Comma added!
    This document's analysis was generated with commit
        5840404e3586d9b657a2da835a19b7b0ce24949a. Commits between then and now
        contain additions and edits for clarity and compliance with thesis
        guidelines. It was brought up to date with this commit by hand in order to
        meet the thesis-submission deadline :).
```

Dedication

To you.

This was written for you, that you might learn from our successes and from our errors.

This thesis is a companion to the computer software that automatically generated it. If you have unanswered questions about how most quantities and plots were determined, you'll find more answers there.

The source code is freely available: https://www.github.com/4kbt/PlateWash/ .

If you find an error, or have further questions, please let me know.

Acknowledgements

Susan Ashlock gets special mention, as she has given me endless support with love, great patience, and myriad wonderful trips into the mountains over the latter half of this work. I can't wait to explore the world with her now that this thesis is complete. Furthermore, this measurement's approaches of source-code management, continuous analysis builds, and unit testing were inspired by looking over her software-engineer shoulders. Without Susan's professional influence, this blind measurement would not have been possible.

I thank the entire Eöt-Wash group, the National Science Foundation, and the United States taxpayer for unwavering support, even when a blitzkrieg experiment turned to a protracted and careful campaign.

When a worthwhile project takes a decade, you'll find that a village full of wonderful people have helped along the way. What follows is an incomplete list of the many people and organizations who have made significant direct or indirect contributions to this measurement. Those CENPA members highlighted in bold worked directly on the instrument or on development work for the instrument. Those highlighted in italics contributed physics insight directly to the measurement or its analysis.

Alfred Hagedorn II, Alfred Hagedorn III, Alice Crawford, Allen Myers, Ann Nelson, Barb Fulton, *Blayne Heckel*, Bob Shupe, Brian Balfrey, Catherine Provost, CENPA, Charles Rogers, *Claire Cramer*, *Dan Kapner*, David Hyde, David Kettler, *David Newell*, David Peterson, Debra Nastaj, Don Hartill, Doug Will, Eberhard Bodenschatz, *Eric Adelberger*, George Hagedorn, Greg Harper, Hamish Robertson, **Hank Simons**, James Napolitano, *Jarek Kaspar*, *Jason Detwiler*, ***Jens Gundlach***, **Jim Elms**, John Amsbaugh, *John Greendeer-Lee*, Jon McCoy, Josh Long, Juliet Lovejoy, Karen Daniels, Kate Higgins, Keath Marx, **Kipp Krause**, Ki Young Choi, ***Krishna Venkateswara***, *Kurt Snover*, Laura Bodine, Margaret Roston, Mark Freeman, Mary Ellen Jones, *Matt Dietrich*, *Matt Sternberg,* ***Matt Turner***, Michael Chen, Michael Famulare, Michelle Wang, MicroE, Monday Meeting, Myrna Hagedorn, Nichole Fernkes, *Nikolai Tolich*, OGP Smartscopes, Omega Instruments, *Peter Nelson*, **Prashant Emani**, Raman Sundrum, Rembar, Richard Wiener, Sarah Dasher, ***Scott Pollack***, Scott Schell, Shawn Tanner, Silas Wild, Sky Sjue, ***Stephan Schlamminger***, *Svenja Fleischer*, Susan Ashlock, Tanja Bode, *Ted Cook*, Terry Lynn Helms, Tim Van Wechel, *Todd Wagner*, Tracy Lovejoy, Victoria Clarkson, and *Will Terrano.*

Contents

Chapter 1

Motivation and Background

1.1 Introduction

Four fundamental interactions are known to Physics. In historical order of characterization, they are gravity, which governs planetary motions; electromagnetism, which describes the interactions of electrically charged objects, and the weak force and strong forces, which govern nuclear and sub-nuclear dynamics. The electric and weak forces were unified theoretically by Glashow, Salam, and Weinberg as the electroweak interaction in 1967 [1]. The strong force is described well by quantum chromodynamics. The resulting theory of the electroweak and the strong, precisely tested, finely tuned, and quantum in nature, is known as the Standard Model. Einstein's theory of gravity, general relativity, is consistent with all precision tests, but it contains no known links to the Standard Model. For interactions with sufficiently high energy density, one or both theories must fail, as they predict different phenomena.

The Standard Model is a triumph. It is a complex theory, with 18 or more parameters (particle masses, coupling constants, and mixing parameters) which are tuned to fit observation. The Standard Model had one major remaining prediction - the existence of the 'Higgs boson', a lynchpin for the theory, which was discovered by the LHC in 2012 [2, 3]. After tuning, the Model has predictive power, correctly recovering within experimental and theoretical error the known zoo of observed particles, resonances, and phenomena, with very few exceptions. The exceptions are, therefore, important.

Observations which the Standard Model cannot explain:

- Gravity: There is no tree-level coupling between uncharged objects in the Standard Model. Furthermore, on a charge for charge basis, the interactions in the Standard Model are $> 10^{40}$ times stronger than gravity, perhaps hiding signatures of a unifying theory. The theory of gravity, General Relativity,

is well-tested and in agreement with experiment [4], but does not contain any hint of quantum probability or uncertainty. Our theories of gravity and quantum mechanics are mathematically incompatible.

- The Cosmological Constant Problem/Dark Energy: Astronomical observations suggest that the universe is expanding, and doing so at a rate which is increasing. The phenomenon can be explained by proposing the existence of "Dark Energy", an energy density associated with empty space that has negative pressure. The energy density of vacuum has these features, but, as described by the Standard Model, is 10^{120} times too large [5].

- Dark Matter: Beginning with the study of galactic rotation curves, a compelling body of observations point toward the existence of a novel form of matter, perhaps a new family of particles, that, at accessible temperatures, couples to ordinary matter only gravitationally [6, 7].

- There are perpetual hints of new physics waiting to be discovered. As an important example, the best measurement of the anomalous magnetic moment of the muon disagrees with the best theoretical calculations at the 3σ level [8]. An improved experiment is under construction at Fermilab to address this concern by the end of the decade [9].

Furthermore, the angular power spectra of the cosmic microwave background, along with detailed studies of nucleosynthesis, are highly compatible with cosmology theory that includes dark energy and cold dark matter and are essentially incompatible with theories that do not [10, 11]. These measurements suggest a dark energy density of 4 keV/cm^3, which, when converted to a length using $\hbar$ and c, is roughly 86 micrometers.

1.2 The inverse square law

$$\vec{F}_{21} = -G\frac{m_1 m_2}{|\vec{r}_{21}|^2}\hat{r}_{21}$$

Newton proposed the gravitational inverse square law in 1687 in order to explain the motions of celestial bodies [12]. He suggested that the gravitational force between two objects is always attractive, is proportional to the mass of each object, and inversely proportional to the square of the distance between them. This model was remarkably good at predicting the motion of celestial orbits, the tides, and gave the first quantitative estimate for the mass of the Earth.

The inverse square law (ISL) for gravity has been subjected to many experimental and astronomical tests. The observed stability of elliptical orbits and the apparent validity of Kepler's laws gave immediate support to the theory at solar-system scales. The first laboratory test of the ISL was performed by Cavendish and

Michell in 1798 [13] in support of a measurement of the density of the Earth. See a fine review by Fischbach for the history of these experiments [14], and a detailed review of short range experiments by Adelberger, Heckel, and Nelson [15]. Current tests of the inverse square law are underway at the University of Washington [16, 17], Huazhong University [18, 19], University of Indiana [20], Washington University at St. Louis [21], Stanford University [22], Physikalish-Technische Bundesanstalt (PTB) [23], Apache Point Observatory [24], the University of Maryland [25], Grenoble [26], Birmingham University [27], and Düsseldorf [28].

Early tests of the inverse square law parametrized possible inverse square law violations as

$$F_G \propto \frac{1}{r^{2+\epsilon}},$$

and set bounds on the size of ϵ. This parametrization is hard to reconcile with modern theory, as few, if any, viable theories propose fractional dimensions at all scales.

In modern work, proposed inverse square law violations are parametrized as forces mediated by an exchange particle. For both scalar and vector particles [14], the modified potential takes the approximate form

$$V(r) = -G\frac{m_1 m_2}{r}(1 + \alpha e^{-r/\lambda}), \tag{1.1}$$

where α is the new interaction's strength relative to gravity and λ is the interaction's characteristic length (reduced Compton wavelength of a mediating particle with mass m. $\lambda = \hbar/mc$ [29]). This is not the only possible modified potential, but it is general and has become a traditional parametrization for inverse square law violations.

1.3 Theories

Here, we present the leading theoretical motivations for testing the inverse square law at short distance. There are many more proposed theories which are not covered here. In general, the theoretical guidance to experimentalists is clear: measurements at shorter ranges and with higher sensitivity are valuable. Particular emphasis is placed on shrinking the shortest range at which experiments can see Newtonian gravity itself.

As experimentalists, we take particular joy in testing theories like the 'fat-graviton' theory described in Section 1.3.4, which prescribes both longest and shortest distances over which we must test, making it completely-falsifiable.

1.3.1 String Theory

String theory provided the catalyst for a resurgence of interest in the inverse square law at sub-millimeter scales in the late 1990s. At the cost of as many as seven extra dimensions, the theory offers a promise of unifying gravity with the other forces. The extra dimensions are hidden away from normal observation with "compactification", introducing a length-scale below which extra dimensions are relevant to particle interactions and above which any effects of extra dimensions are suppressed. Critically, the effects of gravity would be sensitive to the change in dimensionality by obeying Gauss' Law at all scales - diluting gravity into the many extra dimensions at short distances, while diluting into only three spatial dimensions at longer scales. This motivates the comparatively low strength of gravity in the three-dimensional macroscopic world. This scenario predicts that gravity is significantly more attractive at scales shorter than the compactification scale.

If parametrized as a particle-exchange theory, string theory predicts two new types of exchange particles, known as the dilaton and the radion. Both would appear gravitationally as Yukawa violations of the inverse square law at scales comparable to the compactification length. The dilaton is responsible for the familiar Gauss' law increase in strength of gravity, while the radion is associated with excitations of the periodic spacetime normally accorded to the compactified dimensions.

In general, the compactification scales can be as small as the Planck scale, $\sim 10^{-35}$ m. It was shown in 1998 by Arkani-Hamed, Dimopoulos, and Dvali that, using the then-current limits on the gravitational inverse square law, the compactification scale of one or more of these extra dimensions could be as large as a millimeter or more [30, 31]. 2007 limits from Kapner *et al.* [32] place a 95% confidence upper bound at 44 μm for one large extra dimension. Greater numbers of extra dimensions are more stringently constrained by collider experiments [33] at scales below 40 μm [34].

1.3.2 Dark Energy

At present, the dominant motivation for sub-millimeter gravity tests comes from the physics of "dark energy". The universe has been observed to be expanding through redshift, CMB, structure, and other measurements. Not only is the expansion occurring, but the expansion rate is observed to be accelerating throughout the visible historical record in astronomical observations. The expansion and changes in its rate can be roughly explained if there is an energy density Λ associated with a volume of spacetime and if this new kind of energy has negative pressure.

If the measurements of the Hubble Constant (WMAP: [35], Planck: [11, 10], Supernova searches) are used to extract a value for Λ, it is found to be ~ 4 keV/cm^3. If this is converted to a length using Planck's

constant and the speed of light,

$$\lambda = \sqrt[4]{\frac{\hbar c}{\Lambda}} = \sqrt[4]{\frac{8\pi G\hbar}{3\Omega_\lambda H_0^2 c}} \approx 86\,\mu\text{m},$$

where $\Omega_\lambda$is the dark energy fraction of the closure density of the universe ($\sim 70\%$) and H_0 is the Hubble constant (~ 70 km/s/Mpc$= 2 \times 10^{-18}\,\text{s}^{-1}$) [10]. If there is new gravitational physics associated with Dark Energy, we should expect interesting behavior at this approximate scale.

1.3.3 Cosmological Constant

Closely related to the dark energy problem is an older problem in physics. In general relativity, there is a free parameter, traditionally denoted as Λ, which plays a similar role as the aforementioned dark-energy Λ in the evolution of spacetime [36]. Roughly, it is the energy density associated with a volume of space. If computed from modern particle theory, the expectation is that $\Lambda_{ParticleTheory} \sim 10^{120}\Lambda_{Observed}$ [5]. Without a physically-motivated infrared cutoff to the necessary integrals, there is no known explanation for this discrepancy.

Any deviation, particularly a weakening, of gravity at short range might provide the clues needed to explain this major theoretical problem. If the dark energy is a cosmological constant, then the motivation for sub-millimeter gravitational experiments is even stronger.

1.3.4 Fat Graviton

Raman Sundrum showed [37] that the Cosmological Constant problem might be resolved if gravity were to become significantly weaker at scales below 100 μm. His "fat graviton" modification of gravity is remarkable because he predicts that it must occur at scales greater than 20 microns. Few theoretical predictions offer the possibility of complete falsification, particularly within reasonably-accessible parts of parameter space. Experimental attempts to constrain the interaction [38] have adopted the functional form $F_{12}(r) = -G\frac{m_1 m_2}{r^2}\left[1 - e^{-\left(0.914\frac{r}{l_g}\right)^3}\right]$ after Figure 8 in Sundrum's paper. Present bounds on l_g are $20\mu\text{m} \lesssim l_g \leq 95\mu\text{m}$, by naturalness and Adelberger *et al.* [38] respectively.

1.3.5 Chameleons and Screened interactions

Khoury and Weltman [39, 40] in 2004, and Feldman and Walsh with Nelson, [41, 42] in 2006, and 2008 respectively, opened a new direction of inquiry. They showed that the addition of a self-interacting term to the Lagrangian governing an otherwise long-range force can hide it from long range experimental searches. Self interactions act to "screen" any new interaction, this self-hiding ability gives these interactions the

name "Chameleons". As self-interaction is a general property of fields, the notion is very reasonable. In its simplest form as a scalar field, a chameleon interaction must be constrained by limiting not only the exchange particle's mass m but also the self interaction strength β. Because chameleon interactions are self-screening, the screening length becomes the relevant length scale at which the new interaction may emerge from camouflage. The possibility of self-interaction-based screening opens up considerable theoretical freedom for new theories of dark energy and dark matter in the face of existing experimental constraints.

Chameleon interactions are ready-made to evade traditional inverse square law tests. All precision tests of short-range gravity at gravitational strength to date, including our experiment, place a metal foil between the source and detector masses to suppress non-gravitational interactions. Because of the self-screening within materials the foil significantly diminishes the apparent strength of any chameleon interaction. When searching for chameleons, isolating shields must be made as thin and of as low density a material as possible, or dispensed with entirely.

Because of the self-interacting nature of chameleons, existing theoretical methods have limited regions of parameter space over which the effects of chameleons can be reliably predicted. The remaining calculable region is now severely constrained, and current-generation experiments may exclude it entirely [43].

Self-interacting theories are not the only proposed interactions which exhibit screening behavior. Other scalars, such as "Symmetrons" and "Galileons" achieve similar screening in regions of high density. Symmetrons screen through the spontaneous restoration of a symmetry in regions of high density; Upadhye published an extremely helpful calculation of the expected torques in a torsion balance [44]. Galileons are scalars that appear in massive-gravity theories that obey a certain shift symmetry; they become strongly self-interacting in regions of high density. For in-depth review of these interactions, we refer to these reviews [45, 46]. Fortunately for experimentalists, the general predictions of these theories are the same: an apparent anomalous force at short range.

Our experiment's parallel-plate geometry is of particular use in these screened-interaction classes of theories, as calculations which require challenging numerical models for hole- and wedge-pendulum geometries can be more straightforward with planar attractor and detector masses, particularly in an infinite-sheet approximation.

1.3.6 Unparticles

Georgi kicked off a flurry of theoretical work in 2007 by noting the possibility of interactions with novel dispersion relations, with a scale-invariant particle propagator instead of the usual $\frac{1}{k^2+m^2}$ [47, 48]. In particular, the properties of an interaction might be scale-invariant without requiring that the propagating

particle have zero mass. Unparticle interactions could appear in tests of short range gravity [49, 50]. The generic signature of a new unparticle interaction is an additional power-law-potential fifth force; it does not alter the gravitational potential to be $V_G \propto 1/r^{1+\epsilon}$, but rather $V_G \propto (1/r + C(R_G/r)^{2d_I - 1})$, where C is a dimensionless constant (weakly dependent on d_u), R_G is a characteristic scale, and d_u is the potentially-fractional "anomalous unparticle dimension". The existing body of inverse square law tests place tight constraints on these interactions. Indeed, the first publication on unparticles appeared in March 2007, and these constraints, from other theorists, were in place by August. The tight constraints on new theory that stem from precision tests of gravity are the guaranteed payoff that come with incremental improvement in the laboratory.

1.4 Survey of Previous Experimental Work

Tests of gravity using test bodies with known masses have now spanned about thirty orders of magnitude in spatial extent. Gravity itself has been observed to be consistent with an inverse square law over the roughly sixteen orders of magnitude from the solar system scale down to a few tens of micrometers. Useful reviews are more comprehensive than will be presented here [15, 34, 51]. Of timely note is [52], which features a comprehensive and uniform accounting of the field's modern history.

Precision measurements of the distance dependence of gravity may have begun with Kepler. Kepler's Laws [53, 54] are now taught to students as necessary corollaries of Newton's theory, but Kepler's observations were essential to the formulation of the theory itself. Astronomical tests of gravity remain the most precise, providing direct tests of General Relativity in addition to Newton's Laws. In order of decreasing range and increasing precision, the leading solar system tests of the inverse square law are the Pioneer spacecraft [55], planetary orbits [56, 14], and the very-precise lunar laser ranging [24, 57]. The lunar results are the most precise tests of gravity at any scale, limited by uncertainties in the astronomical modeling code for solar-system dynamics. The experimental data are consistent with the inverse square law at better than the 10^{-10} level [15] over lengthscales comparable to the lunar orbital radius of $\sim 4 \times 10^8$ m.

In the laboratory, direct observations began with Michell and Cavendish's torsion balance in 1798 [13], observing gravity at scales of tens of centimeters to meters. Successive improvements on that method and tests with improved gravimeters have extended tests of gravity out to scales of the size of Earth [14].

At scales of meters and below, tests of the inverse square law become more specialized. Experimental designs become more and more carefully crafted to suppress systematics that would otherwise hide the signature of gravity entirely. Prior to 2001, the shortest distances at which gravity had been observed was $\sim$ 0.5-1 cm, by Spero *et al.* in 1980 [58] and $\sim 800 - 900$ μm by Mitrofanov and Ponomareva in 1988 [**?**

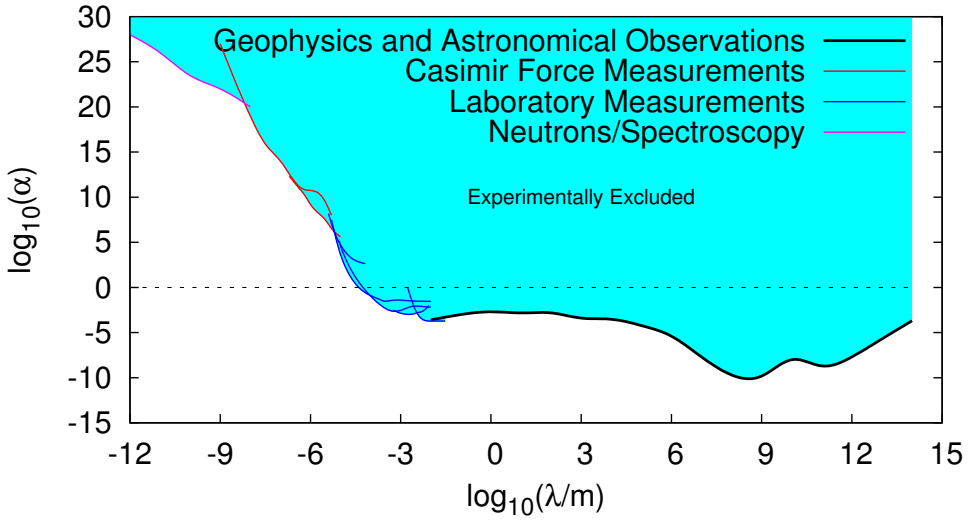

Figure 1.1: ISL existing limits

]. A novel superconducting Gauss' Law test by Moody and Paik in 1994 remains the best published limit for scales of about a meter [59], though an improved result from Boynton, Newman, and collaborators is expected soon [17].

Initially inspired by the suggestions of large extra dimensions, the Eöt-Wash group at the University of Washington began its series of four horizontal torsion balance experiments, yielding limits at $\alpha = 1$ of $\lambda = 218\,\mu$m [60] (2001), $\lambda = 197\ \mu$m [61] (2004) , $\lambda = 56\,\mu$m [32] (2007), and $\lambda = 42\ \mu$m [16] (2013). In parallel, a 1 kHz tungsten oscillator experiment at the University of Colorado achieved their maximum sensitivity at $\alpha \sim 5$, $\lambda \sim 100\ \mu$m [62] (2003). An atomic force microscope-inspired series of experiments at Stanford have provided the tightest constraints to date on gravity at scales below 10 μm with publications in 2005 [22] and 2008 [63, 64]. Most recently, there has been a new experiment from Huazhong University using a vertical torsion balance geometry that set an $\alpha = 1$ limit at $\lambda = 76\,\mu$m [18]. A cryogenic experiment at the University of Maryland with a superconducting isolating shield has reached scales of $\lambda \sim 120\,\mu$m [25]. Experiments at Cornell [65], Dusseldorf [66], the PTB [23], and Humboldt State [67] have not yet reported exclusion limits.

At distances shorter than $\sim 5\,\mu$m and longer than ~ 6 nm, the best limits on gravity come from measurements of the Casimir force [68]. Any disagreement between observations of the Casimir force and the Casimir theory could possibly be due to new physics. So far, no clear deviations have been observed, so

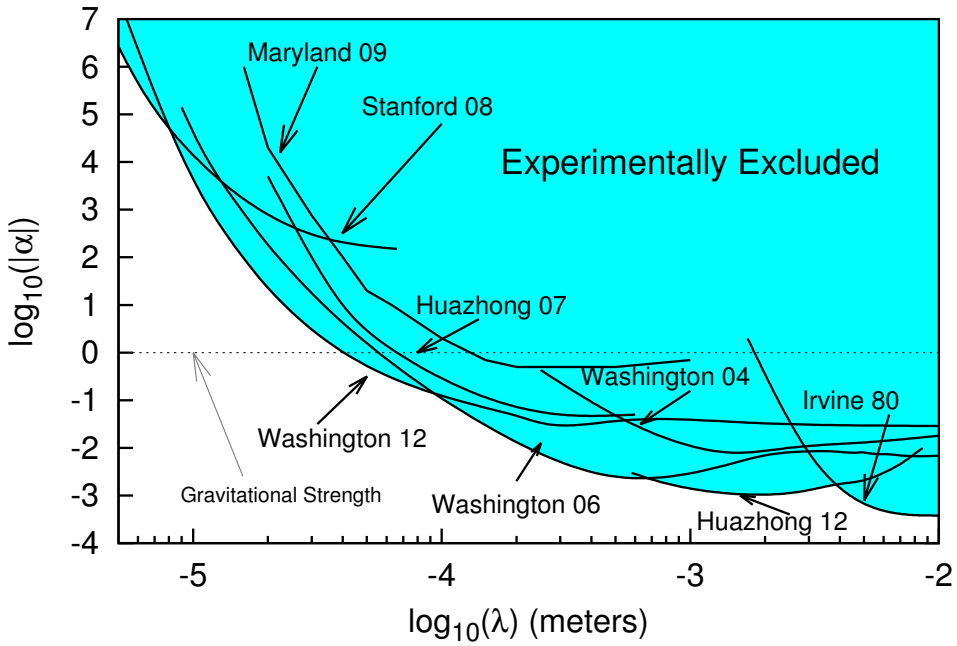

Figure 1.2: Laboratory Tests of Gravity

the measurement uncertainties are traditionally converted into exclusion limits for gravity. A new variation on the Casimir technique will use ultra-cold Strontium spectroscopy to test theoretical predictions of the Van der Waals interaction [69]. Other experiments at these scales include ultra-cold neutrons "bouncing" quantum-mechanically on tabletops [26, 34]; while these experiments are less sensitive to gravity, they are completely insensitive to any confounding electromagnetic Casimir background. A very different experiment attempts to place limits on new interactions by searching for on-shell bosons from the sun [70]. At still shorter scales, limits are collider-based, restricted by the non-observation of unexpected physics, beyond the scope of this thesis.

Gravitational tests for micron-scale extra dimensions with astrophysical sources have been proposed by Simonetti *et al.* [71, 72].

1.5 Experimenter's motivation

If the theoretical arguments in favor of short-range gravity tests are insufficient, perhaps this will suffice: Pick up a caliper and dial it to 50 μm. If no caliper is handy, pluck a hair from your head and consider its width. We know almost nothing about gravity on that scale. For every other known interaction of nature, our understanding reaches far below the diameter of a proton. That's more than 10,000,000,000 times smaller.

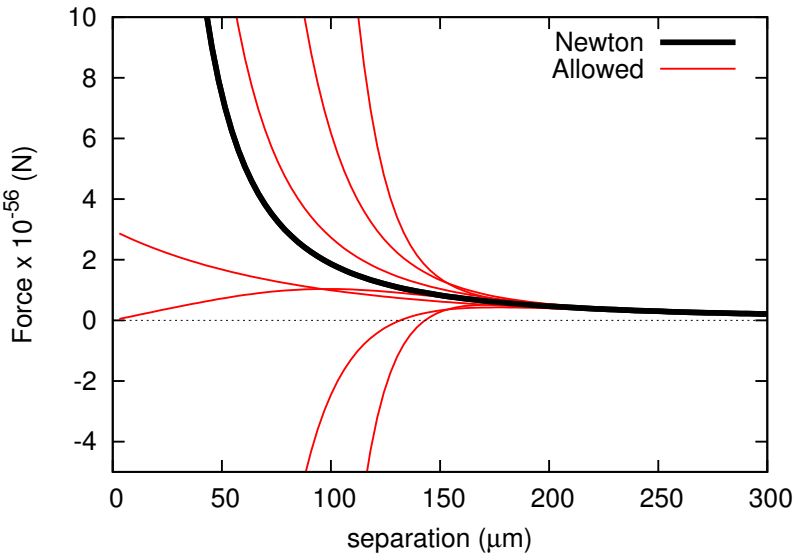

Figure 1.3: A kindergartener's perspective: the inverse square law, with allowed possibilities plotted. Force laws are drawn from the edge of the Yukawa and "fat graviton" constraints by Kapner [32, 38].

At scales of a micron, gravity could be either attractive or repulsive and 10^{10} times stronger than Newton predicts. We don't know.

Chapter 2

Apparatus

2.1 Principle - Gauss' Law Test

It is easily shown that, in three dimensions, any force that obeys a $1/r$ potential satisfies Gauss' law

$$-Gm \oiint_S \frac{\vec{r}}{|r|^3} \cdot d\hat{r} = -4\pi Gm.$$

The converse is true, too: If a force law satisfies Gauss' Law, then the field flux through a sphere centered on a point charge must be constant for all sphere radii. As the surface area of a three-dimensional sphere in three spatial dimensions is proportional to the square of its radius, then the field strength must fall as $1/r^2$.

Our experiment is a test of the uniformity of the gravitational field of an "infinite" sheet of mass. It is in this sense that our experiment is a "Gauss' law test of gravity." For any force in three spatial dimensions, the field from an infinite uniform two-dimensional sheet of charge must, by symmetry, be perpendicular to the plane of the sheet. Furthermore, if the interaction obeys Gauss' law, the flux of field remains constant with increasing distance from the sheet. Therefore, the strength of the interaction cannot vary with distance. This fact immediately suggests a simple test for whether an interaction follows Gauss' Law: Measure the force acting upon a test charge at two different distances from a sheet of charge. If the two measured forces are not equal, then the interaction does not obey Gauss' Law. Such an experiment is a "null-test" for any deviation from an inverse square law. Colloquially, it might be referred to as a "Does-the-Inverse-Square-Law-Break-O-Meter".

In idealized form, our experiment compares the gravitational field strength at two distances from an infinite sheet of mass. Finite sheet masses are suspended on opposite arms of a torsion pendulum. The sheet masses are aligned parallel to their moment arms, but displaced in opposite directions perpendicular to the

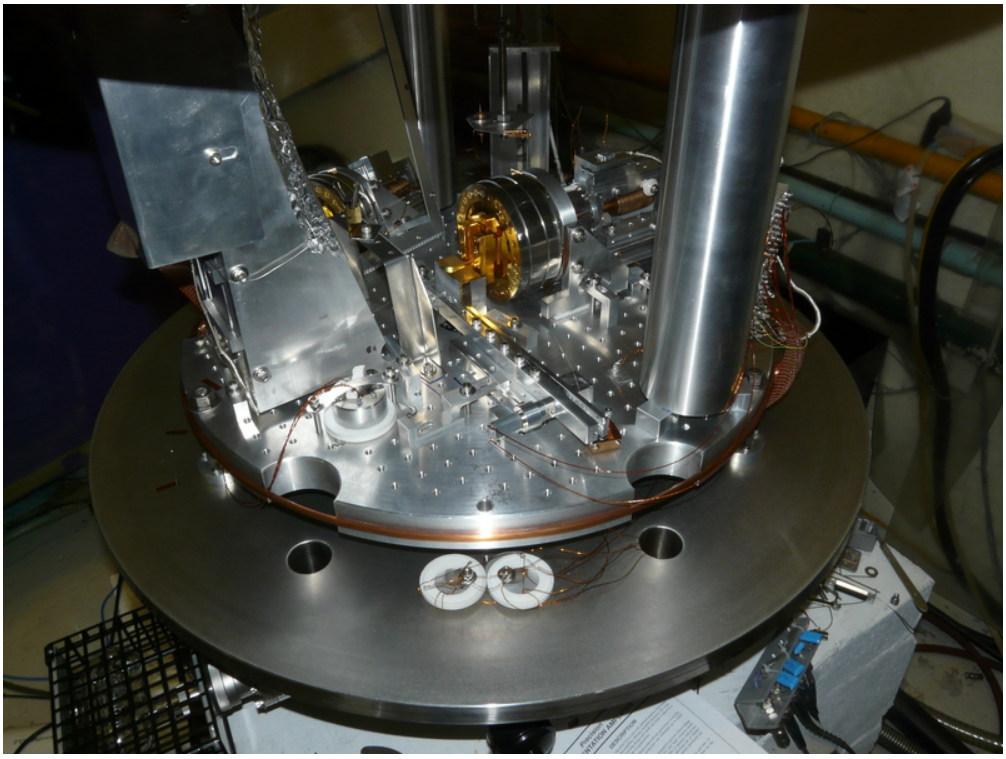

Figure 2.1: The PlateWash Experiment. Last photo before closing for science run, December 21, 2011.

arms (see Figure 2.2). If there is an inverse-square-law violating force between the sheets on the pendulum and an infinite sheet of matter, the pendulum will feel a torque. If the distance between the pendulum and the infinite sheet is modulated, the torque will vary in sync with the modulation. If there are no inverse-square law-violating interactions between the pendulum and sheet, the pendulum will experience *no torque change* as the infinite sheet is translated. The difficulties in the measurement primarily involve ensuring that the "infinite sheet" approximation is valid, that the mechanism moving the sheet adds negligible gravitational interaction, and that gravity (and any new interaction) is the only messenger between the sheet and the pendulum.

2.1.1 Implementation

Before launching into the full description of the apparatus and method, a qualitative description and definitions of nomenclature will provide context for the full description to come. A more-accurate, but still schematic, diagram of the apparatus is shown in Figure 2.3.

- Our laboratory at the University of Washington is named after Baron von Eötvös, an important early torsion-balance innovator. Each experiment in our Eöt-Wash laboratory has a "-Wash" moniker. To highlight the parallel-plate nature of the design, and for levity, the experiment is often referred to as

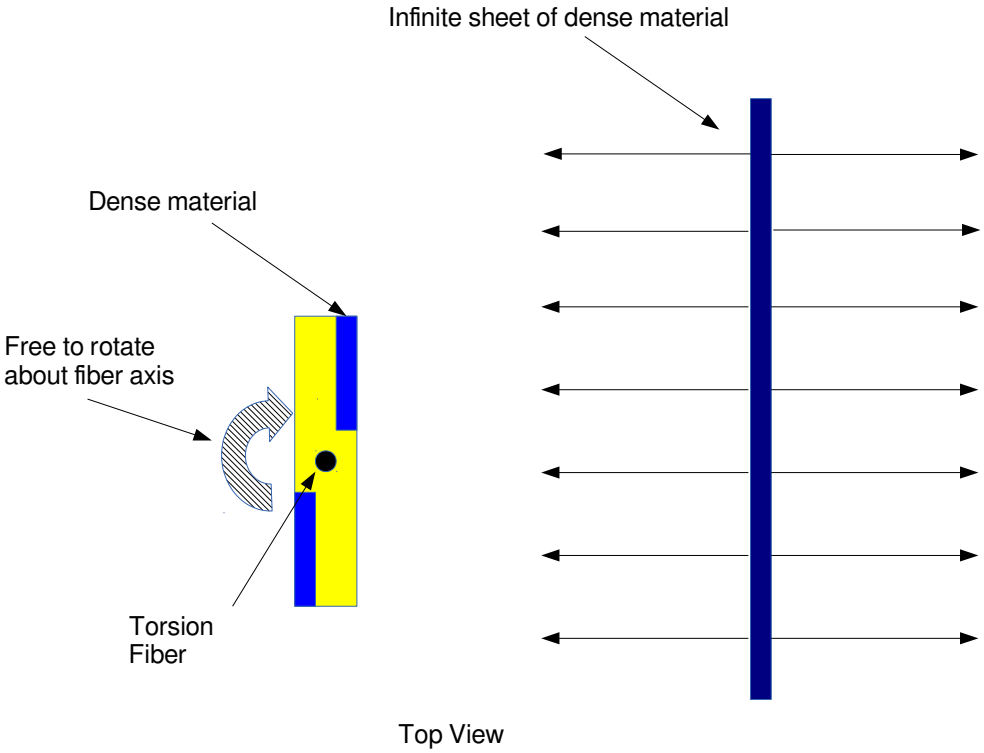

Figure 2.2: Simple Apparatus Diagram

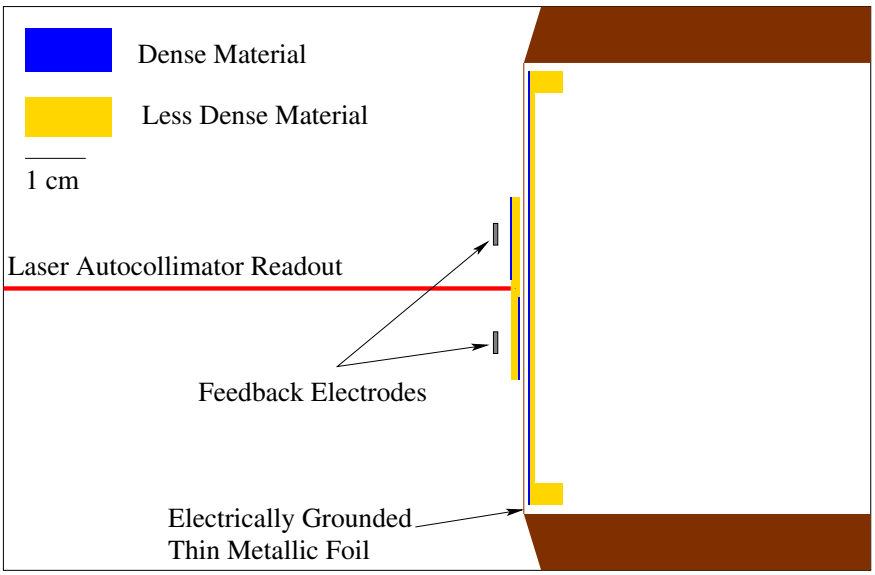

Figure 2.3: Experimental implementation schematic. The only not-to-scale dimensions are the pendulum-foil separation and the attractor-foil separation; they are smaller in reality.

"PlateWash"

- Throughout this thesis the moving-sheet field mass is referred to as the "attractor".
- The first line of defense against any non-gravitational interaction between the attractor and the pendulum is a tightly-stretched metal foil. It acts as a Faraday shield against electrostatic interaction, blocks all electromagnetic interaction at frequencies above ~ 30 MHz, blocks residual gas interaction, and provides physical isolation between the attractor and pendulum.
- The infinite-sheet attractor must become finite; the addition of a rim to the circular attractor can replace many of the gravitational forces that would be present with an infinite sheet. In this iteration of the experiment, the rim is too-large, leading to a small residual gravitational signal that is very linear in attractor position.
- The rotation of the pendulum is measured by an autocollimating optical lever, an autocollimator. Typical sensitivities of our group's autocollimators can be as small as fractions of a nanoradian in a second in the millihertz range [73]. Our prototype designs reach tens of picoradians in a second at higher frequencies [74]. The autocollimator used here can resolve rotations of ~ 30 nrad in a second in the relevant frequency range.
- The pendulum is held in electrostatic feedback: the electrostatic interactions between the pendulum and the foil are far stronger than the torsion-fiber's stiffness, making feedback a necessity. The torque signal is extracted from the feedback loop. The signal strength is continuously calibrated using a separate rotating gravitational attractor.
- The attractor is moved toward and away from the pendulum in a square-wave pattern. "Dead time" is declared during each transition between positions to avoid motion-dependent systematic effects. Groups of ten changes in position are lumped into a "cut" and used for further analysis. The experiment was operated with both per-cut randomized start/stop locations and continuous square-wave motion.
- Great attention was paid to systematic effects. As any foil motion could lead to a parasitic signal, a Fizeau fiber interferometer was built to monitor the foil position. We believe that this is the first inverse-square-law measurement to have a dedicated foil monitor.
- For the primary goal of constraining Yukawa-like violations of the inverse square law, the data analysis makes a conservative choice to treat all systematic effects as Yukawa interactions, with independent interaction strengths and associated lengthscales.

- While the measurement was begun before blinding, the experimental analysis has been performed blind. Early in the process of developing the analysis, the true experimental results were accidentally viewed twice. So many critical parts of the data analysis have changed since that time that, at the time of unblinding, we have no idea what the outcome will be. While the experiment cannot be regarded as truly blind, it has brought our blind-measurement experience up from nothing and directly leads the way to a future measurement.

- In order to facilitate the unblinding of this complex analysis, and to guarantee reproducibility, the entire analysis and the creation of this document are entirely automated. A collection of free software, the source-code repository, and the experimental data are all that are required to re-generate this result. The majority of the numerical quantities stated in this document are automatically either extracted from the configuration files used to initialize the data analysis or are products of the analysis itself. These quantities are either stated in accordance with standard protocol for significant digits or with their associated uncertainties.

- Throughout, we use torque units of fN-m, as the signals we seek are at that order. The smallest signal of interest is approximately ~ 0.1 fN-m. This leads to the rare automated statement of a torque like $(-1175539012 \pm 170855801)$ fN-m/m, where the point is that the signal is large and resolved, not that the uncertainty should be taken seriously beyond the leading two digits. These few oddities occur only where approximate statements are appropriate. Our software libraries properly handle such quantities correctly in scientific notation, but would lose the reader's visceral connection to the natural scale of 1 fN. Handling these corner-cases correctly would add significant complexity, and thus opportunity for error, to our software.

2.2 Expected Signal

2.2.1 Flat plate force

The form of possible inverse square law violations is not known from first principles. However, under the assumption that a new interaction can be modeled by a Yukawa potential (Equation 1.1), the new force between flat plates is calculable. Neglecting edge effects, if one plate is infinite in spatial extent, the other has area A, and the plates have densities ρ_1 and ρ_2, thicknesses t_1 and t_2, and face-to-face separation d, the Yukawa force F_Y is

$$F_Y = 2\pi G \rho_1 \rho_2 A \alpha \lambda^2 \left(1 - e^{-t_1/\lambda}\right) \left(1 - e^{-t_2/\lambda}\right) e^{-d/\lambda} \tag{2.1}$$

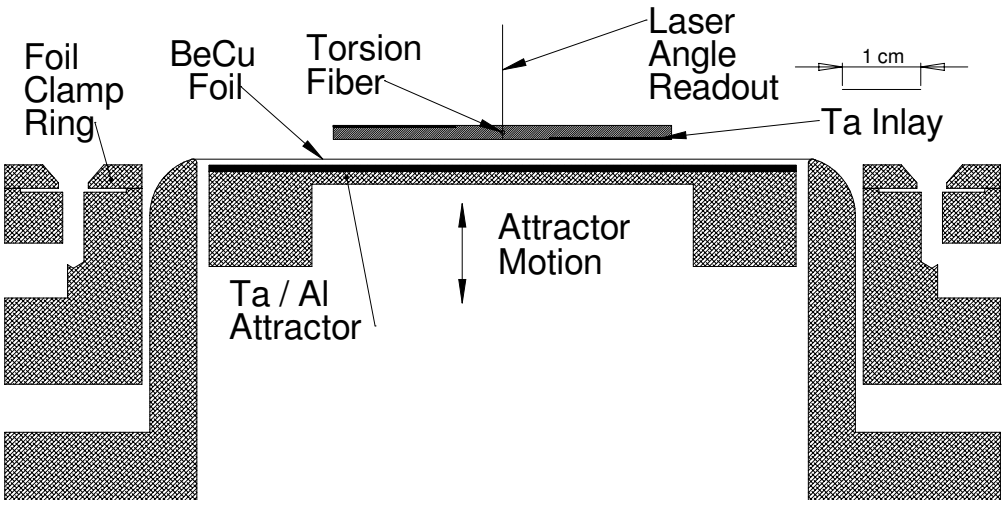

Figure 2.4: Pendulum/Attractor schematic. The image is a top view of a horizontal slice through the center of the pendulum, foil, and attractor. The separation between pendulum, foil, and attractor have been exaggerated for clarity. All else is to-scale. The pendulum and attractor are properly dimensioned. The foil-holder shown is an earlier design, but functionally equivalent to the one used in the experiment. The actual foil clamping ring is < 1.5 mm in thickness.

When the plate size, A, is large compared to λ, the fractional error from end effects should be roughly $\sim 4\lambda\sqrt{A}/A = 4\lambda/\sqrt{A}$. In our experiment, with a parallel orientation, this correction spans 0.1-10% (λ from 10-1000 μm), and always acts to increase the magnitude of the signal.

When $\lambda \to \infty$, the Yukawa interaction is indistinguishable from an altered G. $F_Y \to 2\pi G\rho_1\rho_2 A\alpha t_1 t_2$.

2.2.2 Flat plate torque

For a pendulum of height h, width w, and center-to-step distance j, pendulum body density ρ_l, pendulum inlay density ρ_h, the Yukawa torque on one inlay, neglecting factors from the attractor, is:

$$\tau = \pi\alpha\rho_a G h \left(\frac{w^2}{4} - j^2\right)(\rho_h - \rho_l)\,\lambda^2 e^{-\frac{d}{\lambda}}$$

The torque due to a single attractor sheet of density ρ_a, thickness t_a, and a pendulum inlay thickness t_I, is $\tau_{one} = \rho_a\tau\left(1 - e^{-t_a/\lambda}\right)\left(1 - e^{-t_I/\lambda}\right)$

$$\tau_{one}(x) = \pi\alpha G h \left(\frac{w^2}{4} - j^2\right)(\rho_h - \rho_l)\,\rho_a\lambda^2 e^{-\frac{d}{\lambda}}\left(1 - e^{-t_a/\lambda}\right)\left(1 - e^{-t_I/\lambda}\right)$$

Taking the difference between the inlay torques, with a pendulum thickness t_p, $\tau_{full} = \tau_{one}(x) - \tau_{one}(x + t_p - t_I)$

$$\tau_{full} = \pi\alpha G h \left(\frac{w^2}{4} - j^2\right)(\rho_h - \rho_l)\,\rho_a \lambda^2 \left(e^{-d/\lambda} - e^{-(d+t_p-t_I)/\lambda}\right)\left(1 - e^{-t_a/\lambda}\right)\left(1 - e^{-t_I/\lambda}\right) \tag{2.2}$$

With an attractor stroke (the difference between close and far positions) s (and now d becomes the pendulum-face-to-attractor-face distance of closest approach),

$$\begin{aligned}\Delta\tau_{full} &= \pi\alpha G h \left(\frac{w^2}{4} - j^2\right)(\rho_h - \rho_l)\,\rho_a \lambda^2 e^{-d/\lambda} \\ &\quad \times \left(1 - e^{-(t_p-t_I)/\lambda} - e^{-s/\lambda}\left(1 - e^{-(t_p-t_I)/\lambda}\right)\right)\left(1 - e^{-t_a/\lambda}\right)\left(1 - e^{-t_I/\lambda}\right)\end{aligned}$$

If $\lambda \ll t_p, t_I, t_a$

$$\Delta\tau_{full} \approx \pi\alpha G h \left(\frac{w^2}{4} - j^2\right)(\rho_h - \rho_l)\,\rho_a \lambda^2 e^{-d/\lambda}\left(1 - e^{-s/\lambda}\right)$$

If, further, $\lambda \ll s$

$$\Delta\tau_{full} \approx \pi\alpha G h \left(\frac{w^2}{4} - j^2\right)(\rho_h - \rho_l)\,\rho_a \lambda^2 e^{-d/\lambda}$$

In the numerical implementation of our torque calculations, Equation 2.2 is used as the basis for all calculations.

2.2.3 Angled plate torque

In practice, the pendulum is not exactly parallel to the attractor. Furthermore, rotating one part of the pendulum toward the attractor can increase sensitivity. This analysis, like the flat plate calculations, neglects edge effects. This approximation underestimates the expected torque by an amount that depends on the angle, θ, pendulum dimensions, and λ. See Section 2.2.3.1.

We will compute the Yukawa torque about $x, y = 0$ on a finite plate of height h, thickness t_p, and density ρ_p that spans $x \in [a, b]$ due to an infinite plane of density ρ_a and thickness t_a. Let their face-to-face separation (real or projected along the finite plate's plane) at $x = 0$ be d. Let the angle of twist of the finite plate about the vertical plane be θ. We will work in the limit that $t_p \ll \sqrt{\int_a^b |x| dx}$ (plate is functionally large in x compared to its thickness). We will only consider $|\theta| \ll 1$. This analysis only applies when $\max(|a|, |b|)|\theta| < d$ (plate does not touch the infinite plane)

The Yukawa field from an infinite planar attractor is

$$g(d) = 2\pi G\rho_A \alpha\lambda(1 - e^{-t_a/\lambda})e^{-d/\lambda}.$$

The torque on the finite plate is:

$$\tau = h\int_a^b \int_{x\theta+D}^{x\theta+D+t_p} \rho_p g(y) x dy dx,$$

where the second integral is over the thickness of the finite plate.

$$\tau = h\rho_p \int_a^b \int_{x\theta+D}^{x\theta+D+t_p} g(y) dy x dx$$

$$\tau = -2\pi G\rho_A \alpha h\rho_p(1 - e^{-t_a/\lambda}) \left(1 - e^{-t_p/\lambda}\right) e^{-D/\lambda} \frac{\lambda^4}{\theta^2} \left[e^{-b\theta/\lambda}(\theta b/\lambda + 1) - e^{-a\theta/\lambda}(\theta a/\lambda + 1)\right] \tag{2.3}$$

Critically, for a fixed angle θ, the angular correction has no dependence of D, making it quite compatible with the existing flat-plate analysis.

As an intuition check, we show the case where $a = -b$, and $b > 0$, the case of a uniform plate wiggling about its center.

$$\tau = -4\pi G\rho_A \alpha h\rho_p(1 - e^{-t_a/\lambda}) \left(1 - e^{-t_p/\lambda}\right) e^{-D/\lambda} \frac{\lambda^4}{\theta^2} \left[\cosh(\frac{b}{\lambda}\theta)\frac{b}{\lambda}\theta - \sinh(\frac{b}{\lambda}\theta)\right]$$

While we do not incorporate this calculation into our current analysis, preferring instead to take a distance penalty for the pendulum twist, future iterations of the analysis should do so. Vectorizing this calculation for the Octave analysis requires careful attention to detail, and may have substantially slowed our bootstrapped fits.

2.2.3.1 Estimation of edge effect correction

The "parallel-plate" approximation underestimates the strength of the interaction between the pendulum inlay and the attractor. An estimate of the fractional correction, f_c, scales roughly as the perimeter of the inlay times λ:

$$f_c = \frac{2(w + h)\lambda}{wh}$$

where w and h are the width and height of the inlay. For $\lambda \sim 10^{-4}$ m, this effect is about 2% for our geometry.

2.3 Design

2.3.1 Experimental masses

Our Ti (4500 kg/m^3) torsion pendulum had a thickness t =(1.8 ± 0.1) mm, height h =(32.1 ± 0.1) mm, and width w_{pend} =(43.1 ± 0.1) mm. The Ta (17000 kg/m^3) inlays were (203 ± 51) μm thick and w_{inlay} =(15.5 ± 0.1) mm wide. Uncertainties quoted here are conservative estimates for dimensions; the surfaces themselves are much flatter, as seen in Figures 2.10 and 2.11. The pendulum used here was our prototype, pressed into service when a better-characterized improved pendulum suffered a glue failure. The pendulum was threaded onto a thin copper tube crimped onto a $\sim$ 9 cm long 18 μm diameter tungsten torsion fiber. The upper end of the fiber was connected to a cylindrical (9.5 mm diameter, 28.5 ± 2 mm length, $\sim$ 18 g) copper "prehanger" mass that was suspended by a $\sim$ 9 cm long $>$ 50 μm diameter fiber in the magnetic field of a small ring magnet to damp the pendulum's swing and vibrational modes (see Figure 2.6). Both the fiber assembly (Figure 2.5) and the magnetic damper were attached, through a ceramic electrical isolator, to a motorized two-axis translation and one-axis rotary stage. Electrical contact to the pendulum was established by a thin ($<$0.5 mm) silver-plated copper wire clipped to the otherwise-isolated fiber support.

The moving field mass, or "attractor", was constructed of a (0.76 ± 0.03) mm thick Ta sheet glued to an Al alloy (7075 alloy, density 2700 kg/m^3) backer/gravitational compensator. The backer was (74.85 ± 0.05) mm in diameter, with a thickness of (1.66 ± 0.08) mm. In addition, the monocoque backer's rim extended (10.49 ± 0.05) mm, with an inner diameter of (48.64 ± 0.05) mm. Three aluminum spindles ((73.69 ± 0.08) mm long, (4.60 ± 0.03) mm diameter, (3.182 ± 0.010) g mass) were attached to the back of the backer with aluminum set screws. Each spindle was attached, via Macor machinable glass electrical isolators, to a pneumatically-driven flexure translation stage. The isolators were (26.85 ± 0.03) mm in length, (6.55 ± 0.03) mm in diameter, and had a mass of (1.972 ± 0.007) g.

Both large faces of the pendulum and the front face of the attractor were hand-polished to a mirror finish with silicon-carbide and 3M lapping paper (aluminum oxide, possibly diamond). The individual components of the pendulum were etched before assembly in hydrochloric acid to remove any ferromagnetic impurities. After ultrasonic cleaning and a final rinse in de-ionized water, 30 nm Ti and 100 nm Au were deposited by thermal evaporation. The titanium layer was used beneath the gold to promote adhesion.

The pendulum and attractors used in this experiment are actually our prototypes. A second higher-

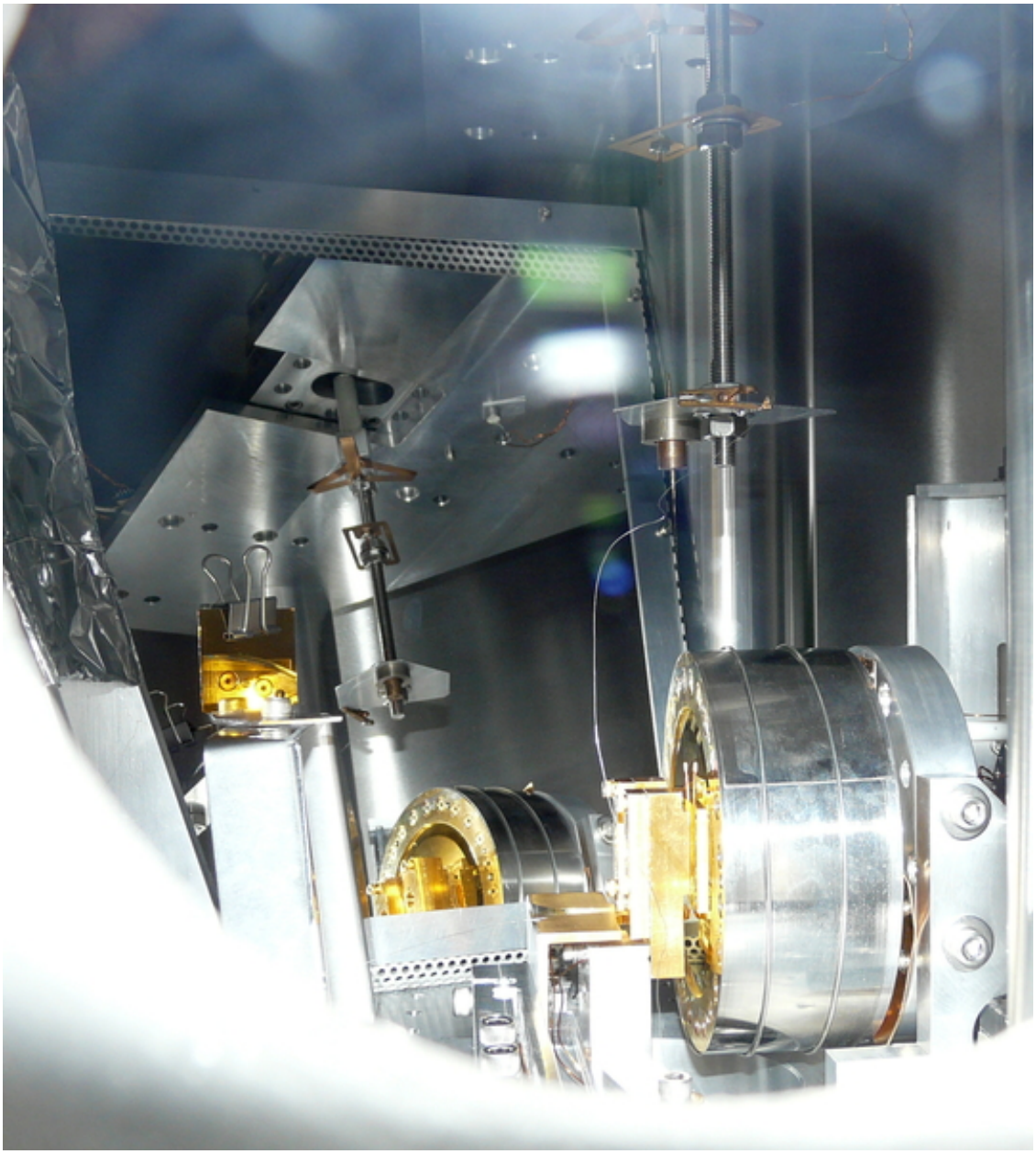

Figure 2.5: A view through the vacuum window: The entire fiber assembly is visible at right and in a reflection in a large mirror at left. Phosphor-bronze or beryllium-copper springs provided vertical isolation of the fiber attachment point. Above the three-fold-symmetric upper spring is the porcelain electrical isolator separating the pendulum from the rest of the apparatus. The pendulum ground wire is visible in the foreground. At left is a small mirror allowing a face-on image of the fiber against a scale scribed into the isolating foil.

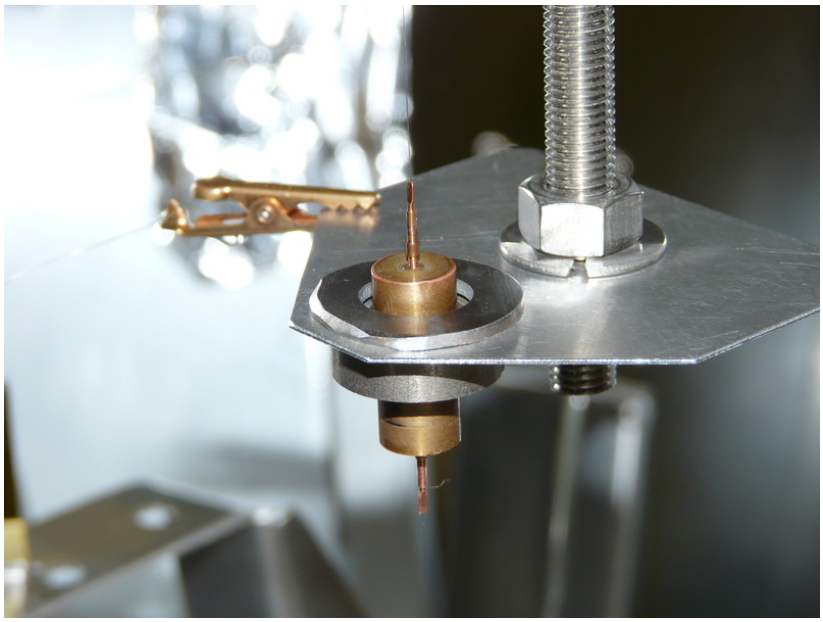

Figure 2.6: Magnetic damper, June 2011. The upper fiber is the "pre-hanger" fiber to provide isolation from tilt couplings, the lower fiber is the torsion fiber. The copper mass provides a mechanism for eddy-current damping of pendulum swing in the presence of the magnetic field gradients from the ring magnet. The alligator clip in the background connects the pendulum/phi-top assembly to electrical ground.

quality Ti-Pt pendulum was built and operated for more than a year in vacuum before it was noted that the inlay glue had failed and the surfaces had bubbled by more than 20 μm. A refined attractor suffered a similar problem. Fortunately, the prototypes are sufficient to deliver an interesting measurement.

2.3.2 Isolating Foil

The electrostatic shielding "foil" was made by stretching (13.9 ± 0.5) μm thick BeCu alloy foil (Micro-Tek, Diamond Bar CA) over the tapered end of a 77 mm inner-diameter Al tube like a drumhead. The foil was flat to better than $\sim \pm 3$ μm over its width (see Figure 2.13). We coated the pendulum side of the foil with an optically-thick ($\sim$ 100 nm) layer of Au over a Ti adhesion layer by e-beam evaporation. The foil was mounted to an electrically isolated frame to allow us to vary its electrostatic potential. The foil was mounted with a screw/flexure arrangement that permitted small tip and tilt adjustment of the foil. Two of the three screws were connected via long aluminum screwdrivers to in-vacuum adjustment motors. The motors were electrically isolated from ground, floating with the foil.

2.3.2.1 Resonant frequency

The resonant frequencies of a clamped circular tightly-stretched membrane with negligible internal stiffness are

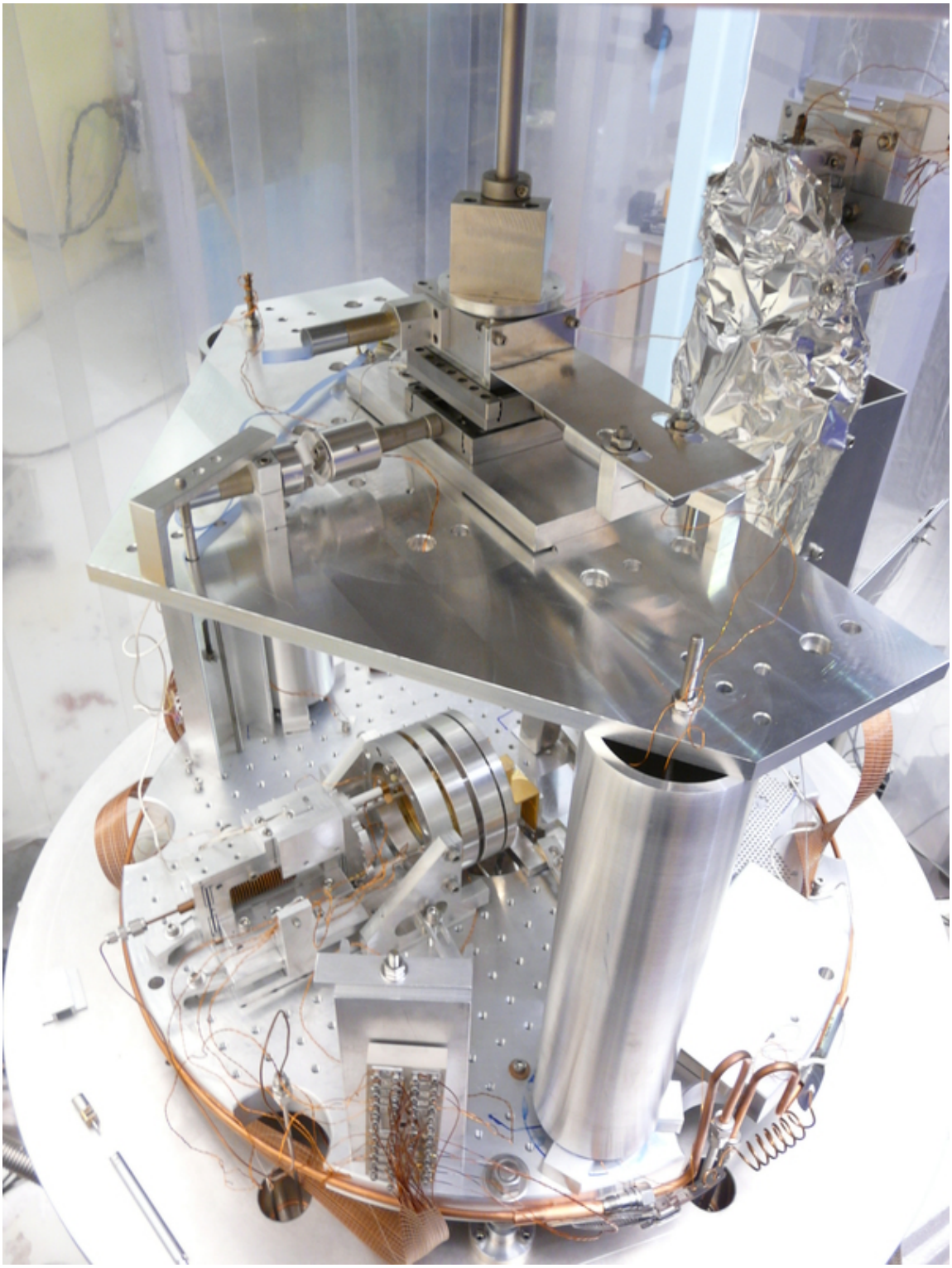

Figure 2.7: Fiber positioning assembly, January 2011 (prior to foil-motor installation): two axis translation, axial rotation through externally-controlled motors. Axial/vertical translation by manual release of passive flexure/clamp. Foil-covered autocollimator visible at right.

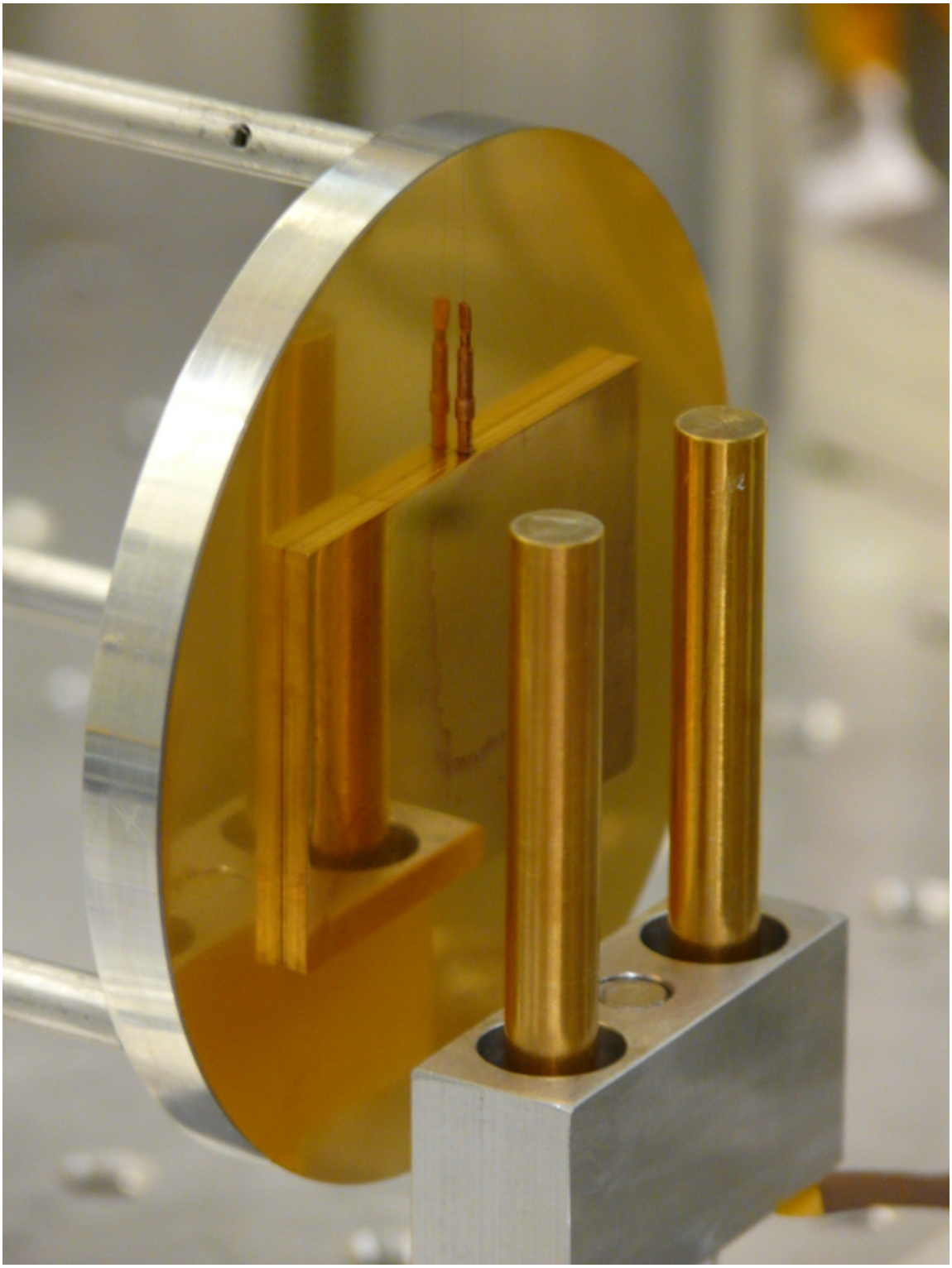

Figure 2.8: Intermediate-prototype pendulum, attractor, and early feedback electrodes demonstrating attractor/pendulum arrangement without the shielding foil. May, 2009. The rod electrodes have since been replaced with guarded planar electrodes for reduced swing coupling and greater feedback control.

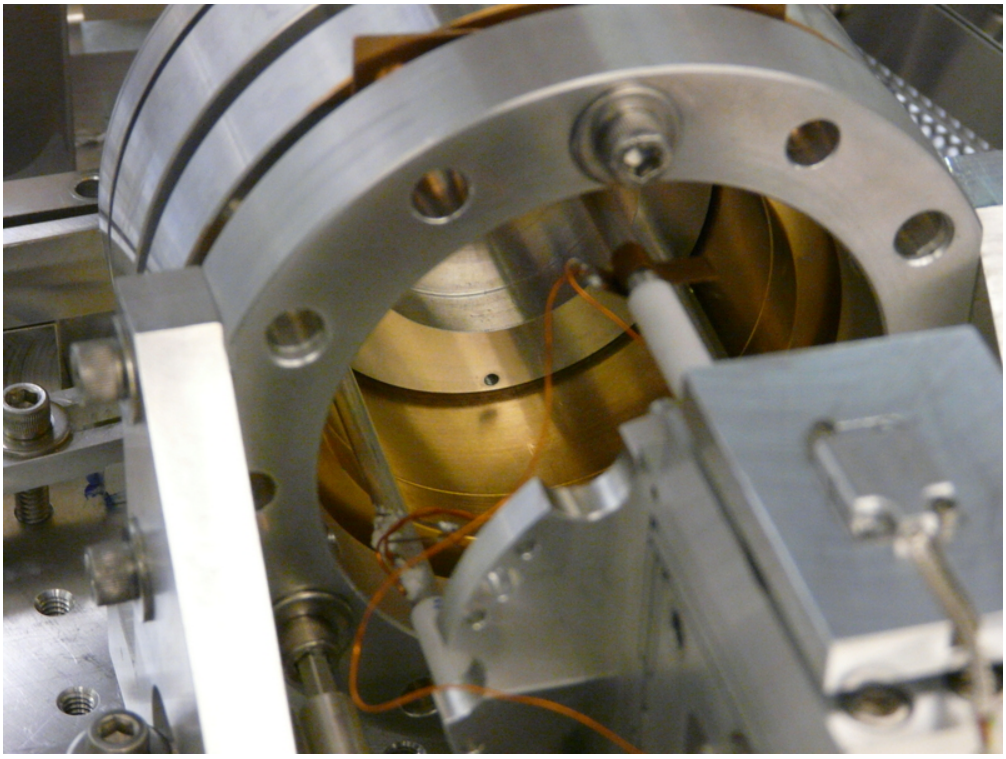

Figure 2.9: Attractor thermistor and heater glued to attractor spindle at center-left, attractor grounding clip at center-top. White cylinders are Macor isolators for the attractor. Beryllium-copper foil-adjustment springs visible at top, foil-positioning motor screwdriver at bottom left. Over-thick attractor rim visible at center, with six-fold-symmetric screw hole.

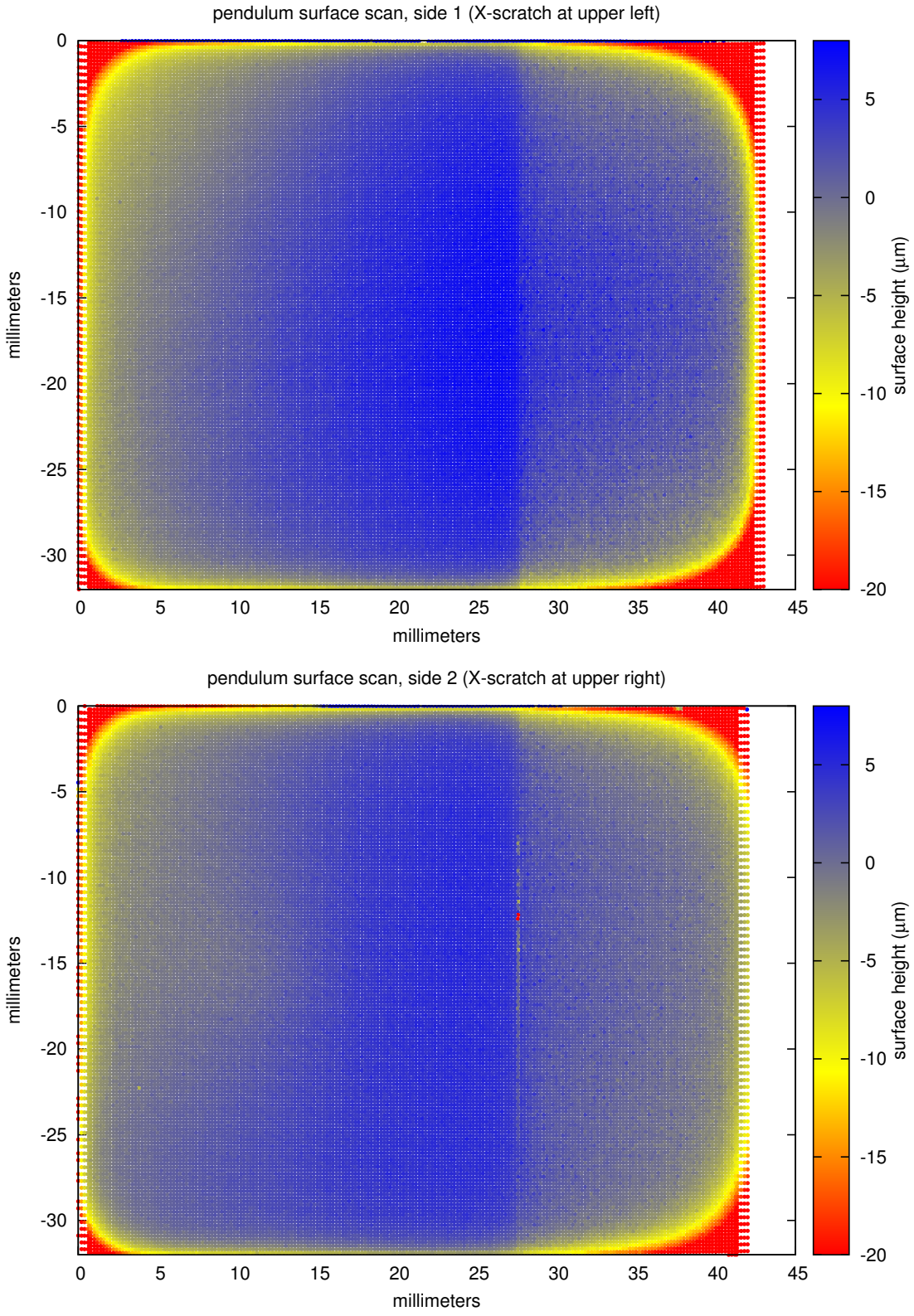

Figure 2.10: Laser scans of the pendulum surfaces. The tantalum inlays are visible as the height-step at right. The side of the pendulum used in this experiment is the one denoted, "X in upper right".

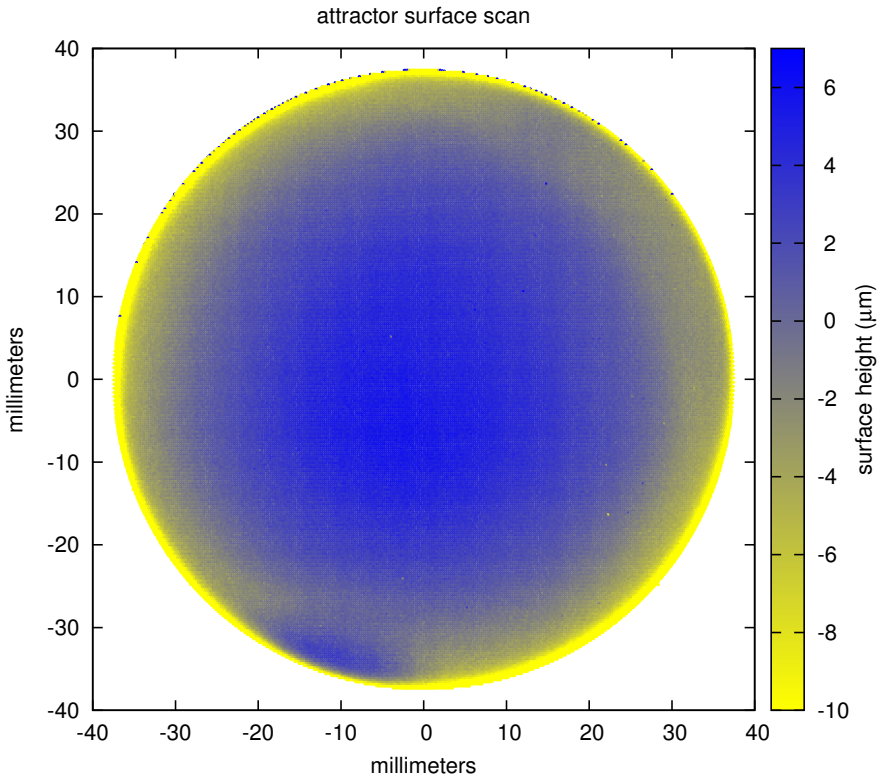

Figure 2.11: Laser scan of the attractor surface – installed orientation does not correspond to alignment shown here

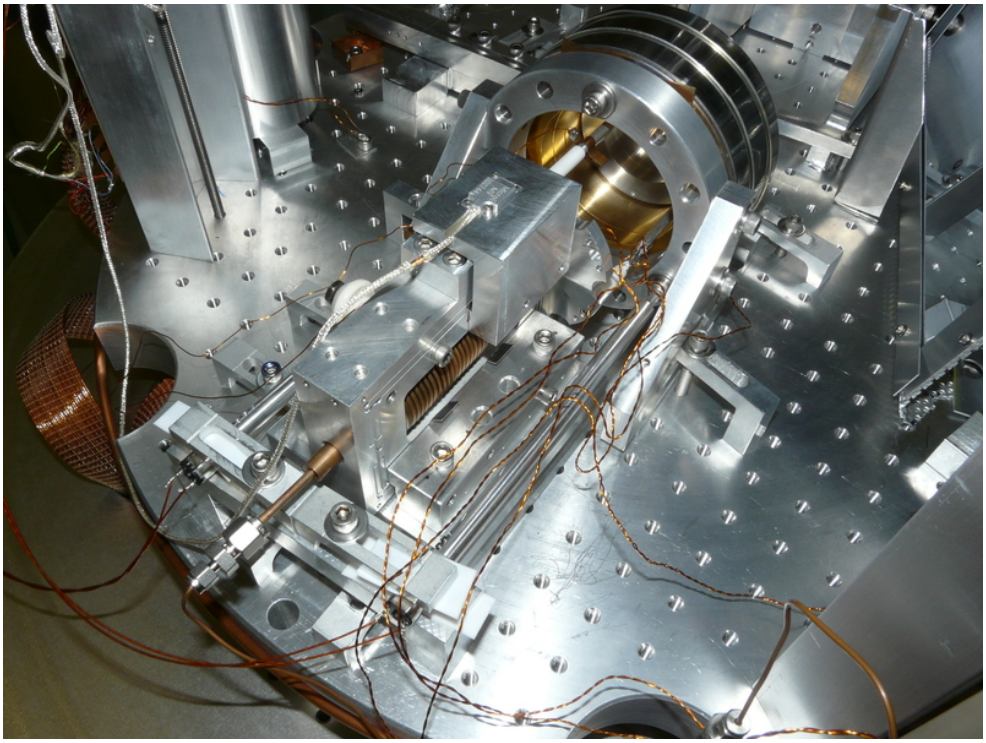

Figure 2.12: Rear view. Shows flexure, encoder, attractor, spindles, foil holder, foil adjustment motors.

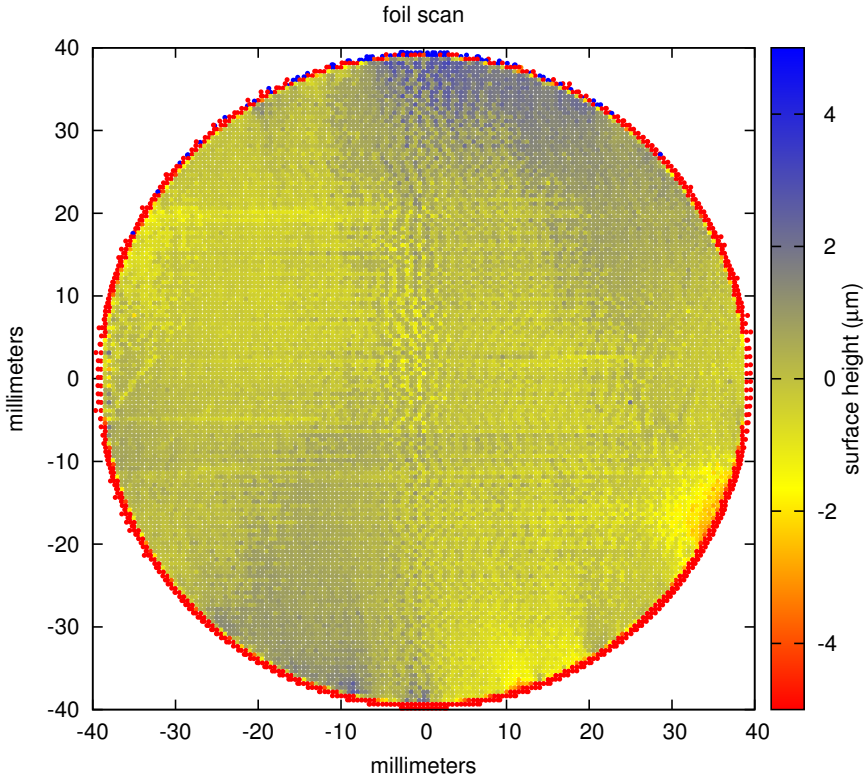

Figure 2.13: Laser scan of the foil surface

$$f_{n,m} = \frac{\alpha_{n,m}}{2\pi r}\sqrt{\frac{T}{\sigma}},$$

where r is the membrane radius, σ is the membrane mass per unit area, T is the surface tension, and $\alpha_{n,m}$ are the Bessel function zeros that satisfy $J_n(\alpha_{n,m}) = 0$ [75]. The most important resonant frequency is the fundamental, for which $\alpha_{0,1} = 2.4048$ [76]. If all the other quantities are known, then the foil tension may be measured as $T = \sigma 4\pi^2 r^2 f_{n,m}^2/\alpha_{n,m}^2$. Using the foil interferometer (2.3.9), the in-vacuum fundamental frequency was measured to be (1.580 ± 0.005) kHz. Assuming a density of beryllium-copper 8200 kg/m^3 [77] and the foil thickness, we find that the foil tension is 2.9×10^3 N/m.

2.3.2.2 Foil-stretcher design

The design of the mechanism for stretching foils is important. The foil must be flat and tightly stretched. More recently, it has become clear that the foil surface must be easy to access and clean. In the past, our group has constructed foils by clamping beryllium-copper foils between strong aluminum rings while the entire assembly was submerged in liquid nitrogen. Upon warming to room temperature, the differential

expansion of the aluminum rings brought the foil to high uniform tension. This arrangement offers superior rigidity, but the clamping rings do not provide sufficient access to the foil face.

The PlateWash foil stretcher is designed to provide good access to the front face of the foil and excellent flatness. It is constructed much as one would make a bongo drum; a clamping ring holds the foil at its perimeter, and a rounded-edge inner "pusher" that pushes on the back of the foil to bring it up to tension. Key to the success of this design is the rounded edge of the inner pusher. The radius of the edge must be large compared to the foil thickness in order to completely determine the state of the foil at the position of last contact. The radius of the edge should continue well past the point of last contact, to ensure a smooth transition, without any burrs or varying radial position of last contact. Furthermore, after turning, sanding, and polishing on the lathe, the pusher should be gently lapped, just for a moment, on a granite inspection plate to make the final planarization of the point of last contact.

In order to enable excellent access to the foil face, but preserve uniform tension in the foil, it is essential that the front ring of the foil clamping ring be very thin. In the case of PlateWash, with parallel fiber/foil alignment, the fiber is separated by only a millimeter from the foil face. The front clamping ring is only ~ 1.2 mm thick. This allows the foil to remain nearly flat when being tensioned, requiring minimum slack in the foil to permit the foil pusher to come into tension while permitting full access to the foil face. In this iteration of the experiment, the front ring does not retreat behind the foil face, requiring a slight deflection of the fiber screw and inclination of the apparatus in order to clear the torsion fiber.

The dimensions of the entire instrument are set by the availability of thin foils. We were unable to find strong metal foils of thickness < 12 μm in widths larger than 4 inches. This constrains the inner bore of the foil stretcher, hence the size of the attractor, and therefore size of the pendulum.

2.3.3 Autocollimator

The twist angle of the pendulum was measured by an in-vacuum laser pseudo-autocollimator with a 400 mm focal length. Light from a near-infrared (~ 780 nm) laser diode was directed through a pinhole, collimated, and directed at the pendulum. Light reflected from the pendulum passed back through the collimating lens and was focused, via a mirror, onto a two-dimensional position-sensitive photodiode. The detector (UDT SC-10D) had a (10.16 ± 0.01) mm (corresponds to 13 mrad) dynamic range in both tip and twist. The autocollimator noise gave a sensitivity of 50 nrad/$\sqrt{\text{Hz}}$, plotted in Figure 2.14. The four photodiode signal currents were measured by transimpedance pre-amplifiers adjacent to the photodiode in vacuum. We refer to the sensor as a 'pseudo-autocollimator', as the traditional beamsplitter of a true autocollimator was replaced by a pair of 45-degree mirrors placed astride the objective lens optical axis. This arrangement

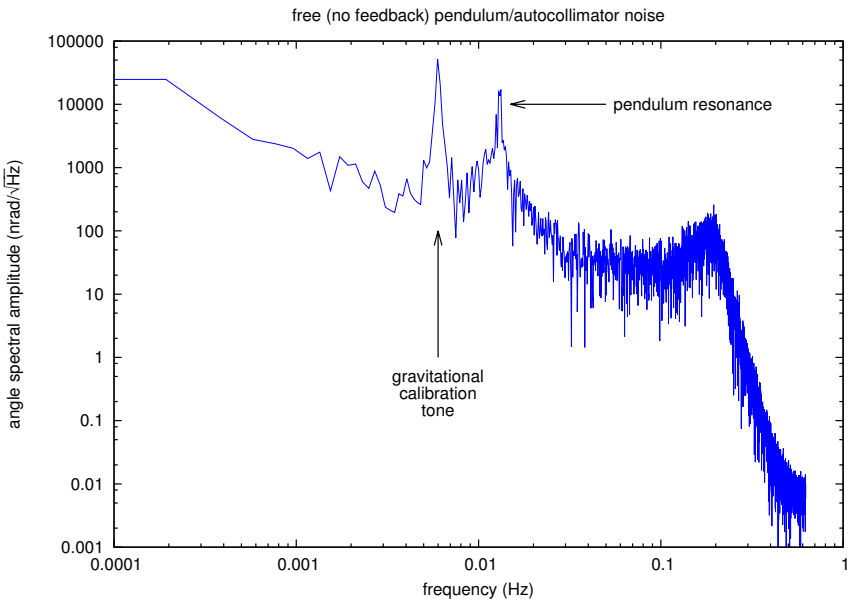

Figure 2.14: Free pendulum/autocollimator noise

preserves 2-4 times as much light intensity at the price of introducing a weak sensitivity to translation of the pendulum mirror along the optical axis; it remains unclear whether this was an advantageous trade. The added translation sensitivity was not burdensome, as excellent entire-instrument performance was easy to reach at large pendulum-foil separations.

The tip and twist of the pendulum were computed in software. As in an analog readout, the angles were determined by taking "Diff/Sum" – for each axis the difference between the left and right (up and down) channels was computed and divided by their sum. This standard procedure minimizes the autocollimator's sensitivity to light intensity variations. Doing the computations in software allowed more-precise leveling of the twist axis signal than can be readily achieved in hardware, as any synthetic combination of the tip and twist angles can be trivially generated.

In addition, the autocollimator was operated with a software lock-in technique. The laser drive current was modulated at 500 Hz, creating an approximately square-wave signal at the photodiode. The photodiode signal was digitized at 25 kHz. In each cycle of the laser drive current, 25 samples were recorded with the laser on, and 25 samples with the laser off. By taking the difference, we suppress the influence of any photodiode dark current. By averaging the locked-in signal, interference at frequencies incommensurate with the lock-in frequency is suppressed.

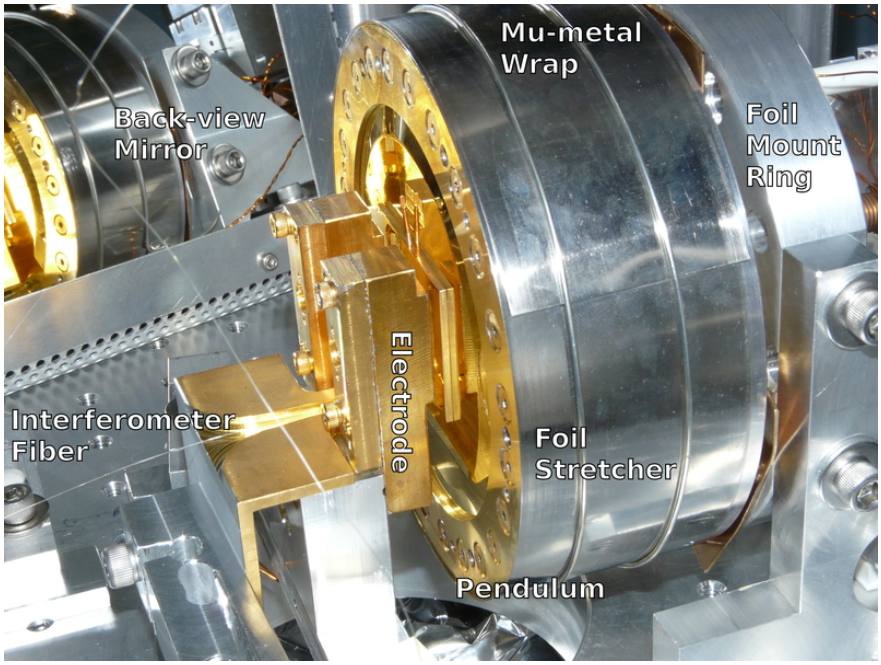

Figure 2.15: Left to right: Interferometer fiber, interferometer holder, feedback electrodes, pendulum, isolating foil, μ-metal shielding, foil adjustment leaf spring.

2.3.4 Pendulum Feedback System

For all of our science data, the pendulum's orientation with respect to the autocollimator was held in electrostatic feedback. Two planar electrodes were placed in close proximity ($\sim$ 3 mm) to the pendulum. The electrodes were housed within grounded 'guard electrodes' to control the effects of stray fields. As the pendulum was electrically grounded, the application of a voltage to an electrode attracts the pendulum in the same way that the plates of a charged capacitor attract one another.

The feedback voltage system was capable of applying voltages from -10 to $+10$ V on either electrode, but in practice, the entire experiment ran on one electrode, with DC feedback voltage variation of less than 20% over the duration of the experiment (see Figure 2.16). Running on one electrode, by imposing a small "pre-torque" from the torsion fiber, has the advantage of avoiding feedback nonlinearities introduced by any offset voltages or contact potential differences that might be present between the two electrodes. From a practical standpoint, however, having two electrodes makes recovery from extreme transients (earthquakes, attractor-foil collisions) more expedient.

2.3.5 Infrastructure

The electric signals needed to operate the experiment were conducted through three vacuum feedthroughs. The autocollimator supply, readout, pendulum motor control, and electrostatic feedback shared one feedthrough.

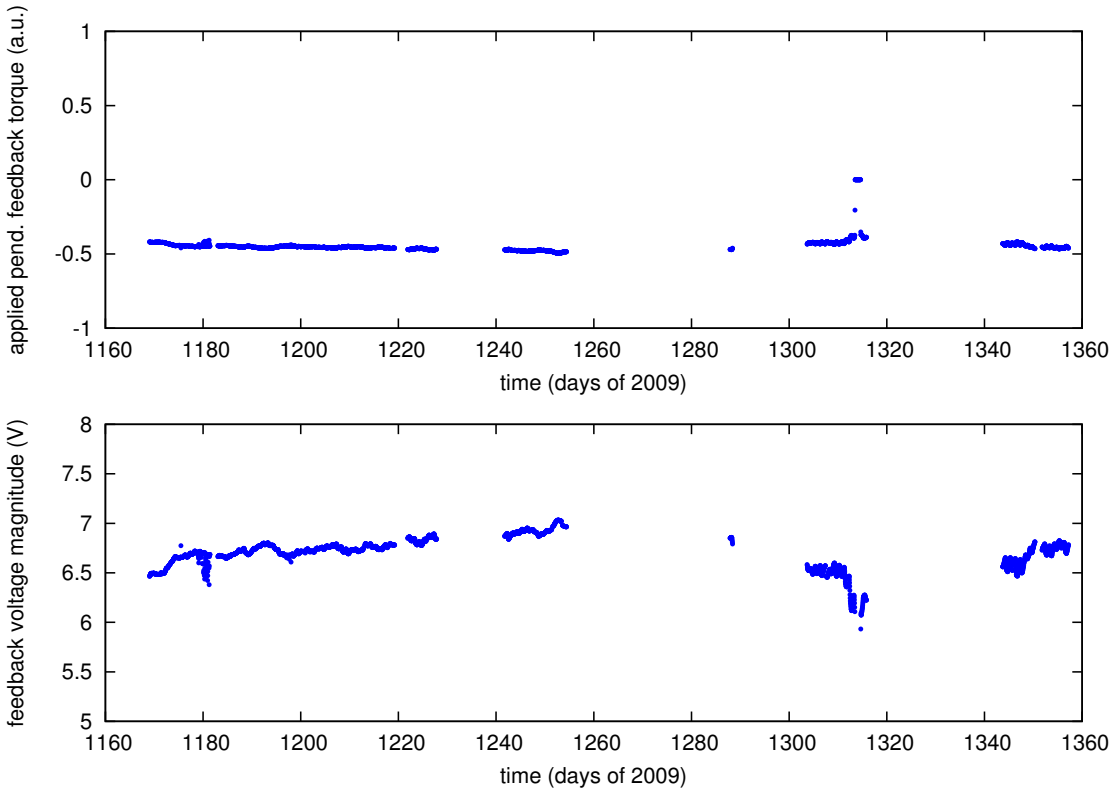

Figure 2.16: Applied feedback signal and feedback voltage over the duration of the experiment.

The attractor encoder, attractor-side thermistors, and pendulum translation encoder shared the other two. The thin copper tube for the attractor flexure bellows was connected via a Swagelok compression fitting to a separate vacuum feedthrough.

The pendulum support, autocollimator, foil, attractor stage, and feedback electrodes were all firmly clamped to a single 3-4 cm thick aluminum optical bench. The bench was secured, via three 1/2" diameter stainless threaded rods, with aluminum adapters on the ends, to the stainless steel baseplate of the bell jar. The jar itself sat atop three aluminum legs with adjustable leveling feet. The entire apparatus was installed upon a U-shaped solid masonry block table that was mortared to the cement floor of the laboratory. The bell jar's lid was sealed with a double O-ring seal. The gap between the O-rings was held under rough vacuum by the foreline pump. All high-vacuum seals except the bell-jar lid were made with copper gaskets. A 6" Leybold turbo pump was connected to and suspended from the center of the bell jar baseplate by a hydroformed stainless steel bellows for vibration isolation. A ceramic electrical isolator separated the turbo pump electrical ground from the bell jar. Suspended from the pump were six lead bricks to prevent the bellows from compressing when the system was placed under vacuum. The turbo pump was backed by a small rotary oil roughing pump; our initial hydrocarbon-free scroll pump was unreliable over the lifetime of the experiment and yielded higher foreline and chamber pressures.

When pumping the apparatus down from atmospheric pressure, the vacuum chamber was baked. Counter-wound wires around the exterior of the jar were driven by a DC power supply to heat the walls of the jar. Electrical heater tape was wound around the exterior baseplate. In addition, copper tubing was spring-clamped around the perimeter of the experiment's internal aluminum baseplate. Hot air of $\sim 40 - 50$ C was forced through the tubing to heat the internal components. After baking at least overnight, the electrical heaters were turned off, and room-temperature air was forced through the baking lines to speed equilibration. After about a day, the air was turned off, and commissioning could begin.

To eliminate the possibility of electrostatic charging of any part of the apparatus, the ion vacuum gauge was left off from pumpdown until the entire measurement was complete. After systematic checks were finished, the ion gauge was turned on, and the pressure was observed to be 2.2×10^{-7} Torr.

Past experience with the bell jar has shown that pressures of 3×10^{-7} Torr are reliably obtained by 2-3 days after pumpdown from atmosphere, and pressures of 4×10^{-7} Torr are obtained within 1.5 hours when pumping down following a brief pressure rise up to roughing pressures from high vacuum [78]. Science measurements began >1.5 hours after pumpdown from roughing pressures following a power outage, and >2 months under vacuum. With the pump valved off, the pressure rise was measured at the end of the science run to be 6×10^{-8} Torr/s, or approximately 6×10^{-5} Torr-l/s.

2.3.6 Attractor position control

The attractor's flexure stage, shown in Figure 2.17, was electric-discharge machined from a single piece of 6061-T6 aluminum sheet by the UW Physics shop. It was actuated by an internal bronze bellows that was driven by a pulse-width-modulated valve/vented accumulator apparatus external to the vacuum enclosure. The stage has 2 mm of travel, linear spring properties, is free of any "stiction" of non-flexure translation stages, and has minimal angular runout (Figure 3.1). When science data were not being taken (see below), the position of the stage could be monitored by a Micro-E 3000V encoder system with a claimed interpolated resolution of 20 nm.

The motion control system has one non-standard property for a precision translation stage: it is pneumatically actuated. In aggregate, getting this system to work well required months of effort. It was worth every day for its systematic-rejection power (see Section 8.2 for further motivation) . The pneumatic actuation system has zero electrical components within the vacuum envelope. While the attractor flexure system was designed to accommodate standard micrometer actuators, if needed, the actuator is a phosphor-bronze bellows. The bellows is fed by a thin copper tube connected to a pressure-regulation positioned outside the vacuum chamber. When the system is taking science data, the position encoder for the flexure is a

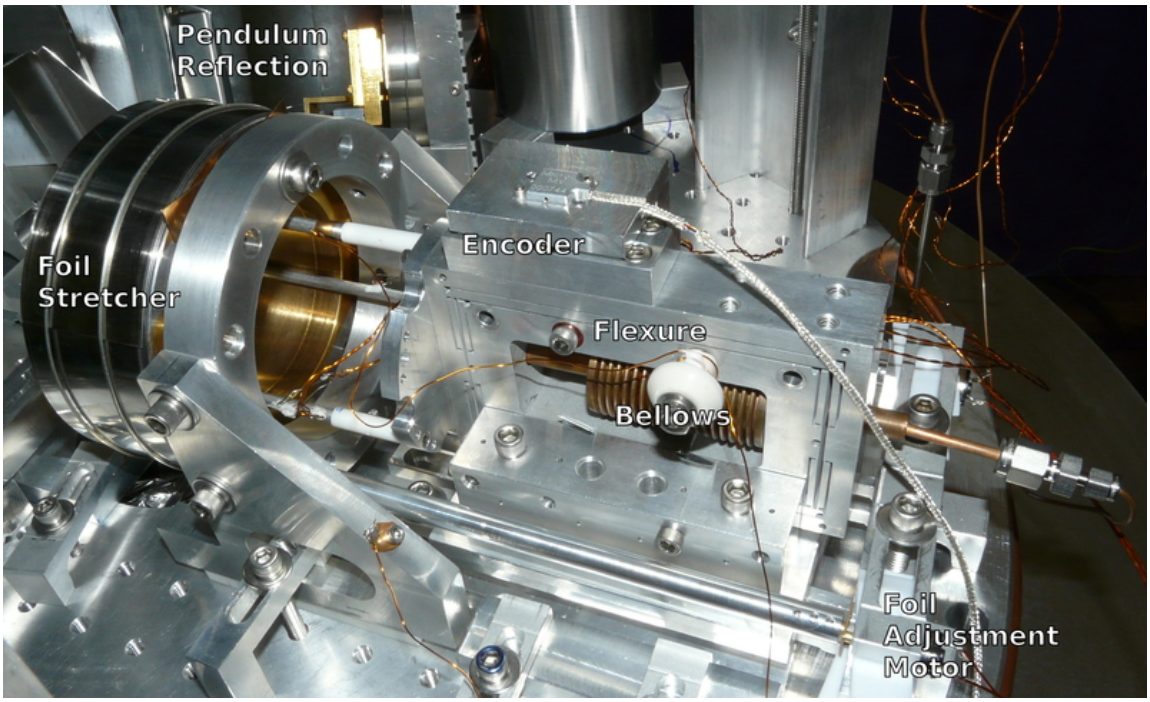

Figure 2.17: Flexure-based attractor translation stage: Single moving part has no static friction and linear response. Pneumatically driven to constrain all possible sources of electrical systematic effects.

precision absolute pressure gauge outside the apparatus. The advantage to this arrangement is apparent when considering systematic uncertainties; by closing a valve between the apparatus and the pressure regulation/measurement system, it is possible to control for all attractor-correlated systematic effects that could be introduced by the data acquisition, sensing, and control systems. This valve isolates systematic uncertainties to only those directly correlated with the physical motion of the attractor or with changes in gas pressure downstream of the isolating valve.

A diagram of the pneumatic drive system is shown in Figure 2.18. The drive system begins with a pressure regulator attached to the building's compressed air line. The line pressure fluctuates on timescales from hours to days as the system's accumulators fill and drain and as loads on the system change. The coarse-regulated air is fed into an Omega PRG101-120 precision regulator, tuned to ~ 50 psig, with claimed 0.02% stability. The "psig" is important - the regulated pressure is not absolute, but relative to atmosphere. The compressed air is then controlled by a pulse-width modulated valve, with a 10 Hz carrier. Several modulation valves were tried. By far the best solution has been automotive fuel injectors; they are very fast, low-leaking (in both directions), durable, and inexpensive. The modulated air is sent into an accumulator, a ~ 0.5 L lecture bottle. The bottle is connected to a permanent leak and to a pair of in-series fuel injectors that together form the "locking valve". Air that flows out of the permanent leak is ducted ~ 4 m away from the apparatus to limit any thermal systematic effect due to changes in modulation. When open, the locking valve permits air

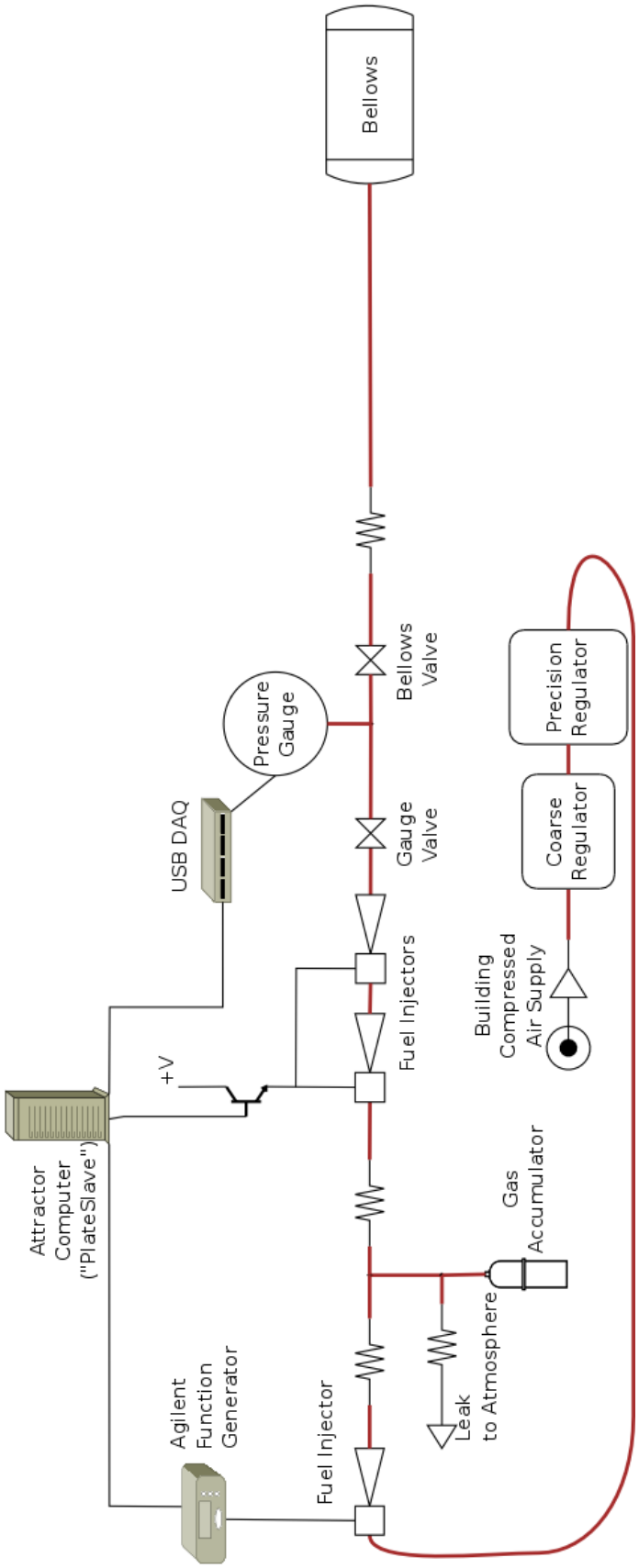

Figure 2.18: Pneumatic Drive System

to flow from the accumulator, through a manual valve (the "gauge valve"), past a precision absolute pressure gauge, through a second manual valve (the "apparatus valve"), through a copper pipe, through a vacuum feedthrough into the apparatus, through a capillary tube, and into the flexure bellows. The flow restrictions and the accumulative capacity of the bellows form a second "RC" circuit element, further low-passing the pressure fluctuations from the pulse-width modulation.

The manual valves are important. As described above, the "apparatus valve" enables complete isolation of the attractor from the outside world and all measuring instruments. The "gauge valve" is the equivalent of a lock on a pair of calipers; whatever air is present in the bellows when the valve is closed is trapped there, and the pressure/position is still measured continuously by the gauge. If there are no leaks and the absolute temperature (K) does not change, then the attractor can be held fixed in any position. As an additional capability, it was useful for foil-interferometer calibration to have a controlled and very slow leak in the attractor drive system; the manual valves make this possible.

The gas-handling system was originally designed without the "locking valve", but it became critical to our success. We had intended to operate with the pulse-width modulation valve operating in continuous PID feedback to pressure or position sensors, this would have enabled fast position changes and excellent control over the attractor's position. Alas, the pulse-width modulation was still measurable and unacceptably large at the attractor – attractor motion during measurement phases opens whole classes of systematic effects we do not wish to consider. The modulator could not be switched at higher frequencies, and any further passive low-pass filters would compromise the speed of our attractor motion. So, we fell back to open loop control of the modulator duty cycle and the "locking valve", which isolates the modulator accumulator from the bellows at times when pressure change is not desired. In this application, the bi-directional isolation of the fuel injector valves is essential.

2.3.6.1 Attractor position change procedure

When an attractor position change is not underway, the locking valve is closed, and the modulator valve is off. When the attractor-control computer is asked to change the attractor position, the modulator valve begins modulating with the desired pulse-width for 10 s before the locking valve opens. The 10 s delay isn't quite sufficient to fully pressurize the pneumatic lines; when the locking valve opens, the attractor systematically jumps about 50 μm away from the foil before moving smoothly to its next position. This effect is easily visible in Figure 4.4. The locking valve is held open for 10 s to allow the system to come into equilibrium. To complete the position change, the locking valve closes, the modulator valve turns off, and the accumulator vents.

Figure 2.19: The gas-handling system. It is remarkable that a system with such a crude appearance can yield such good performance. Low leaks, a high-precision gauge, and the three reliable/fast automotive fuel injectors are key to its performance.

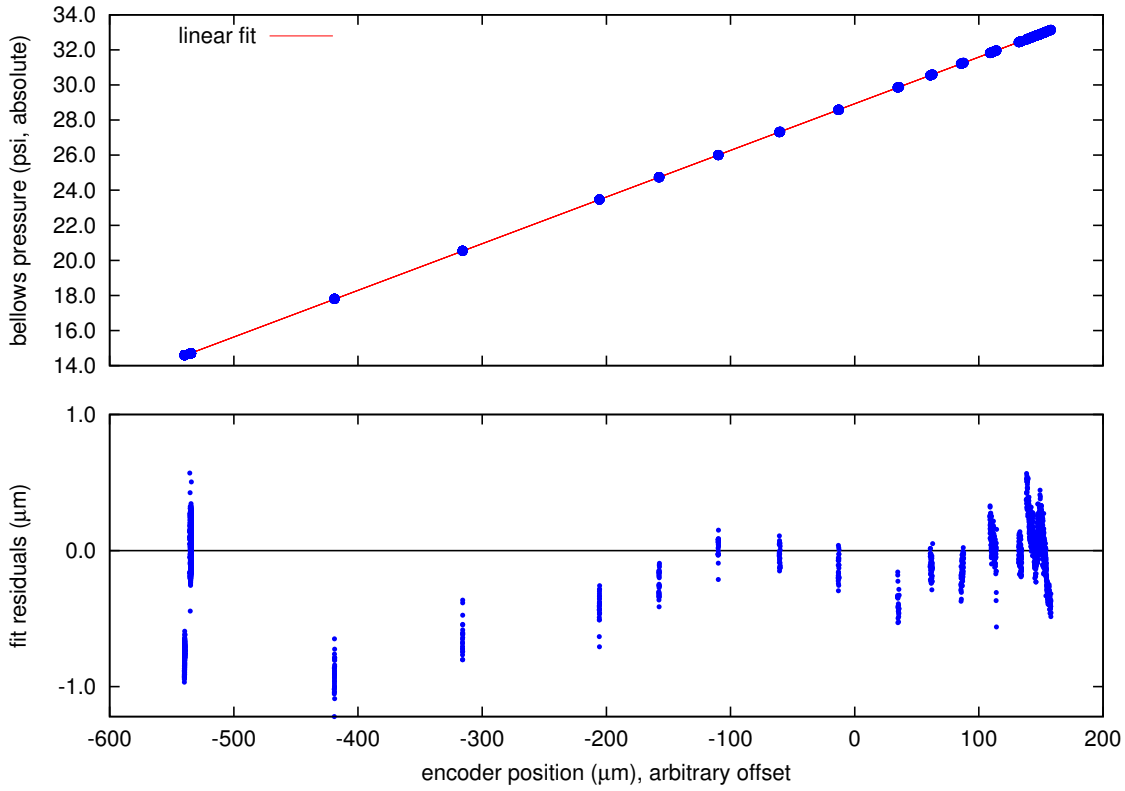

Figure 2.20: Pressure gauge and linear encoder comparison

2.3.6.2 Pressure gauge and position

The original plan was to use the very-precise Micro-E encoder for position readout. The encoder has 20 nm resolution, a nice glass encoder scale, an absolute encoder mark, and low-noise readout. Unfortunately, we were able to trace a systematic signal to the encoder. Pickup somewhere in the apparatus, probably in the autocollimator preamplifiers, was able to sense the encoder grating period, yielding a substantial sinusoidal 25 μm signal. This eluded us for some time, as 25 μm steps are natural gradations for a human experimenter – effects like these are the biggest motivation for the "random" attractor position selections we choose. So, we were forced to use another position-sensing system.

The pressure gauge is an Omega PX409-100A10V absolute pressure gauge. In order to use it as a position encoder, we had to verify its linearity and stability against the Micro-E encoder, shown in Figure 2.20. The result is remarkable, as it requires excellent linearity from each of the flexure, bellows, encoder, and pressure gauge.

2.3.6.3 Position uncertainty

The dominant "noise" in the attractor position system comes from the pulse-width modulation system itself. The modulation valve and the locking valve are not coordinated at timescales comparable to the modulation frequency, so the locking valve shuts off in different parts of the modulation every time. Therefore, the attractor stops in a slightly different location on every move. This dominates the position scatter in attractor position for each "cut" of data.

On longer timescales, the fluctuation of atmospheric pressure is important. The precision regulator that feeds the gas handling system is referenced to atmospheric pressure. The attractor flexure, on the other hand, moves according to the pressure relative to vacuum. As the gas-handling system is nothing but a fancy pressure division device, the open-loop duty-cycle control of the modulation rides along the carrier of atmospheric pressure. This effect can be seen in Figure 4.5, where runs at constant duty cycle fluctuate slightly in absolute position. This effect limited how close we could reliably operate the attractor to the foil over long periods of time with a simple control strategy; if the attractor hits the foil, at least an hour of data are lost (see Section 4.1.2).

2.3.7 Calibration

The torque scale of the apparatus was determined gravitationally by rotating a quadrupole mass distribution from a bearing above the apparatus. By way of the quadrupole coupling, the "q-tester" causes a gravitational torque on the pendulum at twice the rotation frequency. Two (23.88 ± 0.34) kg (uncertainty is statistical only) lead blocks were hung on opposite sides of the vacuum jar by nylon straps attached to a rotating armature which operated continuously during data acquisition. The blocks were weighed with a bathroom scale; calibration to 10% is sufficient, as it only sets the scale for any ISL violation. This measurement requires no absolute corrections. The 4" $\times$ 4" $\times$ 8" bricks were suspended vertically and located with their centers (629 ± 25) mm from the center of the jar and 200 mm above the center of the pendulum. The rotation axis was coincident with the torsion fiber to < 5 cm precision. The rotation was controlled by a PIC microprocessor driving a geared-down stepper motor. The frequency stability of the calibration tone was observed to be comparable to, or better than, the stability of the data acquisition clock. The sinusoidally varying gravitational torque appeared at 6.0 mHz, twice the calibrator's rotation frequency. The amplitude of the calibration torque was taken to be 17 fN-m. The q-tester was mounted on rails with positioning stops to expedite moving it out of the path of the crane-opened bell jar and ensure repeatable positioning of the q-tester with respect to the pendulum.

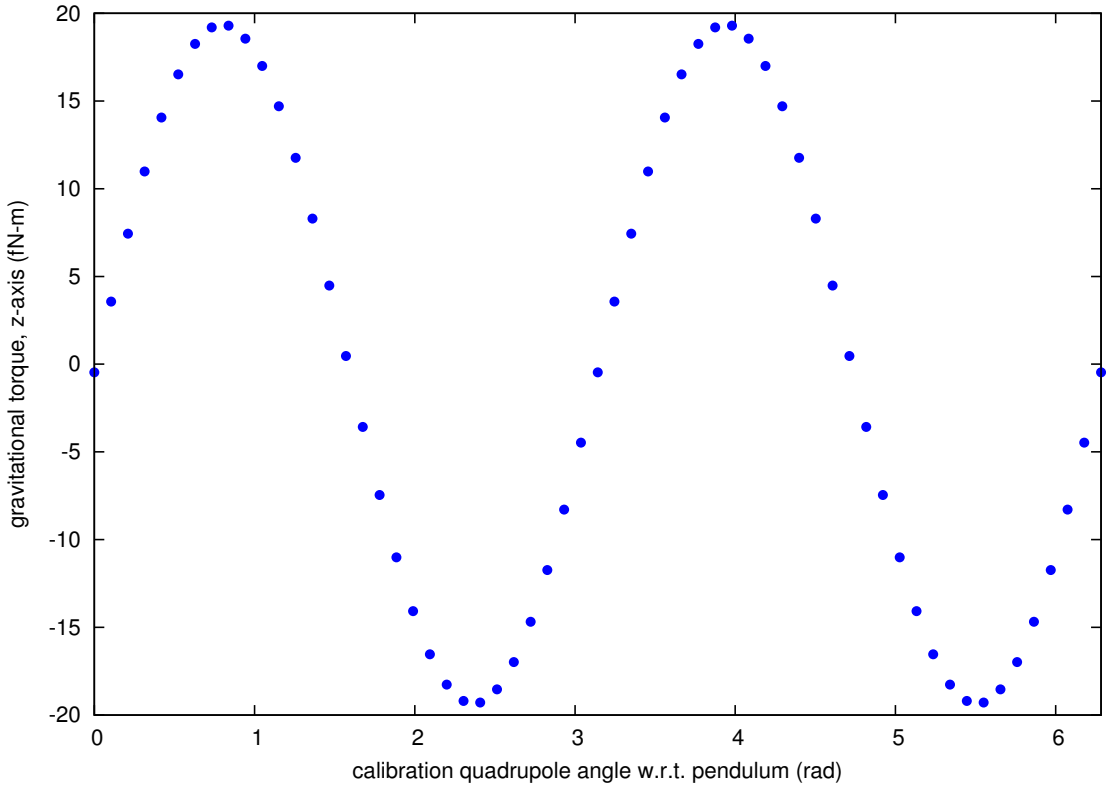

Figure 2.21: Torque between pendulum and lead-quadrupole calibration bricks, as computed by the Newtonian simulation package (Chapter 7). Note the slight phase shift with respect to zero; because of the inlays, the pendulum's mass-quadrupole axis is not parallel to the pendulum face.

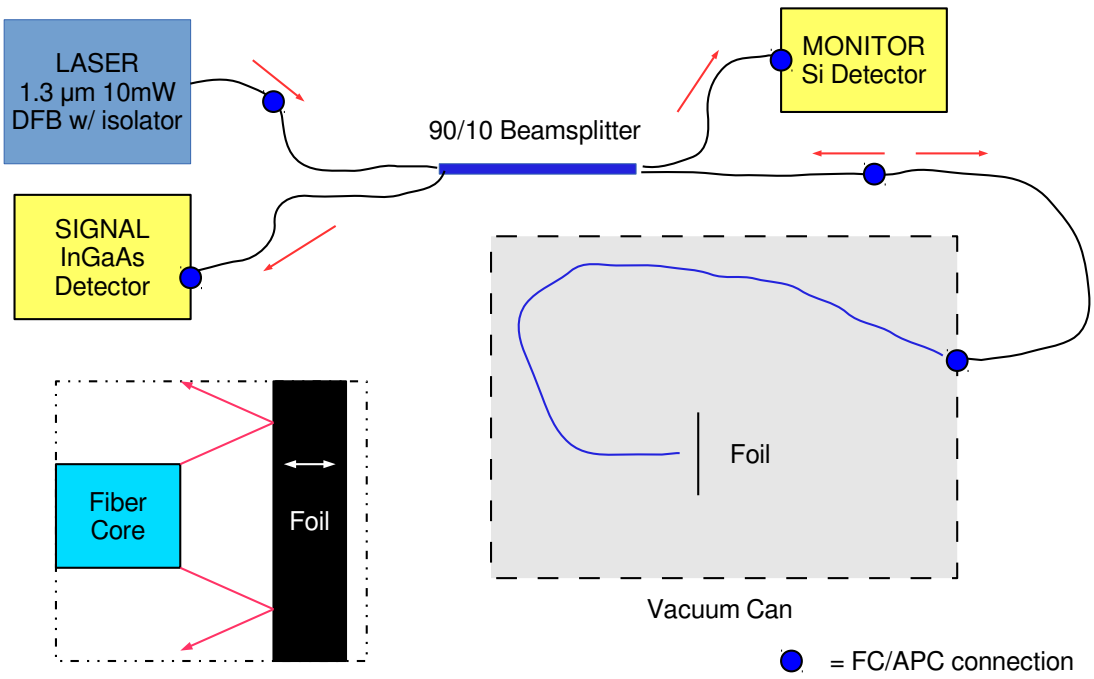

Figure 2.22: Foil Interferometer Layout Diagram

2.3.8 Cleanliness

The apparatus itself was housed inside a "clean room" to control dust. Any dust between the pendulum and the foil or the foil and the attractor could have prevented bringing them into close proximity or, worse, introduce a hidden systematic effect. Air from the outer lab was passed through filtered blowers and forced down from the ceiling of the room, keeping it at slightly positive pressure and flushing airborne dust out of the room through an opening along the floor. When the apparatus was under vacuum, the blowers were turned off; the enclosure then provided limited thermal isolation and isolation from moving air.

2.3.9 Foil Interferometer

The isolating foil is the most important defense against systematic error. It is a physical and electrical barrier between pendulum and attractor. If it moves in an attractor-correlated way, the pendulum's electrostatic environment will change. In order to assess the foil's motion, we built an interferometric optical sensor system.

The foil interferometer is a Fizeau interferometer [79]. Inspired by the designs and sensitivities of early atomic force microscope (AFM) readout systems [80, 81], the device interferes light from a reference surface

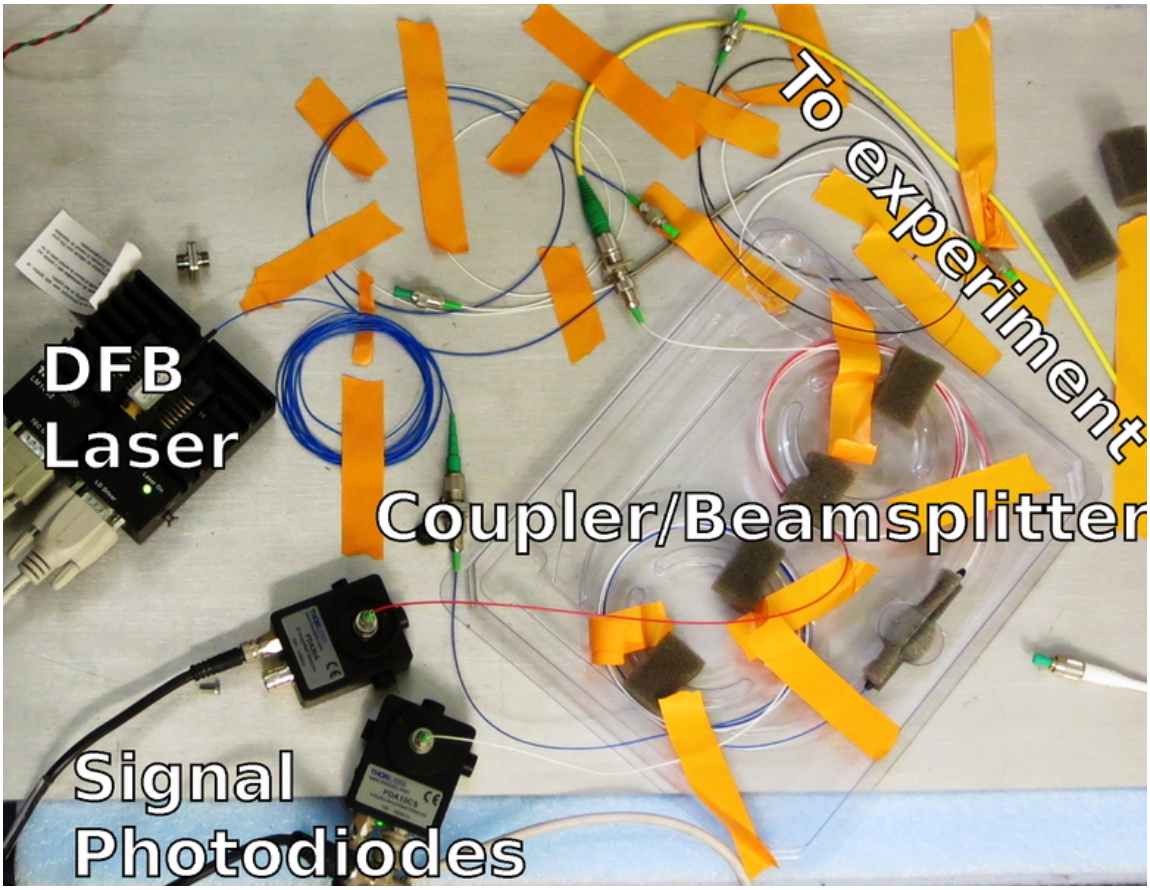

Figure 2.23: Interferometer optical assembly: All fibers have APC ends, operating at 1300 nm. The yellow signal line travels >3 m to the apparatus, accounting for the majority of the seismic pickup.

and a reflecting target surface. In this case, the reference surface is the end of an optical fiber, and the reflecting target surface is the PlateWash isolating foil. As shown in Figure 2.22, light from a 10 mW, 1300 nm fiber-coupled laser (Q-Photonics QDFBLD-1300-10) is directed through a 90/10 fiber coupler (ThorLabs FC1310-70-10-APC). The light that travels down the 90% branch is absorbed by an intensity monitor photodiode. The remaining 10% is coupled into a fiber that goes to the apparatus. The end of the fiber is carefully directed and positioned next to the isolating foil. Approximately 4% of the light leaving the fiber is reflected from the glass/vacuum interface. Due to the divergence of the beam (NA 0.14, mode field diameter 9.2 μm), less light re-enters the fiber core after reflecting from the foil. By adjusting the fiber-foil distance, it is possible to tune the internally and externally-reflected light to have comparable intensity, maximizing the interference contrast. In PlateWash, the operating point is very close to the foil, as seen in Figure 2.25.

The reflected light returns down the fiber to the 90/10 splitter. 90% of the light goes into an InGaAs high-sensitivity photodiode (ThorLabs PDA10CS). The remaining 10% proceeds back to the laser, where it is blocked by a Faraday isolator.

The signal from a Fizeau interferometer is readily understood. Assuming light with wavelength λ, amplitude A, and fiber tip to mirror spacing s, fiber core radius r, tip-to-vacuum reflectivity R_1, tip-to-vacuum transmission T_1, vacuum-to-tip transmission T_2, angular frequency ω, and radial (full width is 2θ) beam divergence angle θ, the intensity traveling back down the fiber is roughly

$$A_{reflected} = ARe^{i\omega t} - AT_1T_2\frac{1}{(\frac{2s}{r}\tan\theta + 1)^2}e^{i\omega(t+2s/c)}$$

$$I_{reflected} = |A_{reflected}|^2 = A^2\left(R^2 - 2RT_1T_2\frac{1}{(\frac{2s}{r}\tan\theta + 1)^2}\cos(4\pi s/\lambda) + T_1^2T_2^2\frac{1}{(\frac{2s}{r}\tan\theta + 1)^4}\right), \quad (2.4)$$

plotted in Figure 2.24.

The optical fiber used is ThorLabs SMF28E: Mode field diameter $\sim 9.2\,\mu$m, numerical aperture ($\tan\theta$) 0.14, index of refraction 1.47. The reflection coefficient for normally-incident light passing from glass with refractive index n into vacuum (refractive index 1) is $R = \frac{1-n}{1+n}$. Note that there is a phase flip at the foil, but not at the internal reflection, hence the minus sign in $A_{reflected}$.

2.3.9.1 Calibration

In order for the intensity values measured at the photodiode to be meaningful, we must develop an intensity-change to displacement conversion. As the intensity at the output port varies sinusoidally with foil position,

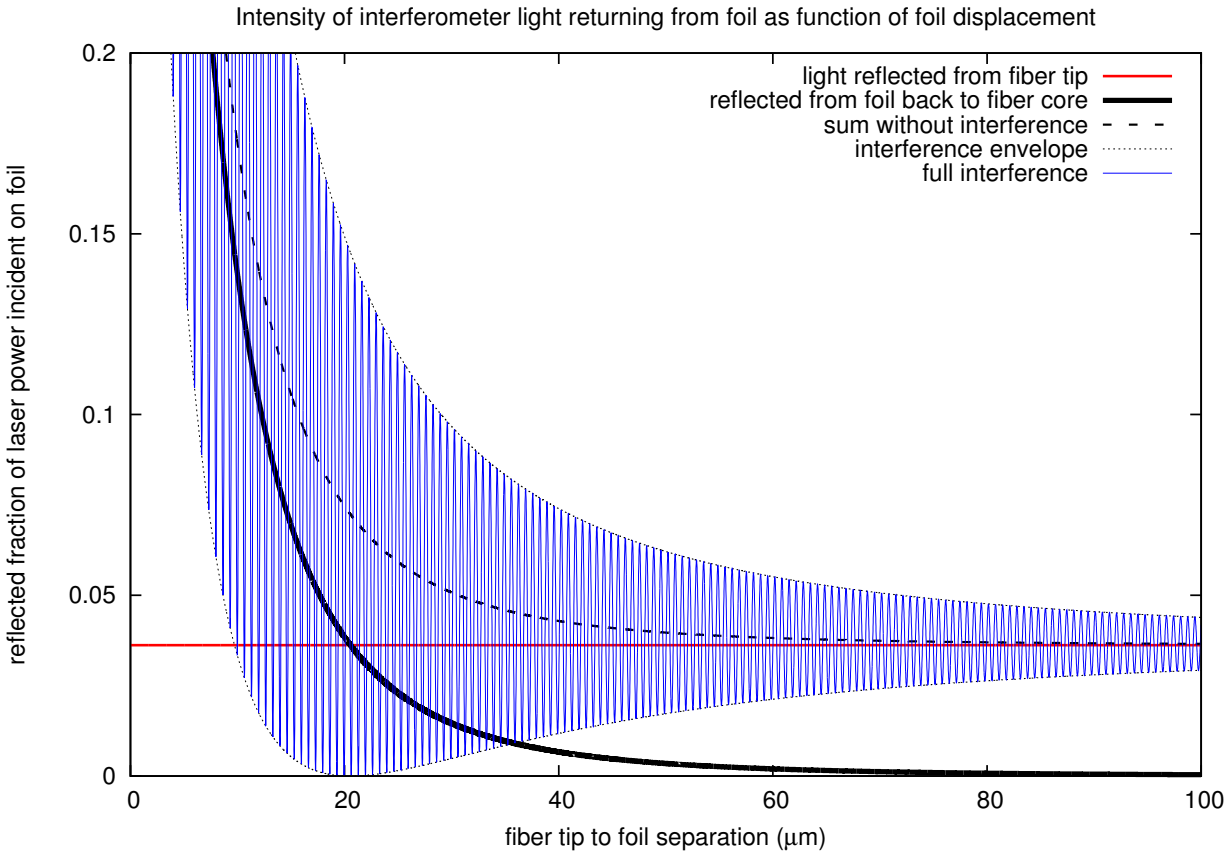

Figure 2.24: Theoretical expectation for the foil interferometer signal as a function of foil displacement (Equation 2.4). The irregular envelope for the fringes is due to the varying amount of light that returns to the fiber core from the foil.

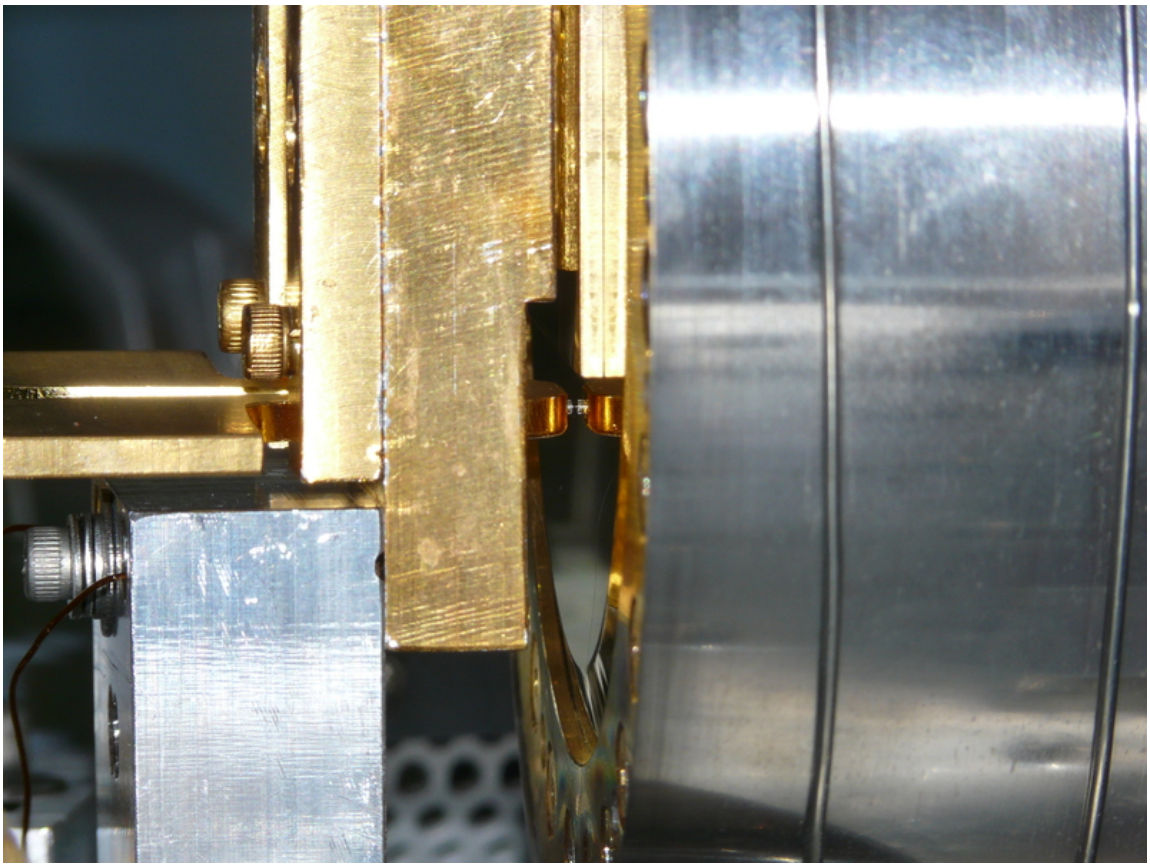

Figure 2.25: Image of the interferometer tip facing the mirror-like foil beneath the pendulum. The narrowest-diameter glass cladding, almost invisible, is 125 μm.

the sensitivity of the instrument is spatially-varying. With the initial visible-light laser, we attempted to lock the fringe phase into the location of maximum sensitivity by varying the laser temperature temperature, and hence frequency. With the infrared distributed-feedback laser, the temperature coefficient is so small that such an approach cannot work. Future versions of the experiment may include positioning feedback for the fiber mount in order to maximize sensitivity.

Determining fringe contrast can be done by displacing the foil. The attractor can be pressed into the foil by lightly over-driving the translation stage. If valves are manipulated to permit a slow retreat of the attractor (or using the very slow leak of the isolation valve), a fringe pattern may be obtained, as in Figure 3.3.

The first few fringes after the attractor hits the foil are our best measure of fringe contrast. We can identify the turning points and sensitive regions of intensity.

In Fig 2.26, we have the interferometer signal plotted over the duration of the experiment. Given the spatial extent of the materials involved and the heterogeneous nature of the assembly, the <1 μm differential displacement over a period of more than a month is excellent. Much of the experiment took place near a fringe turning point, which limits our ability to constrain the foil motion for each individual run.

Interferometer-to-displacement conversion As it is challenging to infer anything about an interferometer's signal near a fringe minimum or maximum, we don't. We only consider times when the interferometer was in the middle 71% ($\sin(\pi/4) = \sqrt{2}/2 = 0.707$) of the range of fringe contrast, i.e. the range over which the interferometer output is approximately linear. Over that range, the slope of the fringe-to-displacement sinusoid varies by at most 21% from the mean (average slope $2\sqrt{2}/\pi$, maximum 1, minimum $\sqrt{2}/2$). As we are only interested in an estimate (and hopefully an upper bound) on the foil's contribution to pendulum torque, this is sufficient.

Within this range, we use an approximation for the sinusoid's slope, if A_0 is half of the peak-to-peak fringe amplitude, $\frac{\lambda}{8\sqrt{2}A_0} \sim \frac{230 \text{ nm}}{\text{fringe (peak-to-peak)}}$, to convert measured output to displacement.

2.3.9.2 Noise

Leading noise source At sub-Hz frequencies, noise in the interferometer is believed to be polarization rotation in the fiber. Differential stress in optical fibers can rotate the polarization throughout the entire Bloch sphere [82, 83, 84]. As the light emitted from the laser is, in general, polarized, the beamsplitter is likely to be polarization-sensitive, and the cavity and detector may have some polarization sensitivity, any change in polarization throughout the fibers may yield changing results. The effect of simply jostling the fiber is easily noted in the readout of the apparatus; listening to the photodiode output with headphones

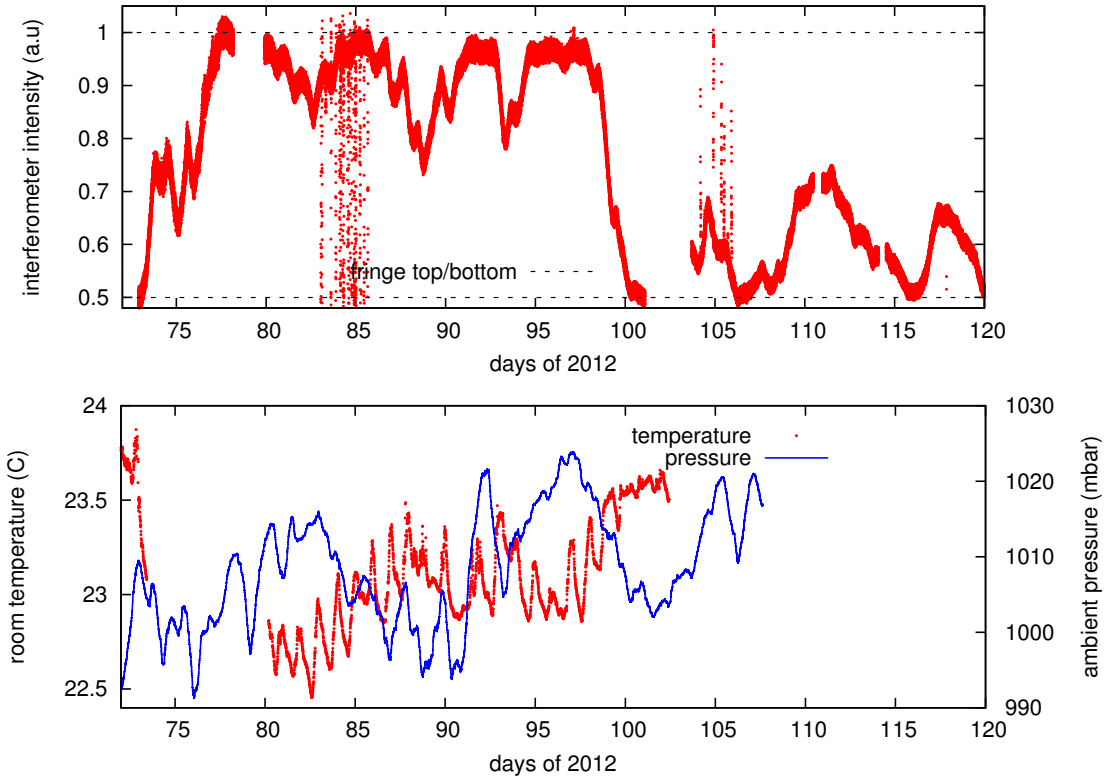

Figure 2.26: Interferometer signal over the duration of the experiment. The instrument drifted by less than a fringe over that time. The periods of substantial scatter (near days 85 and 105) are from times when the attractor was brought into contact with the foil. Only those data from the vertical-middle 71% of the available fringe dynamic range were used to build a quantitative understanding of foil motion. The environmental data are plotted to demonstrate the instrument's apparent lack of correlation with temperature and air pressure variation.

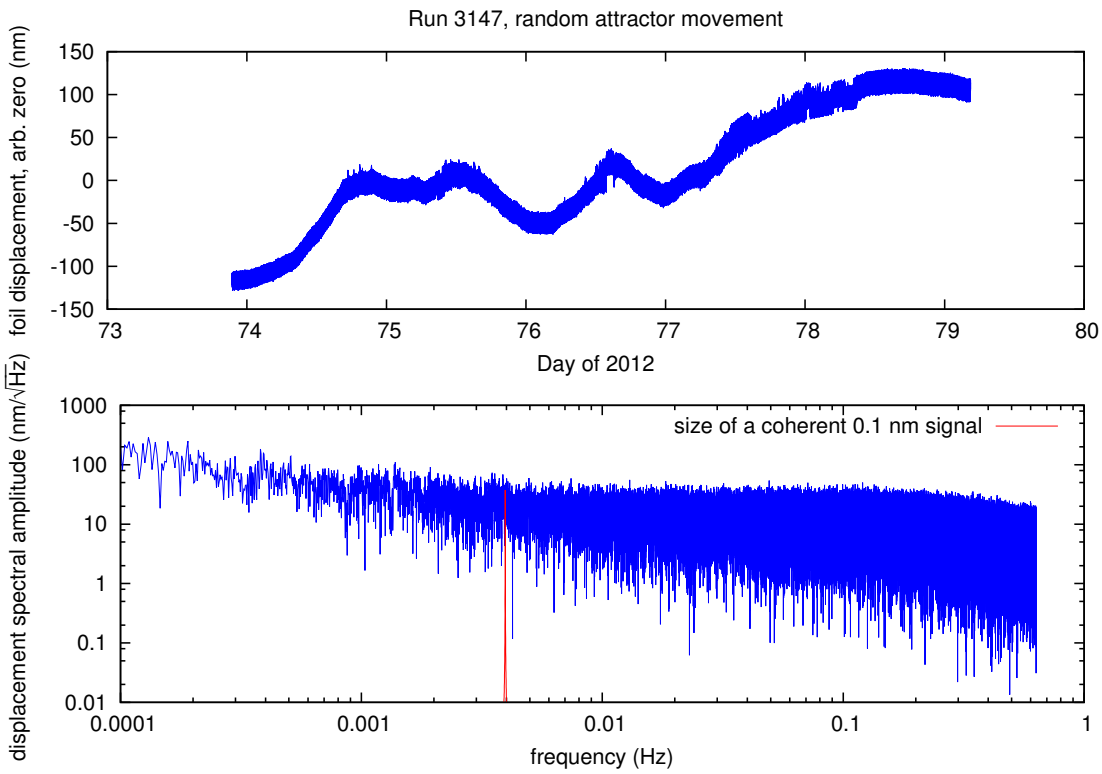

Figure 2.27: Interferometer low-frequency power spectrum

yields a siren-like tone, giving extra credence to the polarization-rotation explanation of the noise. Securing the fiber, where possible, to immovable objects with tape yielded a substantial improvement in sub-Hz noise. Noise from within the science run is plotted in Figure 2.27.

A polarization-maintaining fiber is on-hand for an upgraded version of the experiment. A new fiber coupler, in vacuum fiber, and perhaps even a new fiber-coupled laser are required for optimum polarization-maintaining performance.

2.3.9.3 Translation stage

With the introduction of the foil-positioning motors (see Section 2.3.2), it was necessary to allow control of the interferometer tip position from outside the vacuum vessel. The interferometer tip was connected to a long flexure lever with a linear reduction ratio of $\sim$ 10:1, shown in Figure 2.28. Before pumping down, the fiber tip was translated far from the foil. After the foil and attractor were aligned, then the fiber tip was slowly brought toward the foil until fringes were observed. This must be done with care; going too far will puncture the foil with the fiber tip. The motorized approach is easier, faster, allows better fringe contrast, and damages fewer foils than aligning by hand.

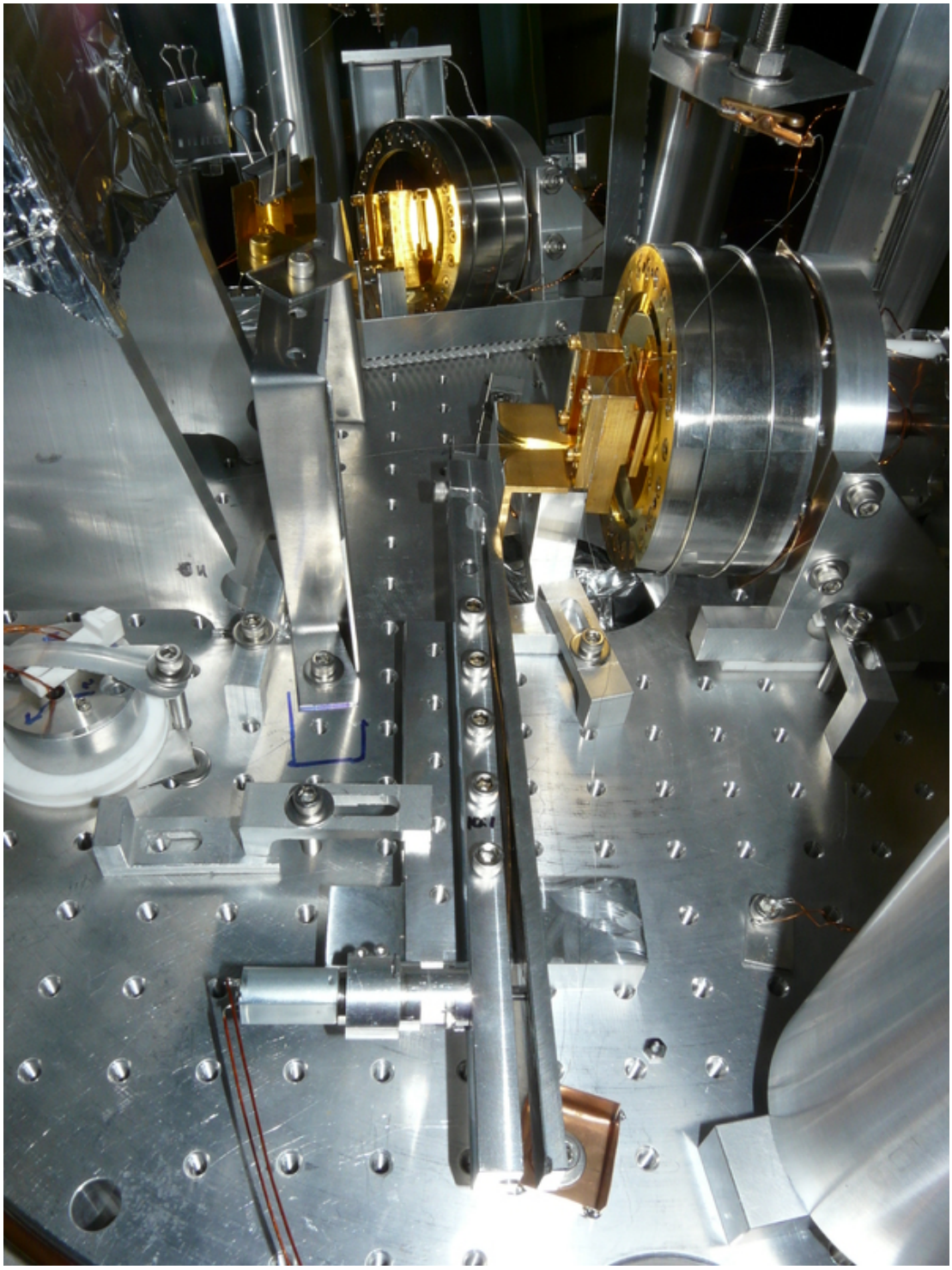

Figure 2.28: Interferometer control lever

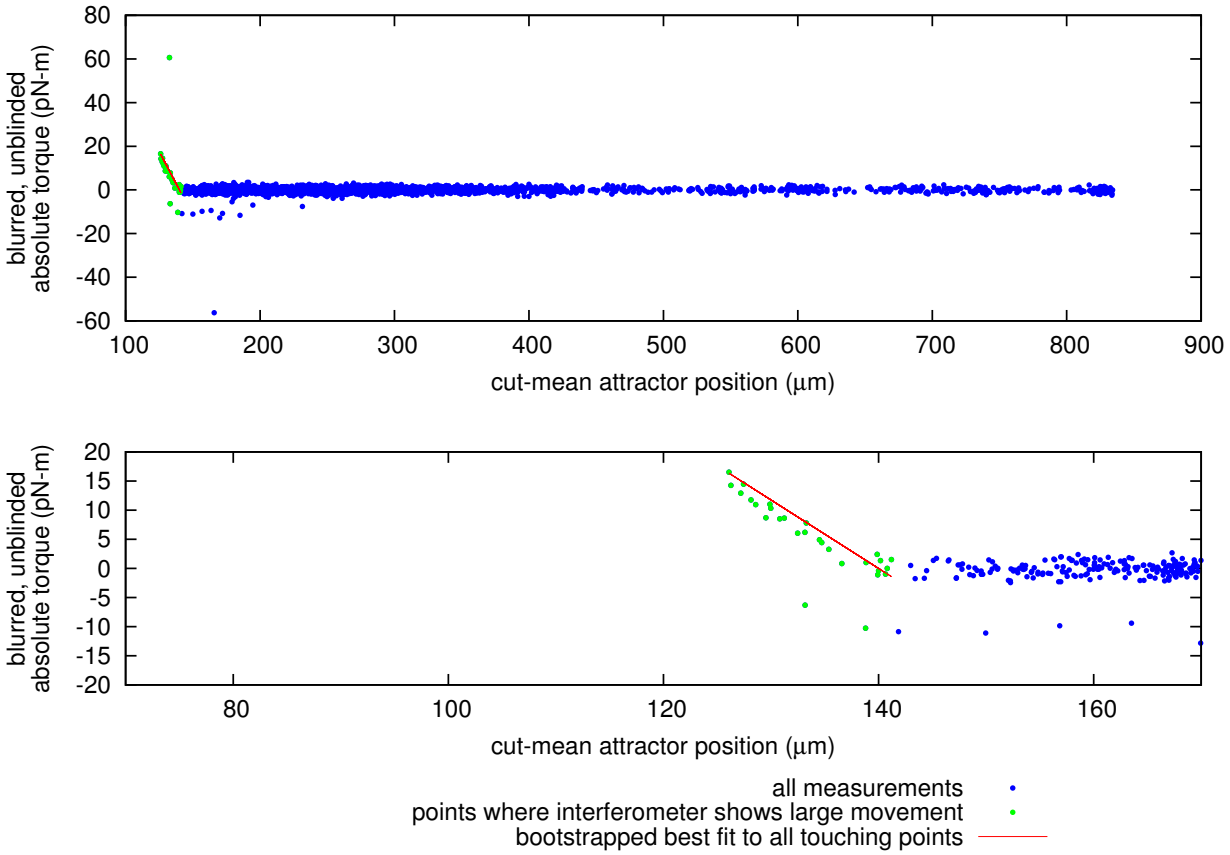

Figure 2.29: Pendulum-Foil Coupling Determination

2.3.9.4 Pendulum-Foil coupling

The critical foil-motion to pendulum-torque coupling was measured by using those science data where the attractor pressed against the foil. As we know from past experience with the interferometer that attractor motion is directly translated into foil motion once the attractor has made full contact with the foil, we can use the attractor position as a proxy for foil motion. In Figure 2.29, we show the absolute torque (no differencing required) measured as a function of attractor position. These data are unblinded, but have been scattered by 1000 fN-m, approximately 100 times the experiment's maximal per-cut error. The points denoted as 'touching' have a clear signal in the interferometer as having been associated with foil displacement. The best fit of $(-1175539012 \pm 170855801)$ fN-m/m is bootstrapped from all of the 'touching' points.

2.3.10 Three+ computers

Great caution was exercised in the design of the data acquisition system. In order to prevent any correlation of attractor-related information with the pendulum readout, the attractor control was placed on an entirely

separate computer. The computers shared only one ground, starred out from the main grounding block (see Section 2.3.11). While the interferometer readout could have been done by the pendulum control computer, it was convenient to use a third computer, with its own data-acquisition board. The data were acquired from these three machines in parallel, each with their own clock. The timing system was arranged to synchronize all three computer clocks every 3141 s to another master NTP server in the lab, later changed to a NIST reference.

The synchronization system (Windows Time) would occasionally fail, and not recover 3141 s afterward. This became a major experimental concern, as we had no built-in synchronization signal, relying on hand-synchronization and attractor-foil interaction for synchronization. A future iteration of this experiment must have an external clock sending optical pulses to all three DAQ systems.

The torsion balance computer used a NI-6259 PCI DAQ card, sampling 32 analog inputs at 25 kHz. Careful attention was paid to the signal grounds. When not used in fully-differential mode, the card allows the use of two analog-input grounds. One ground was referenced to the photodiode readout, and the second to the grounding block. The attractor and interferometer DAQ systems used NI-6211 USB DAQ cards.

In addition, for attractor temperature and temperature-gradient measurements, our room-wide wireless environmental monitoring system was used to read out the attractor thermistors, making for a "fourth" computer.

2.3.11 Grounding

Good electrical grounding is essential. After years of chasing ground loops, we embarked on a grounding purge. A large copper plate was screwed to a nearby plywood wall and defined to be "Ground". The plate was connected by a heavy-gauge braided cable to a metallic connection with a nearby 100-ton cyclotron magnet yoke. From this grounding plate, we implemented a starred [85] grounding pattern. All circuits made exactly one grounding connection to this plate, all through braided copper grounding straps. It is possible to disconnect the grounding strap from the vacuum apparatus and measure >20 MΩ resistance to ground. The vacuum pumping system was electrically isolated by a ceramic vacuum break above the turbo pump. In-vacuum electrical components which required a floating ground were floated and had signal grounds attached outside the apparatus. Control of the pendulum feedback electrodes was done using battery-powered instrumentation amplifiers placed just outside the apparatus in order to simplify grounding to a comprehensible few-component arrangement. The connection from the pneumatic drive to the experiment passed through a short section of plastic tubing. The foil interferometer readout is, beautifully, entirely optical; the nearest associated electronic components were more than a meter from the apparatus.

For the thermal tests, our group's custom wireless XBee sensor readout systems allowed guaranteed ground-loop free sensing of the attractor thermistors. Attractor heating was done by battery-powered systems. Though they were disabled for this iteration of the experiment, the "contact potential pails", battery-powered, externally motorized potentiometer resistive dividers, had the capacity to set the foil and attractor voltages directly relative to the pendulum's ground potential.

In addition to a defined grounding arrangement, paring down the number of sensors to the minimum set required greatly simplified the traceability of system grounds. All of this paid off, as seen in Section 9.4.

2.4 Enumeration of Uncertainties

There are two forms of experimental uncertainty. Statistical uncertainty is random and decreases with additional measurements. Systematic uncertainty is intrinsic to the implementation of an experiment, and does not improve with time. In the context of PlateWash, only attractor-correlated systematic effects are of great concern.

2.4.1 Statistical Uncertainty

There are many sources of fluctuating 'noise' that affect the PlateWash torsion balance. This section is intended to give a general overview of the many sources of fluctuating torques that are of concern to the measurement. Of greatest concern are those sources that appear at small pendulum-foil separations, as they are dominant in any short-range gravity test.

2.4.1.1 Intrinsic fiber damping

As Saulson and González neatly described [86, 87, 88], all mechanical oscillators have intrinsic thermal noise. Through the fluctuation-dissipation theorem, any lossy oscillator at finite temperature drives itself with thermal fluctuations. The intrinsic noise in a torsion balance is known to be consistent with thermal noise from "internal damping", that is, that the spring constant has an imaginary part. $I\ddot{\theta} = \kappa\theta \rightarrow I\ddot{\theta} = \kappa(1+i\frac{1}{Q})\theta$, where κ is the torsion constant of the fiber, I the pendulum's moment of inertia about the fiber axis, Q the oscillator quality factor, and θ the pendulum twist angle. This is in contrast to more-familiar external damping, where $I\ddot{\theta} = -\kappa\theta - \frac{\sqrt{\kappa I}}{Q}\dot{\theta}$.

A torsion balance suspended in sufficiently high vacuum is 'internally damped'. Following Saulson's classic explication, internal damping gives rise to torque noise power with a $1/f$ spectral shape.

In torque units, the expected thermal noise power is

$$\tau^2(f) = 4k_B T \frac{\kappa}{2\pi f Q}$$

2.4.1.2 Gas damping

Origin As we considered in a detailed investigation of "squeeze-film damping" or "proximity-enhanced gas damping" [78, 89, 90], torsion balances in imperfect vacuum are damped by the residual atmosphere. Microscopically, there is only one form of gas damping - collisions between gas particles and an oscillator allow momentum and energy exchange. Experimentally, when the particle mean free path is larger than the dimensions of the chamber, gas damping appears in two forms: 'free' damping and 'proximity-enhanced' damping. Free gas damping is independent of an oscillator's location in a vacuum chamber, and depends only upon properties of the gas and the dimensions of the pendulum. For a rectangular pendulum with moment of inertia I, resonant frequency f_o, width l, and height h immersed in a gas with pressure P, temperature T, and molecular mass m:

$$Q_{unconstrained} = \frac{32\pi I f_0 \sqrt{8k_B T/(\pi m)}}{l^3 h P}\alpha.$$

There remains some disagreement in the literature regarding the coefficient α, which takes values of $16/\pi$, 16, and 24. For both theoretical and experimental reasons, we prefer 16 [78]. For PlateWash the final operating pressure was 30 μPa, implying a $Q \sim 1 \times 10^6$.

Proximity-enhanced gas damping is geometry dependent. It arises in regions of constrained gas flow. Restricted volumes of gas that are compressed/rarefied by oscillator motion have slight over/under-pressures that oppose oscillator motion. If the restricted volume's equilibration time τ_{eq} is comparable to the oscillator period T, the effect becomes lossy. If the constrained flow is determined by a distance of closest-proximity, d,

$$\frac{1}{Q_{PED}} = \frac{1}{Q}\left(\alpha + \frac{\sqrt{8k_B T/(\pi m)}\tau_{eq}(d)}{d}\right)$$

where $\tau_{eq}(d)$ is, in general, a complicated function of oscillator/chamber geometry. The scaling of $\tau_{eq}(d)$ is approximately a fractional power law for most geometries, with connections to an underlying Lévy walk of the gas molecules. Analytical approximations exist for certain geometries, but in any application where these effects matter quantitatively, experimentation and accurate Monte Carlo simulation are essential. Our experiments with PlateWash showed that, for small d, $Q \propto d^{1.6\pm0.3}$; our simulations predicted $Q \propto d^{1.85}$. At

the pressures and separations used in this inverse square law test, the implied gas damping Q is approximately 800. As the pendulum required electrostatic feedback to remain stable at short pendulum-foil separations, we could not test this prediction directly.

The torque noise from free gas damping is white; there is no torsion-balance-accessible characteristic timescale. The torque noise from proximity-enhanced damping is peaked around τ_{eq}.

With the PlateWash atmosphere, inferring from our Lévy measurements at closest proximity to the foil [78], we expect gas damping-driven noise to be roughly $S_\tau(f) \approx 0.7$ fN-m/$\sqrt{\text{Hz}}$.

2.4.1.3 "Patch effect"

The gold-coated surfaces of our pendulum and foil are some mixture of amorphous and polycrystalline materials. The surface environment evolves over time, as crystals slowly change and residual gas particles adsorbed onto the surface move about [91, 92]. This gives rise to a changing electrostatic environment at each surface. Our pendulum and foil are gold coated in order to minimize this effect - gold's work function is only weakly-dependent on crystal orientation [93]. Once the pendulum-foil separation is small compared to the characteristic size of any metallic patches, we expect this effect to be important.

Estimating or measuring the size of this effect is a challenge. Our pendulums may be more sensitive to this effect than available Kelvin probes. It is difficult to compare with surface-physics experiments due to differing surface preparation and vacuum properties, etc.

Experimentally, all of our group's short-range experiments have seen increased noise at small separations. The patch effect is one of the stories invoked to explain it. Improved noise performance is seen from newly-coated foils immediately after pump-down in both PlateWash and our other short-range experiment, but we don't have a satisfactory nor quantitative understanding of this effect.

2.4.1.4 Contact potential noise

Electrical connections generally form small galvanic batteries [94]. Numerous connections in series between, for example, the pendulum and foil, give rise to a small (< 1 V) potential between them that is fairly stable (< 10 mV/month drift), and initially unpredictable in value. For this experiment, fluctuations in contact potential are indistinguishable from noise from a "patch effect", except, perhaps, through the distance-dependence of the noise.

The influence of any contact-potential noise, either through direct fluctuation or through swing-coupling can be minimized by nulling the contact potential. The force acting between plates of a parallel-plate capacitor with area A, plate separation d, and potential difference V_0 is $|F| = \epsilon_0 A \frac{V_0^2}{d^2}$. For small variations in voltage ($V_0 \rightarrow V_0 + \delta v$), $|F| = \epsilon_0 \frac{A}{d^2}(V_0^2 + 2V_0\delta V + (\delta V)^2)$. Thus, nulling V_0 can minimize the effect of

any voltage fluctuation. Similarly, for swing, where $d \to d_0 + \delta d$, $|F| = \epsilon_0 \frac{AV_0^2}{(d_0+\delta d)^2}$, so the swing coupling is quadratically suppressed by minimizing V_0. We carefully designed battery-powered isolated and remotely-operable power supplies to permit the nulling of V_0, and demonstrated control of V_0 at better than the 100 mV level. Swing parallel to the foil, which we expect to have approximately equal amplitude, should cause much smaller variations in electrostatic torques acting on the pendulum, as only the effective lever-arm of the pendulum may be altered. If the contact potential varies across the foil plane, parallel-swing torques will come primarily from that variation.

Furthermore, the presence of an uncompensated contact potential acts as an anti-spring for the pendulum, pulling the torsion pendulum asymptotically away from its zero position. As the pendulum twists away from a parallel orientation with respect to the foil, the V_0^2/d^2 nature of the capacitive force pulls even harder. Once the capacitive torque is greater than the torsion fiber restoring torque, the pendulum runs away toward the foil. As our sensitivity to new physics increases exponentially with smaller pendulum/foil separation, we are driven to operate the pendulum at the smallest distance at which we can preserve stable operation of the pendulum. Compensation of the contact potential allows us to operate with lower noise at closer distances.

Unfortunately, in tracking down the "capacitor systematic" (see Section 8.3), the contact-potential nulling apparatus was intentionally disabled for the duration of the experiment – the pendulum and foil were directly electrically connected. While we cannot measure the contact potential, past experience suggests that it is $\sim 0.2 - 1$ V. Minimizing the size of this potential is the single most effective change we could make to yield improved performance.

2.4.1.5 Autocollimator noise

The opto-electronic system used to measure pendulum twist has its own inherent noise. By design, the autocollimator's relatively white noise floor is chosen to sit below the low-frequency thermally-driven angle noise of 110 nrad/$\sqrt{\text{Hz}}$. The PlateWash autocollimator's noise floor is approximately 50 nrad/$\sqrt{\text{Hz}}$, as seen in Figure 2.14. Because of the pendulum's $1/f^2$ response function at high frequency, angle readout noise is the primary limit on the instrument's torque sensitivity at high frequency.

2.4.1.6 Environment

A number of environmental factors can give rise to extra noise in the balance. Careful design and construction helps to limit these effects. Many of them are transient in nature, and may be avoided by focusing data acquisition into quiet parts of the day, week, and year.

Ground motion There are several ways in which ground motion can cause an apparent pendulum torque:

1. Rotation of the ground: This rotates the instrument relative to the pendulum and appears as a spurious torque. In practice, this effect is much smaller than other noise sources.

2. 'Tilt feedthrough': Asymmetry in the fiber cross-section converts the tilt of the fiber support into a pendulum twist. This is suppressed, but not eliminated, by the fiber 'prehanger'

3. Magnetic damper asymmetry: If the field of the swing damper is inhomogeneous, and the prehanger mass has a magnetic moment or is insufficiently-centered, damper swing will be converted to a torque.

4. Vertical bouncing: up-down motion of the pendulum changes the tensile stress in the fiber. There is a connection between tensile fiber stress and equilibrium position, so vertical shaking may give rise to apparent torques.

5. Electrostatic interaction with nearby surfaces: For PlateWash, and most short-range experiments, this is the dominant effect. Asymmetric position-dependent pendulum interactions with the foil, control electrodes, and nearby grounded surfaces easily create powerful torques on the pendulum. Careful minimization of contact potentials is essential to reduce tilt/swing couplings. With simple modification and bias voltages, the PlateWash balance can become a powerful tilt-amplifying sensor.

Seismic Earthquakes worldwide can affect the balance. Quakes of magnitude 3+ in the immediate region (Northern California to Alaska) can disturb operation for minutes to hours. Large quakes, such as the 2010 Chilean and 2013 Japanese quakes, disturb Earth so violently that it can take a day or more before aftershocks and low-frequency modes of terrestrial vibration have damped.

PlateWash also sees less-palpable seismic effects, such as the 'microseismic peak', an ever-present seismic signature of ocean waves striking the continental coastline, and the effects of nearby humanity (traffic and laboratory noise).

Tilt The effects of tilt are almost indistinguishable from other ground motion in PlateWash. Qualitatively, we expect ground rotation to be our dominant source of external noise at sub-mHz frequencies. As an example, on sunny days, we see a marked change in torque at sunrise when sunlight and its associated heating reaches the hillside in which our laboratory is embedded. Tilt is also the experiment's strongest coupling to human activity. Walking up to the apparatus deforms the floor, leading somehow through the instrument to a tilt coupling and an associated torque. When initially characterizing the apparatus, we determined that this effect was not gravitational in origin by suspending a $\sim$ 40 kg lead mass from our laboratory crane: lifting the mass only a millimeter off the floor was sufficient to produce an instrument response comparable

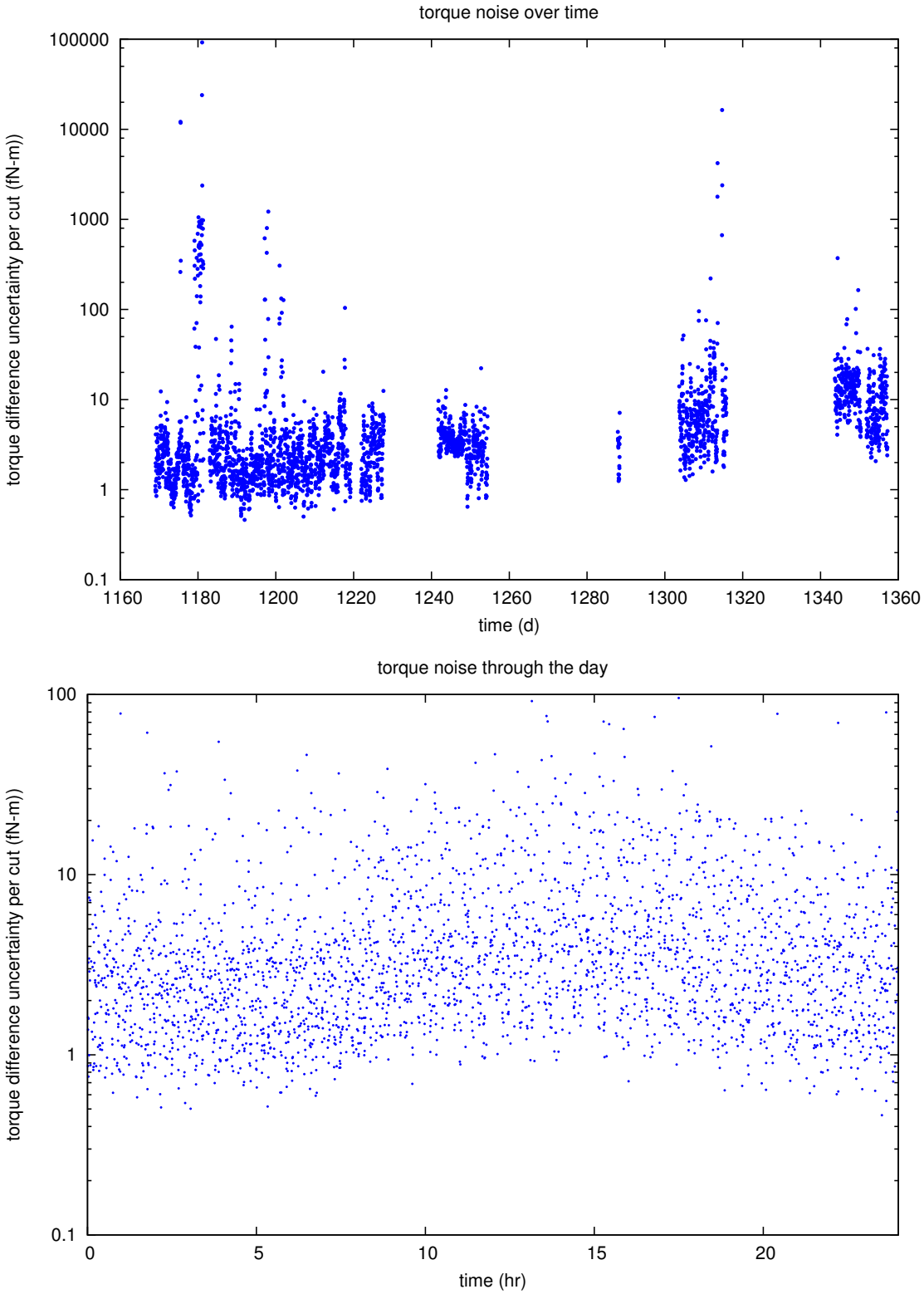

Figure 2.30: Daily noise variation, in torque noise per cut

to that seen from body weight. Lifting it higher had negligible effect. As with many of our external sources of noise, tilt noise in the lab is least at night, giving us a daily 6-8 hr window for our best measurements.

Gravity gradients The majority of Eöt-Wash pendulums are designed to have vanishing low-order gravitational moments. PlateWash, on the other hand, has a sizeable q_{22} moment (it's a flat plate, with masses on the ends). Recent estimates by Krishna Venkateswara and Todd Wagner put the ambient gravitational gradient noise at less than < 0.1 E/$\sqrt{\text{Hz}}$. This noise level converts to torque noise of $<$0.4 fN-m/$\sqrt{\text{Hz}}$, well below the lowest instrumental noise.

Wind While the coupling mechanism to wind is related to some combination of seismic effects, tilt, changing gravitational gradients, or pressure, we aren't certain which effect is dominant. What is clear from our data is a strong coupling between outside wind and pendulum noise. Stormy periods can elevate torque noise by as much as an order of magnitude in comparison with still periods.

Connection to contact potential As elucidated above, the electrostatic environment near the pendulum is important to these vibrational noise sources. The largest and closest surface to the pendulum is the foil. Minimizing the pendulum-foil contact potential, V_{PF}, can have considerable impact on pendulum noise. Because of a non-constant contact potential as a function of pendulum angle, we were unable to trim V_{PF} as well as should be possible in principle. Furthermore, as seen in Section 8.3, we chose to give up control of V_{PF} in order to eliminate a possible systematic effect. With further investigation, the systematic was traced to a bypass capacitor. Direct grounding of the foil is no longer required, and we will re-enable the V_{PF} control. Future pendulum designs and future iterations of this experiment must minimize V_{PF}.

Temperature The PlateWash apparatus is sensitive to temperature. The fiber's equilibrium position and drift rate are perhaps most temperature dependent, but there are many temperature sensitive elements in the apparatus. To limit the effects of temperature variation, the instrument is located in a temperature controlled room. The room temperature varies by less than 1 degree through the year and $<$0.2 degrees daily.

As passive isolation, the vacuum system is an excellent insulator, with a time constant of days for the internal aluminum baseplate with respect to the outside world. As further, rather limited, insulation, the clean-room fans are turned off during measurement periods, trapping a volume of air around the apparatus.

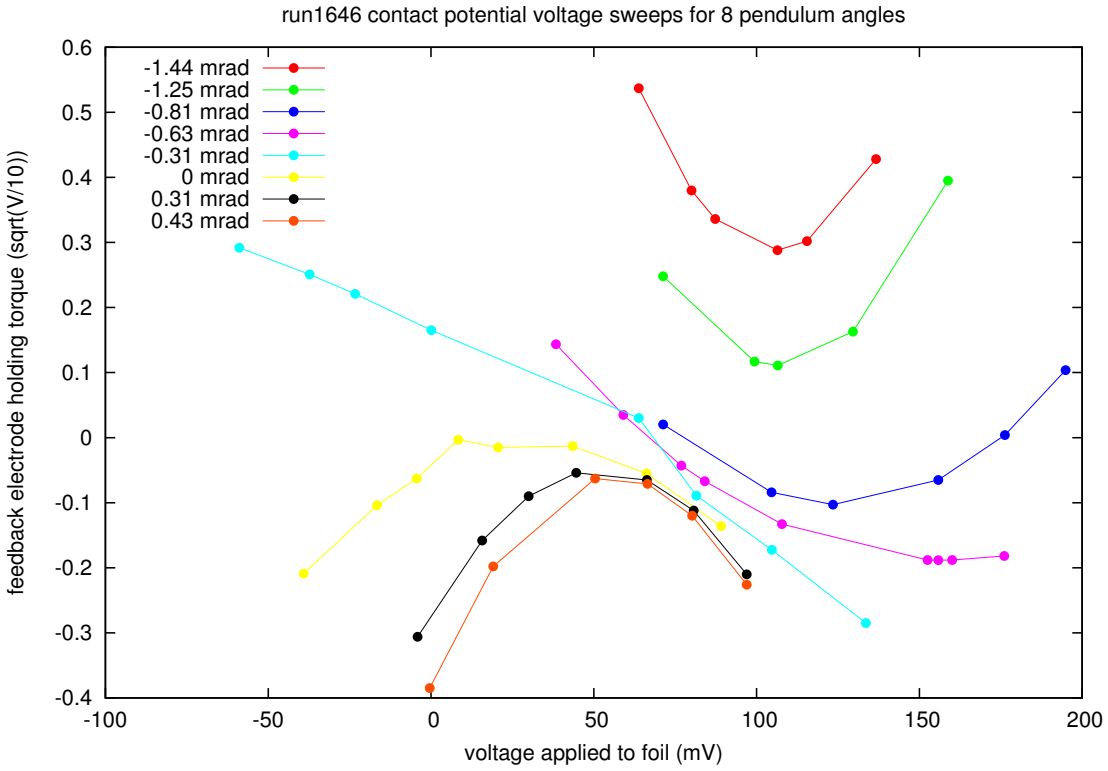

Figure 2.31: Evidence for a "pendulum battery", taken with an older pendulum and foil. The plot displays pendulum torque as a function of an applied pendulum-foil voltage. Each set of points is taken at a different pendulum angle. In the absence of a pendulum-battery-like effect, the vertices of the approximate-parabolas should align vertically, and the curvature-free curve of holding torque as a function of foil voltage (-0.31 mrad above) should be horizontal. The apparent magnitude of the "pendulum battery" is ~ 50 mV.

2.4.1.7 Instrumental noise

Many electronic effects can create excess noise. Most pick-up effects are indistinguishable from autocollimator noise. Because PlateWash normally operates in electrostatic feedback, there is a second channel through which EMF pickup, ground loops, and finite digital-to-analog conversion noise can couple to the pendulum. These effects were assessed by comparing the pendulum noise with and without the feedback electrodes grounded. No excess noise was seen.

2.4.2 Systematics

Systematic uncertainties are the primary challenge of this measurement. Ensuring that no gravitational signals are mimicked nor masked demands great care. This section enumerates the many ways in which the experiment could lead us astray. Approximate quantitative estimates are given where calculable. We will return to the subject in Chapter 9, where the operation of the experiment further informs our understanding of systematic uncertainties. For scale, torques of $0.1 - 1$ fN-m are of importance to this measurement.

2.4.2.1 Electronic Crosstalk

Any effect that is not directly related to attractor motion, but is aware of attractor position, lands in this catch-all category. In general, these systematics are very important in many experiments. In this case, the experiment includes an isolation valve on the air pressure line to the attractor bellows. The valve is positioned downstream of the electronic pressure gauge. The valve may be closed, and every electrical signal associated with the operation of the experiment is unchanged. The data analysis system is also unchanged, making it possible to test for these systematics directly. This ability is the greatest strength of the pneumatic drive system; any electrically-actuated drive cannot separate electrical crosstalk in this way.

The electro-optical encoders mounted atop the attractor flexure are disconnected in normal operation of the experiment.

2.4.2.2 Foil-related

Foil displacement Any attractor-correlated displacement of the isolating foil is a cause for concern. The pendulum interacts electrostatically with the foil with a coupling of $(-1175539012 \pm 170855801)$ fN-m/m. See Section 2.3.9.4.

Dust Any dust between the attractor and foil will cause the foil to move once the attractor-foil separation is smaller than the dust's characteristic size (Pendulum-foil dust is assessed via the "bounce test").

Effort has been made to keep the active surfaces dust-free through the use of clean room procedures. We see the effects of dust in perhaps 50% of our initial cleanings of a surface. The presence of dust is inferred with the interferometer; if dust is present between the attractor and foil, it can cause a weak displacement of the foil before the attractor touches the foil directly.

Attractor-foil contact potential The attractor and foil will, in general, have different values of electrical potential. The electrical connections between them will include a collection of Galvanic-potential generating junctions. The voltages generated at each junction are, as in a thermocouple, temperature-dependent and may depend upon the detailed structure of the junction, leading to temperature sensitivity, temperature-gradient sensitivity, changes with mechanical creep, and perhaps microphonic pickup.

To be safe from the effects attractor-correlated contact potential variation, the pendulum/foil contact potential cannot change by more than 1-10 nV in an attractor-correlated way. This voltage is hard to measure in any way but with a torsion balance, as the effective output impedance of any contact-potential generating phenomenon will be large (cannot drive much current), and the input impedance of the torsion balance as a voltmeter is very small (with the LISA test apparatus, we measured TΩ resistances of a quartz fiber). In defense of the experiment, no part of the moving attractor assembly is in contact with the pendulum or foil support structures. If an effect were to appear, it would most-probably be generated by somehow displacing the wires that connect the foil and pendulum to ground, or by shifting the "ground" potential itself.

Residual gas pressure The diffusion time for gas molecules out of the attractor-foil gap is ~ 40 ms [78]. As we wait more than 1 s between moving the attractor and taking data, any remaining pressure gradient should be, at most, comparable to the ambient pressure itself. The displacement, w, of a tensioned circular membrane in response to uniform pressure is given by Campbell as

$$w(r) = \frac{a}{4}\frac{pa}{\sigma h}\left(1 - \frac{r^2}{a^2}\right) \rightarrow w(0) = \frac{pa^2}{4\sigma h} = \frac{pa^2}{4T}$$

where a is the membrane radius, p is the applied pressure, h is the thickness, σ is the initial tensile stress, and T the tension, computed in Section 2.3.2 [95].

30 μPa acting on a (74.85 ± 0.05) mm diameter circle yields a foil displacement of 4 pm and hence a torque of -0.004 fN-m.

Modulation of ambient foil motion The foil is in constant motion, especially in its resonant modes. The foil has a Q of thousands or more. If those vibration amplitudes change with attractor position, a false signal may emerge in the pendulum. The ambient modes can be damped by proximity-enhanced gas

damping [78, 90] or by electrical effects.

Proximity-enhanced gas damping between a oscillating plate and a nearby surface is known to produce a retarding force [96, 78, 90]

$$F \approx -\frac{AP}{d}\frac{\omega^2\tau^2}{1+\omega^2\tau^2},$$

where P is the pressure, A the region of closest proximity between the surfaces, d the separation between them, ω the oscillator angular frequency, and τ the diffusion time for gas to leave the gap. This approximation buries the complexity of the problem into the estimation of τ. While the best-known approach is Monte Carlo, a rough approximation for the relevant circular geometry can be found [90] using

$$\tau = \sqrt{\frac{\pi}{2}}\frac{R^2}{dv_T \log\left(1+(R/d)^2\right)}$$

where the thermal velocity $v_T = \sqrt{k_B T/m}$, R is the attractor radius, d is the attractor/foil separation, and m is the ambient gas molecule mass.

The quality factor, in this approximate model, is then

$$Q = \frac{T_{foil} d}{AP}\frac{(1+\omega^2\tau^2)}{\omega^2\tau^2}$$

where T_{foil} is the foil tension. The expectation for PlateWash is plotted in Figure 2.32.

Perhaps the most-likely electrical effect is the varying capacitance between the attractor and foil; for a fixed voltage V_0 between plates of area A spaced by a distance d_0 and a small vibrational amplitude B at frequency f, the power dissipated in an electrical resistance R between them is proportional to $1/d^4$.

$$I \approx -\frac{V_0\epsilon_0 A}{d^2} fB \Rightarrow P = I^2R = \frac{V_0^2\epsilon_0^2A^2}{d^4}f^2B^2R$$

If the foil has thickness t, density ρ, and oscillates at a resonant frequency f_0 then the foil quality factor Q is

$$Q = 2\pi f_0 \times \frac{\text{stored energy}}{\text{power loss}} = 2\pi f_0 \frac{\frac{1}{2}\rho A t (2\pi f_0 B)^2}{\frac{V_0^2\epsilon_0^2A^2}{d^4}B^2 f_0^2 R} = 4\pi^3 f_0 \frac{\rho t d^4}{V_0^2\epsilon_0^2 AR} \sim 10^{30\pm2}\frac{d^4}{R} \text{ in SI units}$$

With the installed 20 MΩ resistance between the attractor and the foil, we should expect this electrical loss to become larger than the intrinsic loss of the foil when the attractor is within ~ 10 μm of the foil. This estimate makes an unfounded assumption of a low-impedance current source in the contact potential and bounds the loss from above. If the contact potential cannot sustain V_0, the loss would be much less.

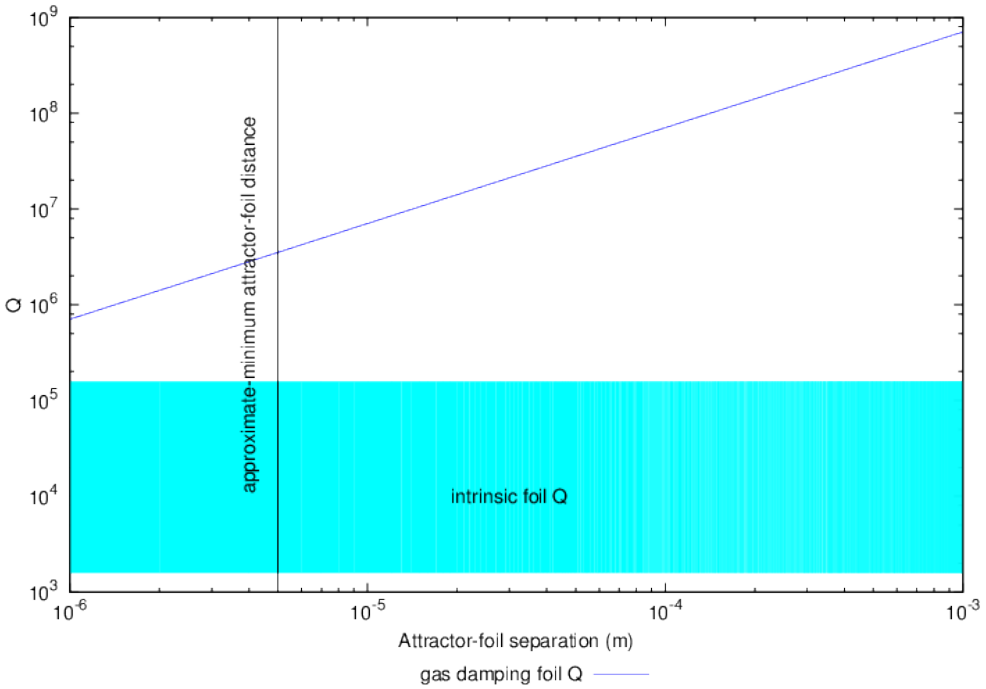

Figure 2.32: Proximity-enhanced gas damping as a function of attractor-foil separation. If the gas-damping quality factor (Q) is larger than the intrinsic Q of the foil (determined conservatively here from the $\sim$ 10-20 s ringdown time of the foil), gas damping will be an inconsequential source of systematic uncertainty.

Attractor vibration coupling via gas or electromagnetic effects to foil From the interferometer, we can see that the attractor, at the end of its spindles, vibrates with 60 Hz and its harmonics considerably more than the foil alone. The attractor must also have its own seismically driven vibrational modes. It is possible to place limits on the appearance of both kinds of modes using the foil interferometer; the raw data needed to do so were collected, but analyzing these data remains a substantial effort, and will not be reported here.

Interferometer Opto-coupling Foil motion inherently changes the ambient light leaving the vicinity of the interferometer tip. Varying stray light incident on optically sensitive elements of the autocollimator could lead to a spurious foil-correlated signal in the autocollimator.

2.4.2.3 Newtonian Gravity

In the search for non-Newtonian effects, it is essential to understand the effects of "normal" gravity that stem from imperfections in the experiment's design and execution. Some of these effects are essential, such as the necessarily-finite size of the attractor. Other effects have been made small by careful design and construction, but residual effects must be quantified. Uncertainties stated here reflect the difference between a simple linear+quadratic fit and our simulated model. They do not take into account absolute uncertainties

in density. At primary issue is the linearity of residual Newtonian gravity, which these numbers constrain.

"Non-infinite" attractor As the attractor does not have the spatial extent of an infinite sheet, the gravitational field of the attractor has both distance and misalignment dependence. To estimate the size of this effect, we developed a Newtonian modeling software library to compute the forces and torques directly (see Chapter 7). As expected from flat surfaces in the near field, the torque vs. distance curves are predominantly linear. For a centered pendulum, the expected torque per unit attractor displacement is (-1253 ± 1) fN-m/m. Similarly, the quadratic term is (124220 ± 418) fN-m/m^2. There is a left-right asymmetry of the curve due to the asymmetry of the pendulum, but for a displacement of the pendulum (0.76 ± 0.25) mm horizontally, the linear term becomes (-4399.77 ± 0.03) fN-m/m and the quadratic term (221582 ± 12) fN-m/m^2. The uncertainties quoted in this paragraph are derived from fit residuals, and do not reflect systematic uncertainty.

Voids/defects in materials Small non-uniformities in density of the attractor or pendulum materials can mimic or mask a signal. The flat plate geometry of the experiment renders this effect largely second order. In the absence of an attractor non-uniformity, a pendulum void should be expected to have no effect at all. In the absence of a pendulum non-uniformity, an attractor void will cause a lateral position-dependent torque. If the attractor void were to occur near the outer corner of the dense pendulum inlay (worst case), the unmodulated torque would be $< 10^{-24}$ N-m/ (missing cubic micron of attractor); the modulated torque would be smaller. An attractor void sufficient to cause an unexpected torque would be easily eye-visible – we see none, even under a microscope. In the unlikely event that two voids are laterally coincident, voids could be >80 μm cubed before the maximum total torque at closest-approach would exceed the 10^{-16} N-m level.

Horizontal displacement of the pendulum is an excellent potential test for local effects, which we did not do in the reported sequence of runs. We did, however, do horizontal displacements prior to beginning this sequence, without breaking vacuum in between; no prominent effect was observed. As we observe no voids in the surfaces under a microscope, the largest voids are certainly <10 μm in depth.

A random distribution of voids across the surface can, with sufficiently-many voids, be viewed as a reduction in the density of the material. This is especially true if the voids are smaller than a characteristic length-scale of interest. We see no signs of a distribution of voids at the > 5 μm level; therefore such a correction is not required for our purposes.

Attractor mount/flexure/bellows gravity The motion of the attractor mount, spindles, flexure, and bellows introduces a small, largely linear, torque background. The linear background is expected to be

(10.699 ± 0.006) fN-m/m. The spindles contribute the dominant nonlinear term, which is, in aggregate, (158 ± 2) fN-m/m^2. It is possible that the bellows could deflect sideways when it expands axially; we do not expect the nonlinearity of this effect to be observable.

Weight of drive gas The flow of gas into the attractor drive bellows changes the mass distribution near the pendulum. 34000 mm^3 of gas at 50 psi has a mass of 120 mg. The expected total torque on the pendulum due to the appearance of 120 mg 180 mm from the pendulum is -0.00012 fN-m. Furthermore, this effect is quite linear with pressure/displacement, as the center of mass of the bellows displaces by 1/2 of the displacement of the attractor and about 1/400th of the distance to the pendulum.

Glue gaps The dense inlays of the attractor and the dense material of the attractor face are attached with a thin layer of epoxy. It is possible that an air bubble may have become laterally wide when the glue layer was pressed. Assuming a 1 cm radius by 5 micron thick glue bubble in both the pendulum and attractor, the associated total torque between them, not torque-difference, is <0.0001 fN-m. With density near 1170 kg/m^3 [97], it's simply not dense-enough to matter.

Spindle screw voids The attractor's aluminum body is attached to aluminum spindles using aluminum set screws. The screws and spindles are matched in mass to better than 1 mg. The screws and spindles are installed such that the mass distribution is most mirror-symmetric about the vertical plane perpendicular to the attractor face and coincident with the torsion fiber. The set screws were screwed all the way into the attractor, then the spindles were screwed onto the screws. In addition, as there are only three spindles, but the threaded attractor holes are six-fold symmetric, we also model the unfilled holes.

The remaining voids in the attractor and spindle housings are expected to give rise to torques of (5.403 ± 0.004) fN-m/m and nonlinearities of (-10 ± 2) fN-m/m^2, where the uncertainties are extracted from fit residuals – systematic uncertainties may be larger.

Surface flatness The attractor and pendulum were lapped to < 10 μm (attractor) and < 25 μm (pendulum) absolute flatness. The surfaces were rounded by the lapping process, which dominates this number. In the central region of both pendulum and attractor, flatness was better than 5 μm (attractor) and 10 μm (pendulum). The lapped surfaces were mirror-like and had centimeter-scale roughness comparable to the limits of our measuring microscope, at the level of a few microns. In addition, the pendulum materials' differing hardnesses led to a step at the dense/light interface of ~ 5 μm. Plots of the surfaces are shown in Figures 2.10 and 2.11.

The torque due to all known flatness variations (roughness, irregularities, and dishing of surfaces) of attractor and pendulum were estimated numerically (Section 7.8.3) to be less than 0.1 fN-m for all possible pendulum positions and azimuthal rotations of the attractor. This estimation was made using the pendulum and attractor scans (Section 7.8.3). Models that include the smoothness of the torque curve can yield a better limit if necessary.

2.4.2.4 Tilt/apparatus deformation

Coupling to tilt is a general affliction of torsion balances. Perhaps because of the non-circular cross-section of fine torsion fibers, all balances exhibit "tilt feedthrough". In short range interaction experiments, the tilt feedthrough is exacerbated by the importance of spatially varying electrostatic potentials. The apparatus has a two-axis tilt sensor, but it was intentionally turned off during data-taking to eliminate any possible electromagnetic interference. The pendulum autocollimator is sensitive along both vertical and horizontal axes, and provides a continuous measurement of pendulum tilt.

Nonlinear tilt feedthrough The apparatus may have attractor-related tilt for a number of reasons, including the displacement of the center of mass of the attractor assembly, deformations of the attractor mount, and changing geometry of the attractor air line. The flexure was mounted from its center in order to minimize any deformation of the apparatus as a whole. The deflection is expected to be both small and linear, limiting the impact on the measurement; we do not anticipate that non-linear tilt feedthrough, particularly nonlinearities that should have anything to do with attractor-foil proximity, will be a major concern. It is a challenge to study this effect experimentally, as the response to small injected tilts will be quite linear, and the response to large injected tilts may be inapplicable.

Flexure non-linearity One interesting systematic effect that can exhibit considerable nonlinearity is the nonlinearity of the flexure itself. Most tilt-related effects are essentially blind to the attractor-foil distance. In order to maximize our use of the flexure's limited (< 2 mm) travel, the flexure's internal travel limit was located not far from the point at which the attractor pressed hard into the foil. If the double-compensating mechanism of the flexure experiences nonlinear stress near the (artificially-placed with mechanical stops) end of its travel, nonlinear stresses might be communicated to the baseplate. We designed the flexure-holding fixture to suspend the flexure from near its center, minimizing the opportunity for unexpected stresses to affect the rest of the apparatus.

The changing effective length of the cantilevered attractor assembly will also deform the experiment's baseplate. A simple elasticity-theory estimate of the bending of the 1.25"-thick, $\sim$ 27"-diameter aluminum

plate due to the placement of a kilogram mass in the center of the baseplate gives a 10 nm total displacement at the center. The attractor assembly weighs less than half a kilogram and displaces by less than a millimeter over the entire range of travel. In any realistic scenario we've considered, the *linear* component of the pendulum's tilt-driven horizontal displacement should be smaller than the nanometer-scale signal we see in the foil interferometer. Any residual nonlinear variation should be much smaller. Without any quantitative way to estimate the change in pendulum-foil separation, we must rest on these estimates.

2.4.2.5 Magnetism

Magnetic systematics are among the most difficult to assess in these experiments. The foil provides limited, if any, shielding against low frequency magnetic fields. Because of the considerable strength of the magnetic interaction, great care must be taken to ensure that magnetic effects do not mimic or mask a gravitational signal. For the shortest-range effects, the torsion balance may be the most sensitive probe for their existence, complicating things considerably. Short range gravity research is becoming increasingly sophisticated in this realm; previous tests have only considered some of these effects.

Magnetic contamination of materials

Point sources Single ferromagnetic impurities may interact with one another or with the finite permeability of the opposing material to produce a force/torque on the pendulum.

Diffuse contamination If impurities are sufficiently numerous, then their presence will only appear in the force/torque once the pendulum/attractor separation is smaller than the characteristic separation of the impurities. Only one surface must be contaminated for this effect to appear, as all materials in the problem have magnetic permeability different from vacuum.

Pendulum magnetic moment The pendulum has a magnetic moment. It may be measured by applying a time-varying uniform magnetic field to the apparatus.

Magnetic materials on the attractor The attractor/flexure/bellows assembly contains magnetic materials, primarily stainless-steel screws positioned more than an attractor diameter away from the pendulum. Any magnetic fields from these sources should vary almost linearly with attractor position,

Modulation of background magnetic fields The magnetic permeability of the attractor and pendulum materials will affect any background magnetic fields. In addition to Earth's field, the largely non-magnetic

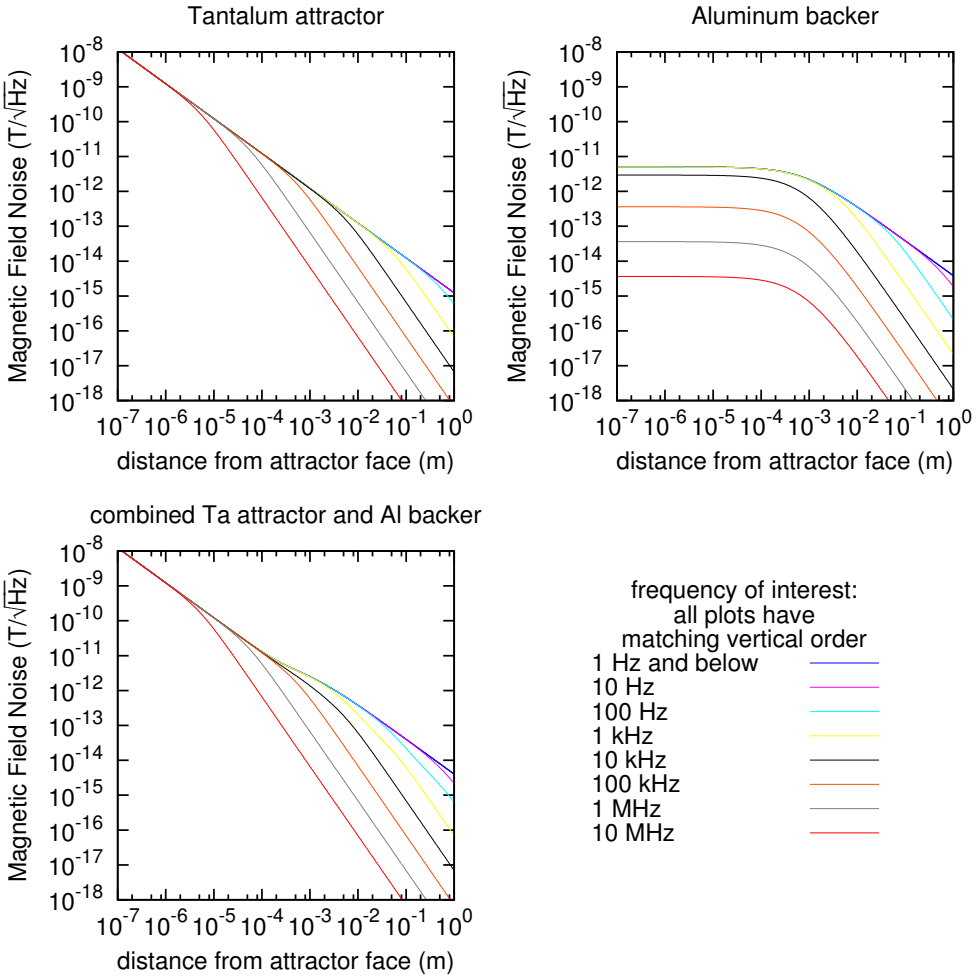

Figure 2.33: Expected thermally-driven magnetic fields, computed using the formalism begun by Varpula *et al.* [98]. The fields are small, but reach easily through the foil.

apparatus will still have a residual field. The magnetic damper is quite close to the pendulum, and is only partially shielded.

As discussed by the group at Stanford [64], fields with a component perpendicular to the attractor surface will be altered by the presence of the attractor. Our magnetic systematic test (Section 9.2) addresses this concern directly.

Thermal currents in materials Charge carriers in any material at finite temperature are in constant motion. These random currents create magnetic fields. They have been experimentally observed in several contexts [98, 99]. The interaction of thermally-generated fields may give rise to a false signal. While it is difficult to estimate the forces due to of such an effect, the expected field strengths are plotted in Figure 2.33.

The thermal fields have low frequency components that will pass through the foil. The skin depth [100] of a material is $\delta = \sqrt{2\rho/\mu\omega}$, where μ is the magnetic permeability, ρ the resistivity, and ω the angular frequency of interest. For Beryllium-Copper, the resistivity is a factor of 2-10 larger than pure copper's 17 nΩ/m [101]. For frequencies below $\sim$ 100 MHz, the isolating foil is effectively transparent to RF.

These thermal fields can produce pendulum torques in several ways:

- Thermal fields induce currents/fields in the opposing mass, leading to an always-repulsive interaction
- Attractor-to-pendulum: direct interaction between the thermal fields
- Attractor modulation of foil fields: thermal fields in the foil can be modulated by the non-zero permeability of the attractor
- Attractor fields interact with pendulum contaminant
- Pendulum fields interact with attractor contaminant

While no previous inverse-square-law experiment of this class has considered these interactions, and we are not certain how to compute their strengths with confidence, we enumerate them here to spur future investigation. The literature regarding the thermal corrections to the Casimir force may hold the keys to determining the magnitude of these forces.

Attractor drift The attractor pneumatic drive system leaks slowly, as seen in Figure 2.34, causing the attractor to retract at 2 nm/s. There are many possible velocity-dependent systematic effects that one might consider. We believe, but cannot prove, that 2 nm/s is sufficiently slow to render such effects negligible.

Attractor vibration Even with a "static" attractor, attractor vibration can induce magnetic effects. As mentioned in the section regarding foil motion, the attractor picks up ambient vibration. Audio frequency motion of the attractor or of attractor impurities can potentially create oscillating magnetic fields that are minimally attenuated by the foil.

2.4.2.6 Thermal

Foil-Attractor temperature difference If there is a temperature gradient between the foil and the attractor, attractor-correlated temperature variations in the foil temperature may occur, giving rise to a false signal. Heaters and thermistors were installed on board the attractor spindles to magnify and study this effect. Heat transfer is enhanced over very short distances in vacuum by near-field electromagnetic

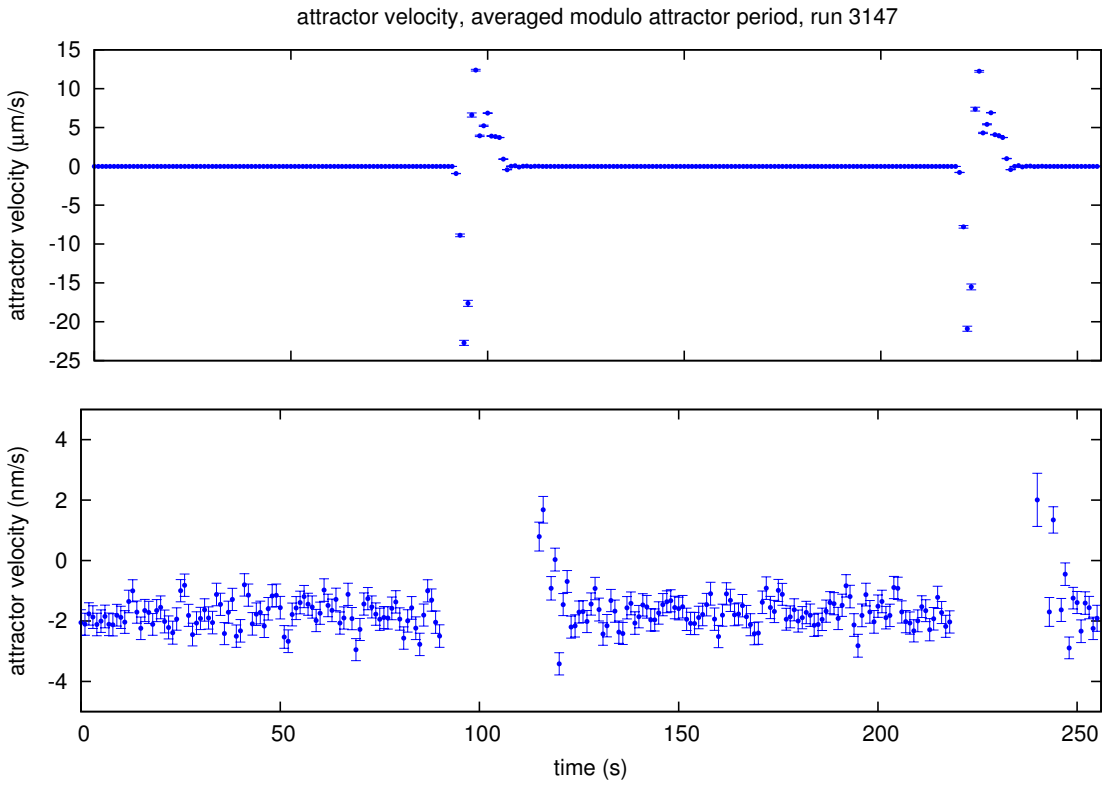

Figure 2.34: Attractor velocity, in 1 s bins

effects, somewhat like the Casimir interaction, as in [102]. Gas effects similar to proximity-enhanced gas damping may also have an effect.

We tested this effect directly by operating with a fixed elevated attractor temperature and with a thermal gradient across the attractor body (see Section 9.3).

Spot-heating by interferometer The foil interferometer has approximately 0.5 mW of infrared optical power incident on the foil. This spot heating may dissipate at different rates in an attractor correlated way. 95-99% of the light is reflected [103], but this leaves 5-25 μW to be dissipated through conduction in the foil. We attempted to test this effect directly, by operating with and without power to the interferometer laser. Unfortunately, a record-keeping oversight prevents us from making any definitive statement regarding the magnitude of this effect.

Overall temperature change at attractor frequency The temperature of the instrument itself may change as gas flows into and out of the attractor bellows. Thermistors are installed (but not continuously monitored) in several locations throughout the apparatus.

2.4.2.7 Pneumatic Drive Leak

If the pneumatic drive system has a leak into the vacuum chamber, it would change the residual gas atmosphere in an attractor-correlated way. While the ion vacuum gauge was turned off during the experiment to preserve low-noise performance, we can get an indirect estimate of any potential leak using the attractor-leakback data (Figure 2.34). The 2.5 cm diameter bellows leaked back at 2 nm/s, most-probably through small leaks in the locking valves. The internal pressure in the bellows was 1-2 atmospheres, so the maximum pressure rise from an attractor leak into the approximately cubic-meter vacuum chamber was $\sim 1\times10^{-9}$ Torr/s, which should be compared with the base pressure of more than 1×10^{-7} Torr. A variation of this size is well below the $\sim 10\%$ resolution of our ion gauge. While it is not impossible for a 1% pressure variation to couple to the pendulum (10^{-9} Torr applied to only one half of the pendulum would yield a torque of 10^{-12} N-m), the indirect path from bellows/gas system to the pendulum and the pendulum's approximate symmetry (spoiled by the pendulum tip, of course) should act to suppress this effect.

We have no way of varying the bellows pressure without moving the attractor, so experimental limits are hard to come by. Displacing the pendulum farther from the foil could reduce any squeeze-film related effects, but it would simultaneously change many other systematic couplings, making the test inconclusive.

2.4.2.8 RF tuning

Attractor motion may change the electromagnetic resonant structure of the inside of the bell jar. This could matter at the lock-in frequency of the autocollimator, altering the magnitude of interference-driven pickup.

2.4.2.9 Casimir effect

The Casimir effect is not expected to penetrate a 10 μm beryllium-copper foil, according to calculations by Chesler, Yaffe, Andreev, and Jaffe [32]. Furthermore, for the attractor-pendulum separations in this experiment, the Casimir force is too weak. As an example, the zero-temperature Casimir pressure on conducting plates ($P = -\pi^2\hbar c/240a^4$) separated by a =80 μm is 32 pPa, while the gravitational pressure between two 50 μm-thick infinite sheets of tantalum (or equivalently, the pressure due to a $\lambda = 50$ μm, $\alpha = 1$ Yukawa interaction) is 270 pPa. Furthermore, a compelling argument due to Eric Adelberger is one of wavelength. The lowest (most-penetrating) modes that participate in any Casimir interaction will have wavelengths typical of the gap between the plates and shorter. As even millimeter wavelength radiation has frequencies higher than 300 GHz, it's unlikely to substantially penetrate a material that becomes impenetrable at frequencies above 30 MHz.

Finally, the symmetry of the PlateWash pendulum acts to suppress any Casimir torque, as the pendulum

is constructed of gold-coated conducting metals (albeit with differing resistivity under the surface). If the pendulum were oriented parallel to the attractor and the resistivities of the pendulum body and inlays equal (Ti: 390 nΩ-m, Ta: 122 nΩ-m [77]), then the Casimir torque on the pendulum would be zero at all separations.

2.5 Typical Behavior

The first eight thousand seconds of torque and attractor data are plotted in Figure 2.35. This represents about 0.1% of the total data used in this analysis. They are presented without any blinding to give the reader a clear impression of the data. The torque peak at 6 mHz is the gravitational calibration tone. The torque spectral response above the pendulum's 13 mHz resonance is under-represented in these plots, as the torque plots are derived from the autocollimator and do not correct for the pendulum response function (see Section 4.2).

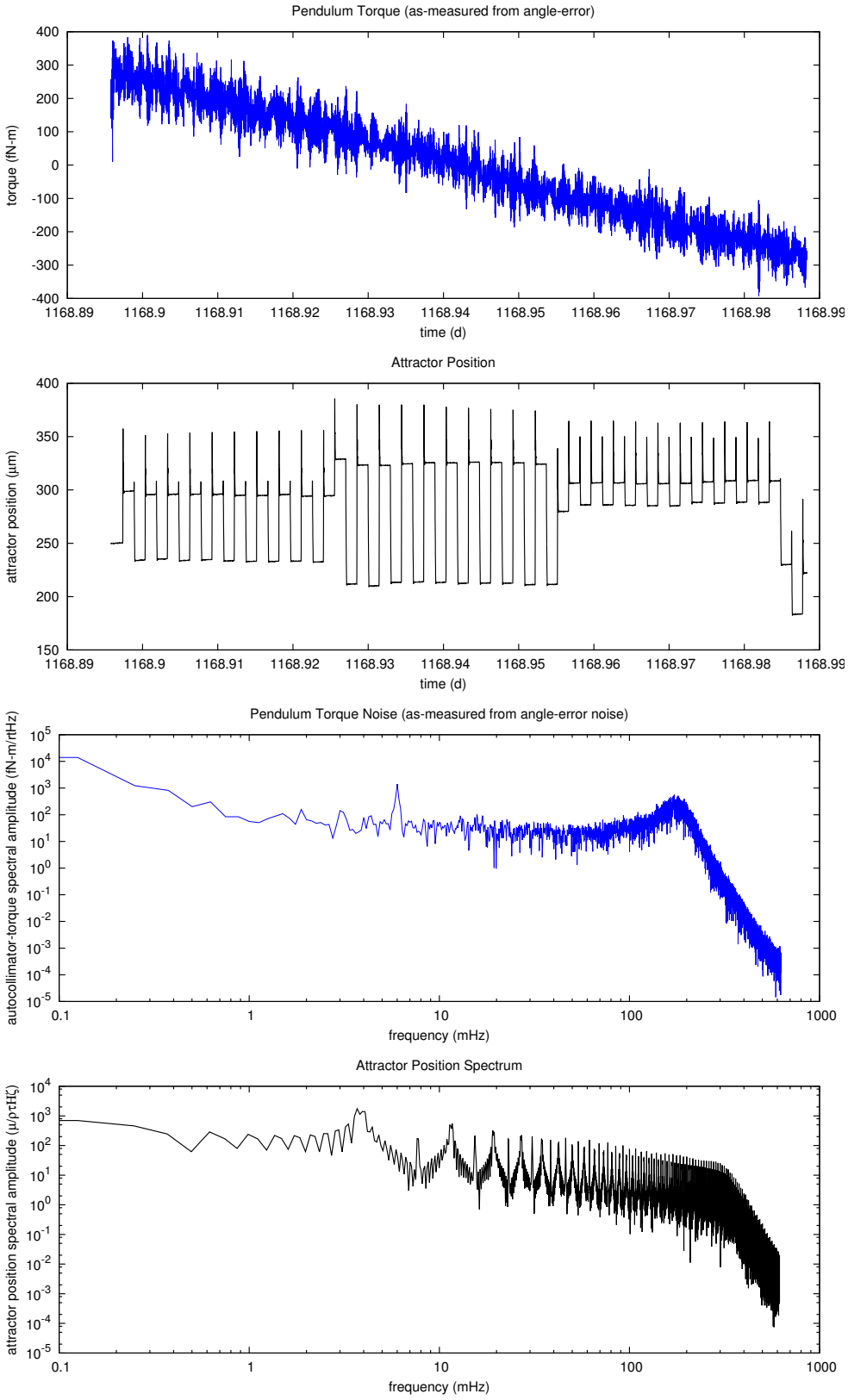

Figure 2.35: The first 8000 seconds of science data, unblinded.

Chapter 3

Metrology

Understanding the separation between attractor and test masses is essential to any inverse square law test. However, because of our experiment's nearly-null parallel-plate geometry, the implications of a sufficiently-small error in metrology affect only the stated exclusion limits, but not the presence nor absence of an ISL violation. This is an important contrast to non-null experiments, where an error in attractor/sensor separation alone can lead to a Newtonian-gravitationally faked or masked signal for new physics.

For PlateWash, metrology is relatively straightforward. We must ascertain the extent to which the pendulum and attractor are parallel and obtain an estimate of their relative separation. As in most inverse-square law tests, this is confounded by the isolating foil stretched between them. As the foil is quite flat (see Section 2.3.2), we break the problem in half, measuring the attractor-foil separation and the pendulum-foil separation.

3.1 Metrology

3.1.1 Initial Alignment

The experiment was initially aligned using the autocollimator. The attractor was fixed in place against two registry blocks and firmly clamped. The autocollimator was then centered on the attractor and aligned such that the light returned to near the center of the sensor. The up-down and left-right angles were recorded. The attractor was then translated through its full range of motion, and the position on the autocollimator recorded, see Figure 3.1. The attractor was adjusted to a point ~ 20 μm from the desired working position. The foil and attractor were cleaned with dry nitrogen and with clean room wipes/isopropanol. The foil was then placed in front of the attractor and iteratively aligned to the attractor using a combination of the

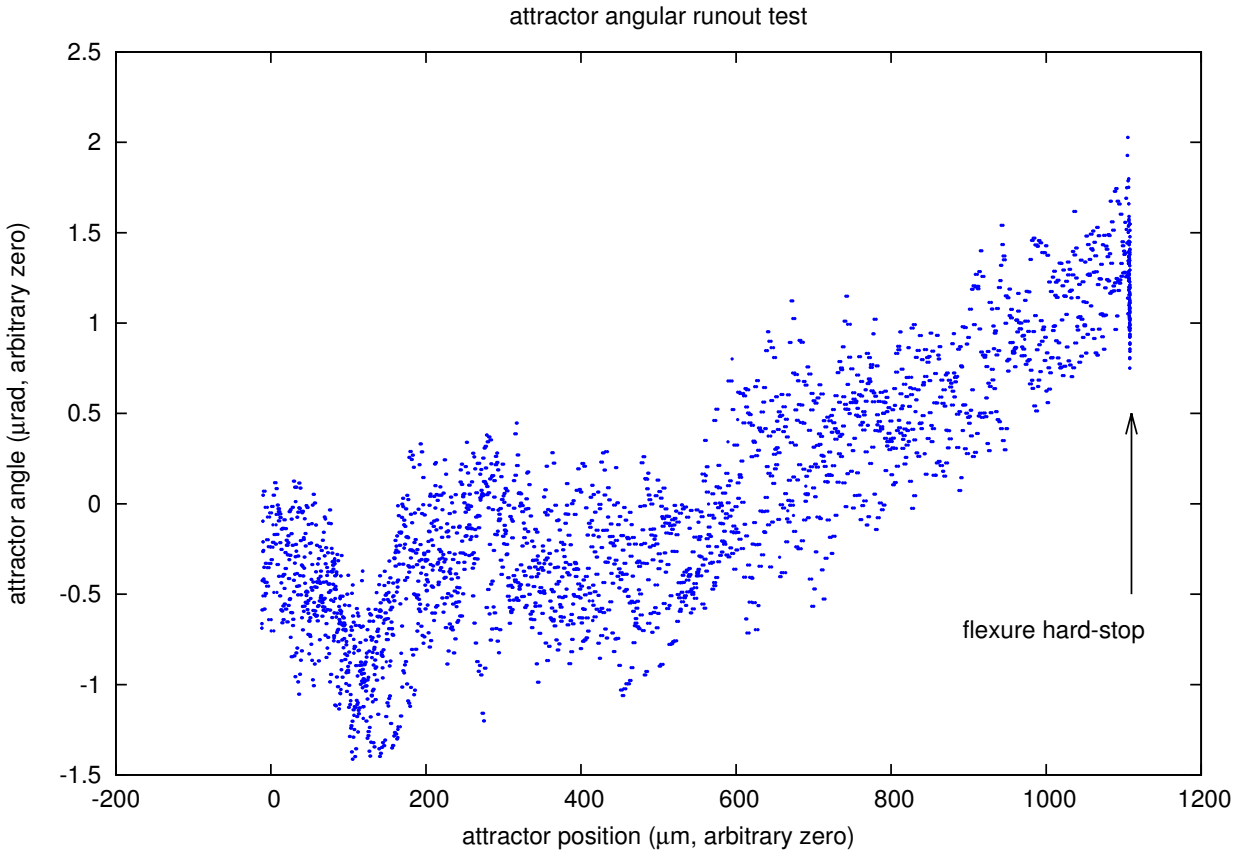

Figure 3.1: Flexure runout, as measured by the autocollimator. The science data were taken using positions largely (perhaps entirely) within this region. The hard stop at right is the physical limit of the flexure's motion.

autocollimator readout and a measurement of the attractor-foil capacitance. Gross alignment was done by hand. Fine adjustment was done using the Belleville washers on the foil mount. When optical alignment reached $\sim$ 100 μrad, a capacitance of $\sim$ 3-400 pF was achieved, and the capacitance seemed to be at a minimum as a function of attractor-foil angle, the system was declared to be coarsely aligned.

3.1.2 Cleaning

Once the attractor and foil were aligned, it was essential to guarantee the absence of debris/dust in the gap between them. To characterize the amount of dust in the gap, and to verify alignment, the attractor-foil capacitance was measured as a function of attractor position, including positions where the attractor touched the foil. Touching was, at this time, determined by noting the position where the capacitance reading became unreliable. The maximum capacitance recorded without signs of touching gave a proxy for the minimum attractor-foil separation.

The attractor assembly was removed and the attractor face cleaned with isopropanol, clean room wipes,

and dry nitrogen. The same was done with the attractor side of the foil. Both attractor and foil were visually inspected for dust. The attractor assembly was replaced against the registry blocks and re-clamped. The attractor-foil capacitance was measured as a function of attractor position. The cleaning procedure was repeated until no gains in capacitance were observed with subsequent cleanings. Typical maximum capacitances were $\sim$ 2 nF; because these measurements were carried out at atmospheric pressure, they should not be used to determine the distance of minimum approach.

3.1.3 Online Alignment

The critical distances between and relative orientations of the pendulum, foil, and attractor were determined with the bell jar closed, under vacuum, and after bake-out. Once the system's atmosphere was evacuated, both the pendulum's and foil's motion were essentially unaffected by gas-related effects. The foil-alignment motors (see Section 2.3.2) were then adjusted to maximize the highest measurable capacitance-before-shorting between the attractor and foil. The pendulum's tip angle relative to the foil (about a horizontal axis parallel to the plane of the foil) was minimized by imaging through the bell-jar window and the "plateau test", as described in Section 3.4. Finally, the pendulum-foil distance was determined both by imaging and by the 'bounce test' detailed in Section 3.3.3.1.

3.2 Attractor-side

This section assesses the parallelism and absolute separation of the attractor and foil. As the positions of the attractor and foil are determined by rigid metal fixations and kinematic mounting, their relative separation is more-easily determined than that of the pendulum and foil.

3.2.1 Parallelism

We used several methods to assess the parallelism of the attractor and foil. When aligning the apparatus, we tipped the apparatus, placing the pendulum onto the foil, pressed the attractor into the back of the foil, and used the autocollimator-measured angular displacement of the pendulum as a proxy for parallelism. That got us close; the fine alignment was performed capacitively, tipping the foil slightly until a maximal capacitance could be measured without electrically shorting the attractor and foil together. With that complete, we performed the science experiment.

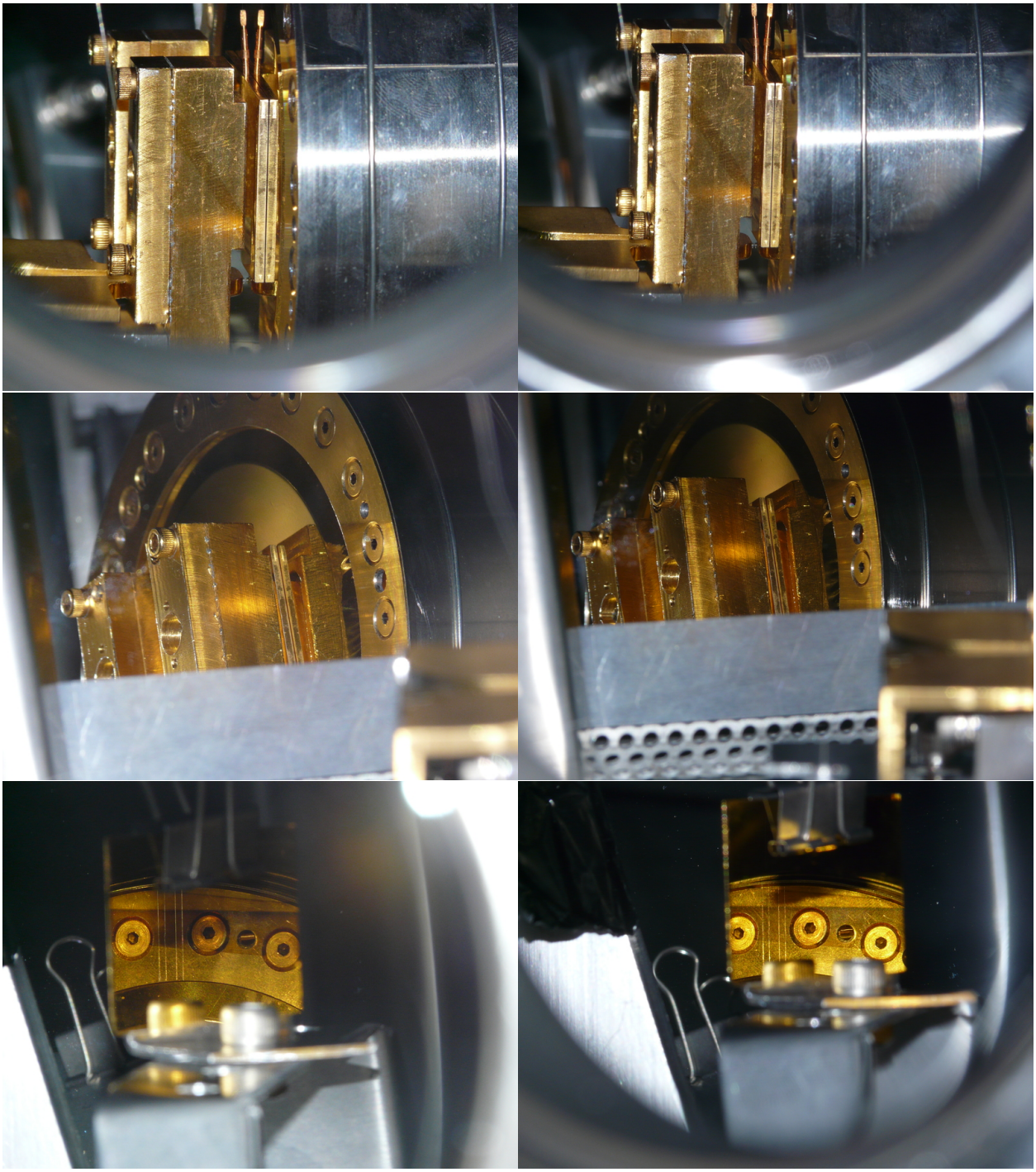

Figure 3.2: Alignment photos before (left) and after (right) science run. The upper photos look directly through the window, the middle photos are the mirror-image of the side opposite the window, and the lower photos are the reflected image of the fiber position against the foil's alignment marks.

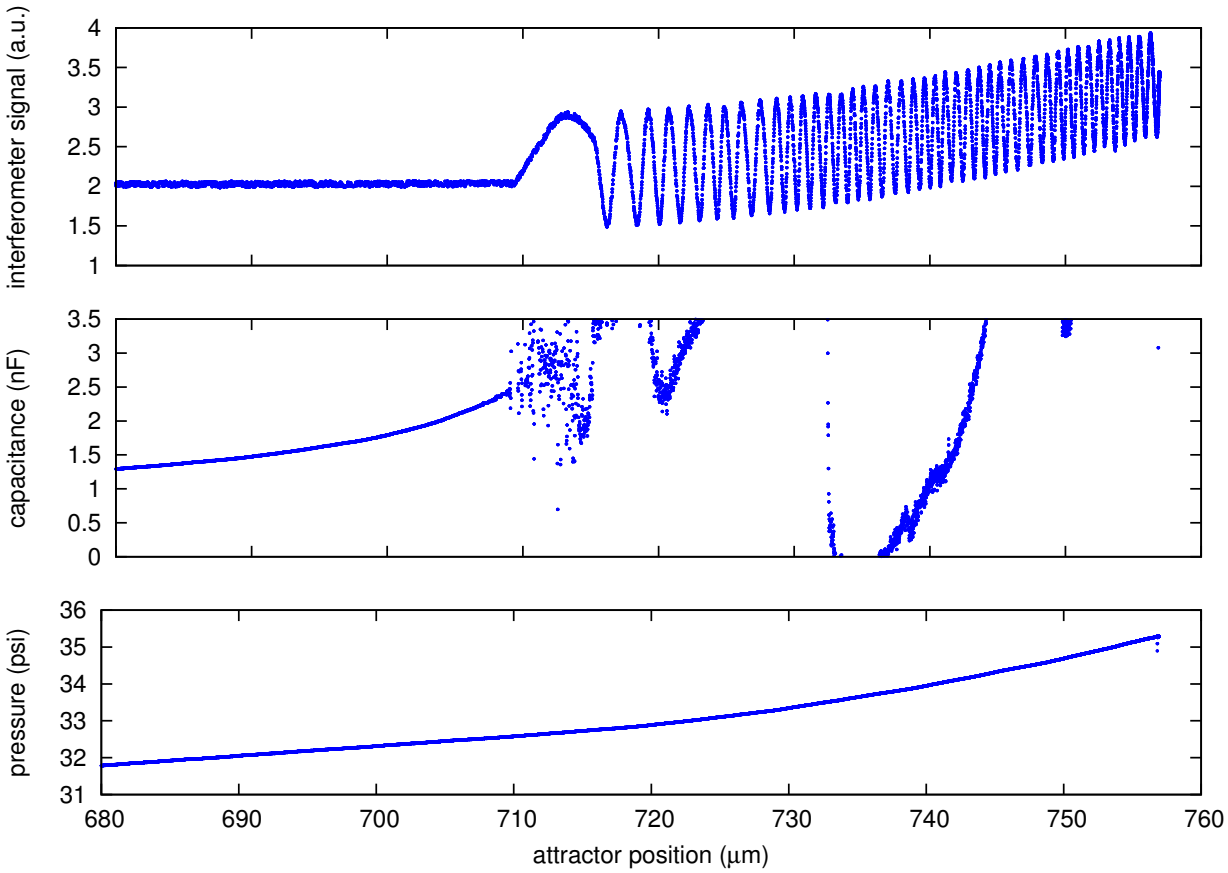

Figure 3.3: Raw data used to infer the attractor-foil relative angle. Each plot is a function of attractor displacement as measured by the optical encoder, with a common arbitrary zero. The upper panel shows the interferometer signal, with prominent touching signal at displacements greater than $\sim 708\ \mu m$. The middle panel shows the capacitance signal, which becomes unreliable in the same location. The bottom panel shows the pressure in the flexure bellows; the slope should increase when the attractor makes contact with the foil.

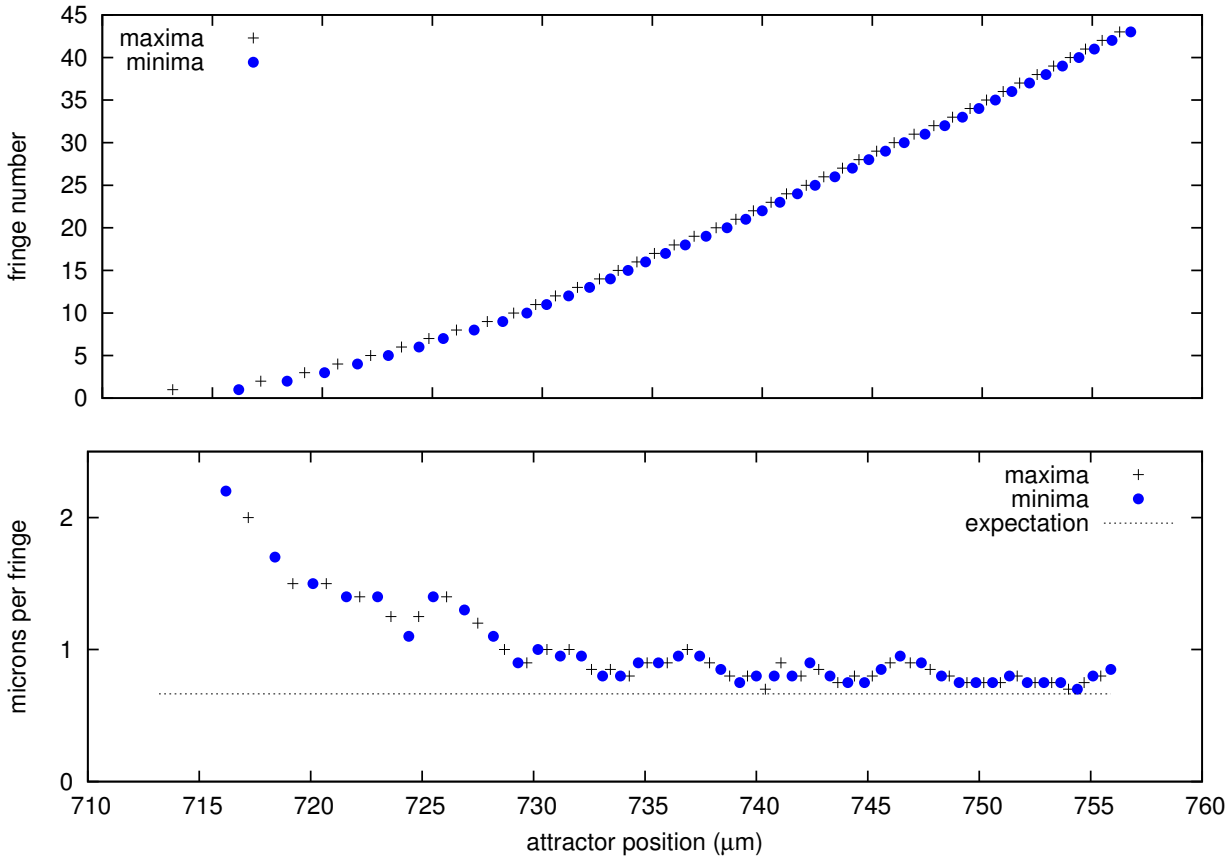

Figure 3.4: Fringe locations taken from the upper panel of Figure 3.3. The upper panel shows the locations of maxima and minima. The lower panel shows the displacement per fringe. It should approach $\approx \lambda/2$ as the attractor makes full contact with the foil. The range of displacements where the displacement-per-fringe value is significantly different from $\lambda/2$ should correspond to incomplete attractor-foil contact, and hence the misalignment of the two surfaces.

Concerns regarding the difficulty of calculating fringing fields led us to view the capacitance measurement with skepticism. In January 2015, the untouched apparatus was brought back to life. The attractor and interferometer computers were turned on, and the interferometer laser and SRS SR720 capacitance meter were reinstalled. Unlike normal science operation, where the encoder can inject systematic interference, it was possible to connect and turn on the attractor optical encoder. The attractor bellows was pressurized to push the attractor into the foil by tens of microns, and allowed to leak slowly back over several hours, producing a smooth slow displacement over time. The results are shown in Figures 3.3 and 3.4.

From Figure 3.4, we conclude that the overall non-parallelism of the attractor and foil is about 20 μm over the (74.85 ± 0.05) mm diameter attractor. It is conservative to add the 20 μm spacing directly into our computed attractor-pendulum separation. As seen in Figure 3.1, the variation in attractor angle throughout its range of travel is far smaller than the otherwise-fixed misalignment.

3.2.2 Attractor-Foil touching

In addition to the attractor-foil parallelism, we must know the absolute distance between the attractor and foil. This is determined through observing the attractor-displacement position where the attractor and foil touch. This is essential both for metrology and to provide a sharp cutoff against any foil-motion related systematic effects.

The determination of the touching point is done using the science data from the interferometer and attractor position on a per-measurement level, rather than per-cut (See Section 2.1.1 for a brief definition of a "cut" or Section 5.1.2.5 for full detail). In contrast to all other analyses, the mean cut position is not used. Instead, as each cut is an aggregate of individual measurements, we use the minimum position of individual measurements within a cut, as plotted in Figure 3.5.

3.3 Pendulum-side

The most-challenging metrology was on the pendulum side of the foil. The pendulum was hand-lapped, rounding its edges and making its corners challenging to image. The attractor's location was fixed, so the pendulum's horizontal (Section 3.3.1) and vertical (Section 3.3.2) positions were referenced to measurements made prior to the installation of the foil. After the foil was installed and aligned, both the pendulum position (Section 3.3.3) and angle relative to the foil were measured (Section 3.4).

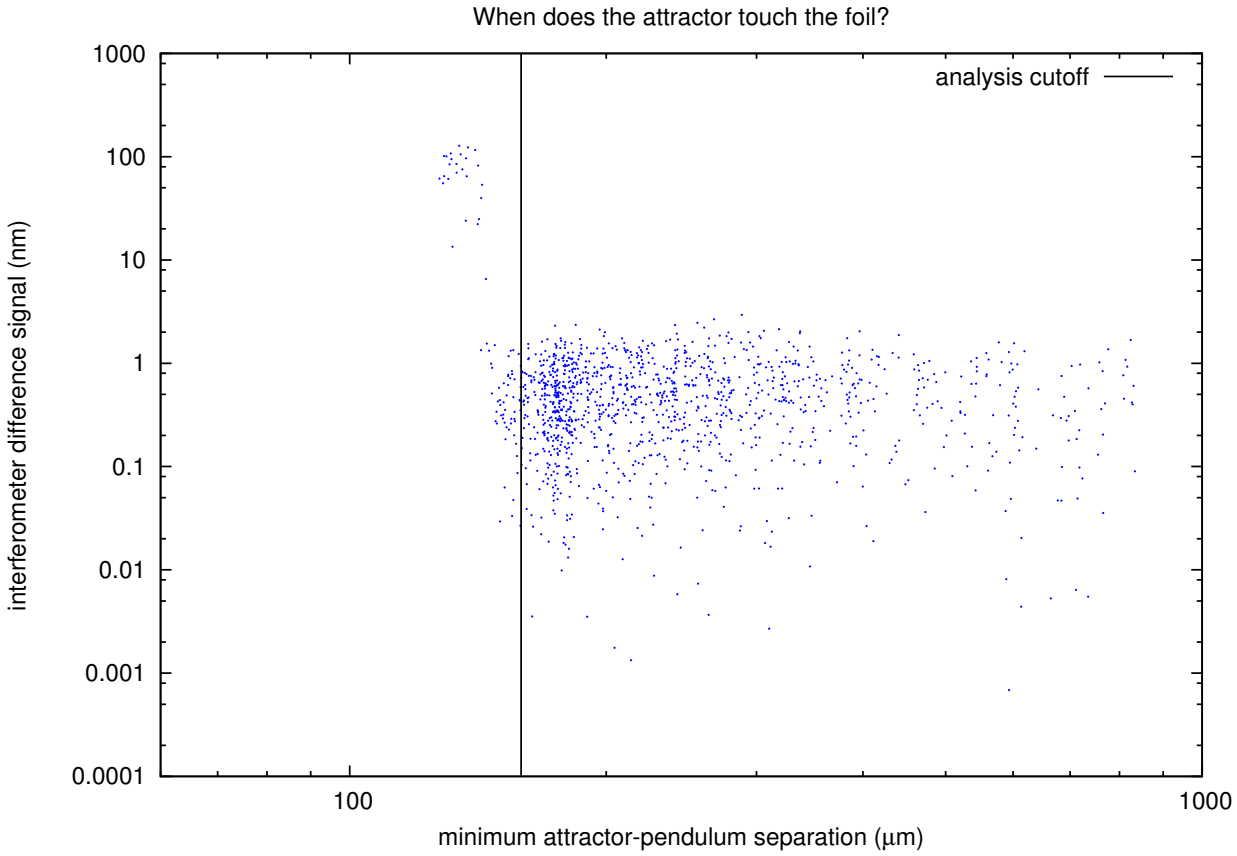

Figure 3.5: Attractor-Foil touching determination. The minimum-accepted attractor-pendulum separation is conservatively determined in Table 3.1.

3.3.1 Horizontal offset

In addition to visual access to the pendulum's separation from the foil, we installed a mirror that allows us to compare the horizontal (left/right) position of the torsion fiber to a ruled grating on the foil-stretcher. When the experiment is initially aligned, the fiber is centered on the top of the attractor, where "top" is denoted as farthest from the plane of the optical bench. The pendulum is translated attractor-axially away from the attractor, and the foil is installed. After the foil is fully aligned, but the fiber has not yet been translated horizontally, the pendulum is again brought near the foil, and the location of the fiber on the foil-stretcher's ruled grating is photographed via the viewport and the dedicated mirror. Once this is complete, then any horizontal translation of the fiber can easily be compared with this reference position in order to determine the horizontal position of the pendulum with respect to the top of the attractor.

The reference photo and in-position photo for this measurement are shown in Figure 3.6. This entire set of measurements was done with a substantial horizontal position offset; the pendulum had been displaced horizontally in runs prior to the beginning of the science data in an effort to understand a systematic ISL-violating signal that would later be traced to the bypass capacitors (Section 8.3). As can be seen in Section 7.8.2, the impressive null-ness of this attractor geometry allows us to get away with such a gross misalignment without important consequence.

3.3.2 Vertical position

The vertical position of the pendulum was determined using a height-gauge referenced from the optical bench. Before installing the foil, the pendulum was centered vertically on the attractor, to a precision conservatively better than 250 μm. The pendulum's suspension system was fixed by a flexure-based clamp that went untouched throughout the entire science campaign.

3.3.3 Pendulum-Foil separation

We assess the pendulum-foil separation in two different ways. The most quantitative is the 'bounce test', but it is potentially confounded by both pendulum tip (misalignment with the foil about a horizontal axis parallel to the foil face) and by dust between the pendulum and foil. The most reliable, but less sensitive, approach uses direct imaging of the pendulum-foil distance. These two methods were found to be in reasonable agreement with capacitance measurements, but do not require any detailed modeling nor consideration of fringing fields.

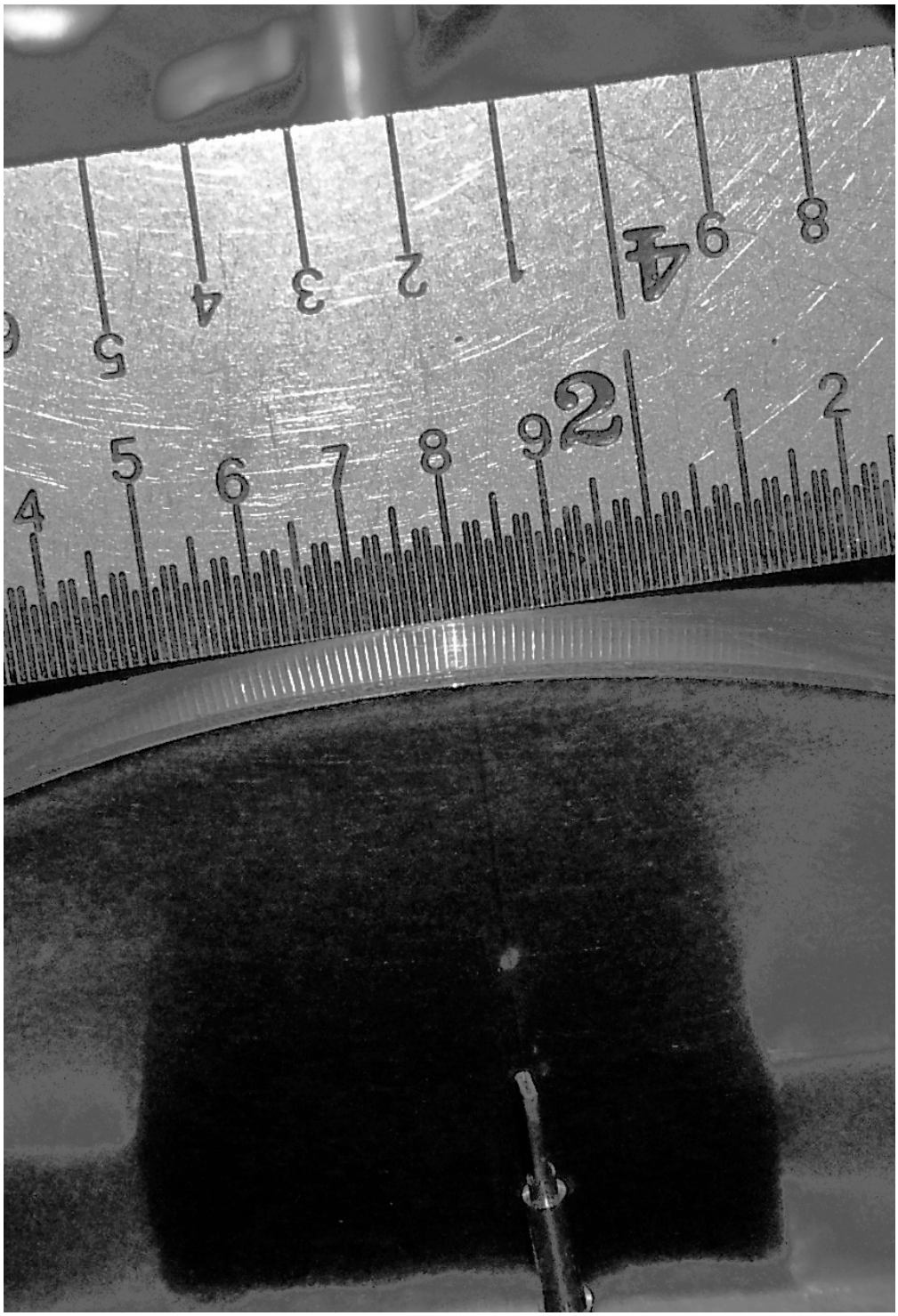

Figure 3.6: Horizontal fiber position. The rough lateral scratches at the top of the attractor mark its highest point. The fiber is visible as a faint out-of-focus, nearly vertical, line on both the attractor and the machinist's scale.

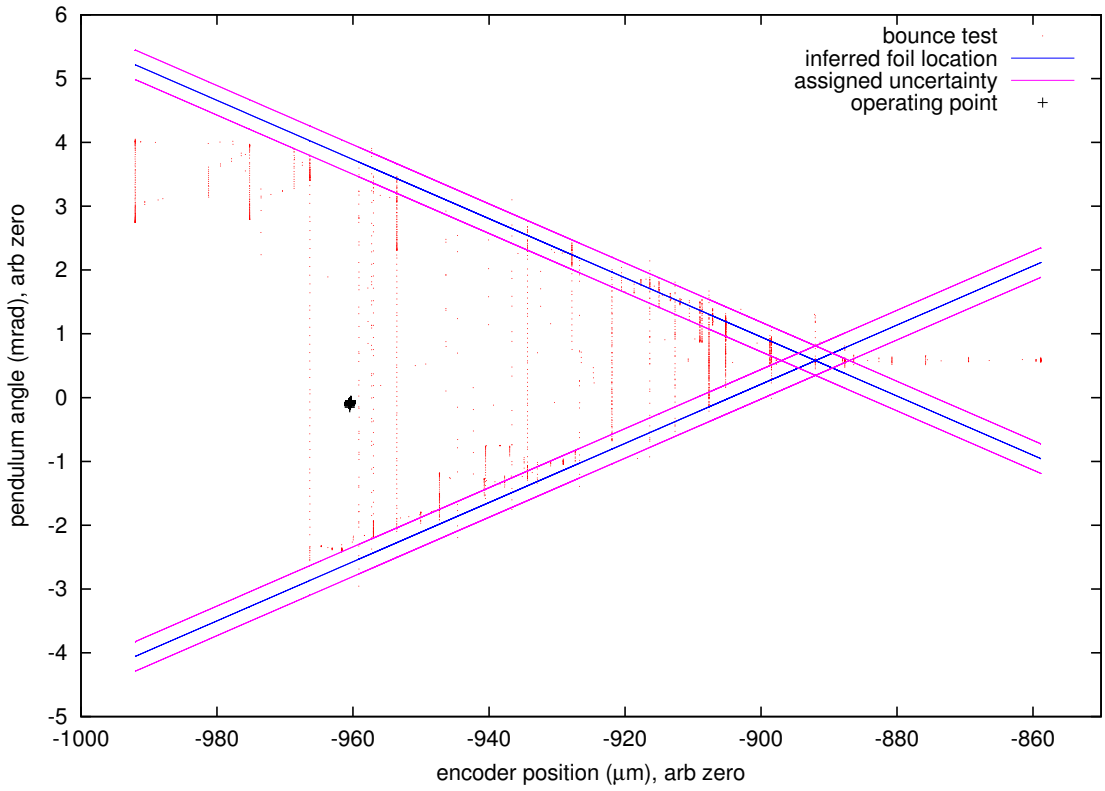

Figure 3.7: Pendulum bounce-test. The pendulum swings freely in angle (vertical axis) as the pendulum support is translated slowly toward and away from the foil (horizontal axis). The bands reflect the assigned uncertainty; the slope of the band is determined only by the pendulum width and the distance translated.

3.3.3.1 Bounce test

The pendulum is allowed to swing freely as the fiber support is moved toward and away from the foil/attractor. The autocollimator trace, plotted as a function of support-position shows a clear pattern from which the point at which the pendulum touches the foil may be extracted. The results of the pre-science bounce test are shown in Figure 3.7.

3.3.3.2 Photos

The PlateWash apparatus permits visual access to the pendulum at all times through a glass vacuum viewport. Visual access provides a reliable and robust cross-check for our more-quantitative metrology techniques. With the addition a mirror, we are able to see both (left and right) sides of the pendulum and photograph them. Using the known thickness of the pendulum, we can compare photographs of the pendulum taken during data-taking with photographs taken when the pendulum is pressed lightly against a flat surface (attractor or foil) to determine the data-taking pendulum's distance from, and angle with respect to, the foil. For the photos used in this determination, see Figure 3.8.

For each photograph, the thickness of the pendulum, the pendulum reflection, and the gap between them were measured at repeatable points near the top, middle, and bottom of the pendulum. The widths of the gaps, divided by two to correct for reflection, were calibrated by the known pendulum thicknesses. The mean difference in gap size between the pre/post-science photographs and the inclined/on-foil photographs was computed to generate an estimate of the pendulum's displacement from the foil of (66 ± 17) μm.

3.4 Pendulum angle

The alignment angle of the pendulum with respect to the attractor is important to the sensitivity of the instrument, in both vertical tip and in twist. Here, we define "tip" to be the angle made between the pendulum and the foil about a horizontal axis parallel to the foil. We define "twist" to be the angle made between the pendulum and the foil about a vertical axis parallel to the foil.

3.4.1 Twist angle

3.4.1.1 Bounce test

We can use the bounce-test data to measure the effective angular difference, about a vertical axis tangent to the foil, between the pendulum's planar orientation and the foil. Comparing the angle at which the experimental data were taken with the angle where the inferred lines cross in Figure 3.7, we find that the angle between pendulum and foil is (0.8 ± 0.2) mrad.

3.4.1.2 Photos

Just as with the pendulum distance (Section 3.3.3), we can use the pendulum photos to extract the effective angle at which the pendulum sits. The pendulum-thickness-calibrated difference between pre/post-science and inclined/on-foil photographs for each view of the pendulum was computed. The ratio of the two measured pendulum-foil separations to the pendulum width yields a pendulum twist angle of (829 ± 814) μrad.

3.4.2 Tip angle

We expect the pendulum tip angle to be a stable quantity, determined primarily by the initial alignment. Over the duration of science data acquisition, the observed change in pendulum tip was less than 100 μrad (Figure 3.9, light blue). We measured the absolute tip angle in two ways: the "plateau test" and directly using photographs.

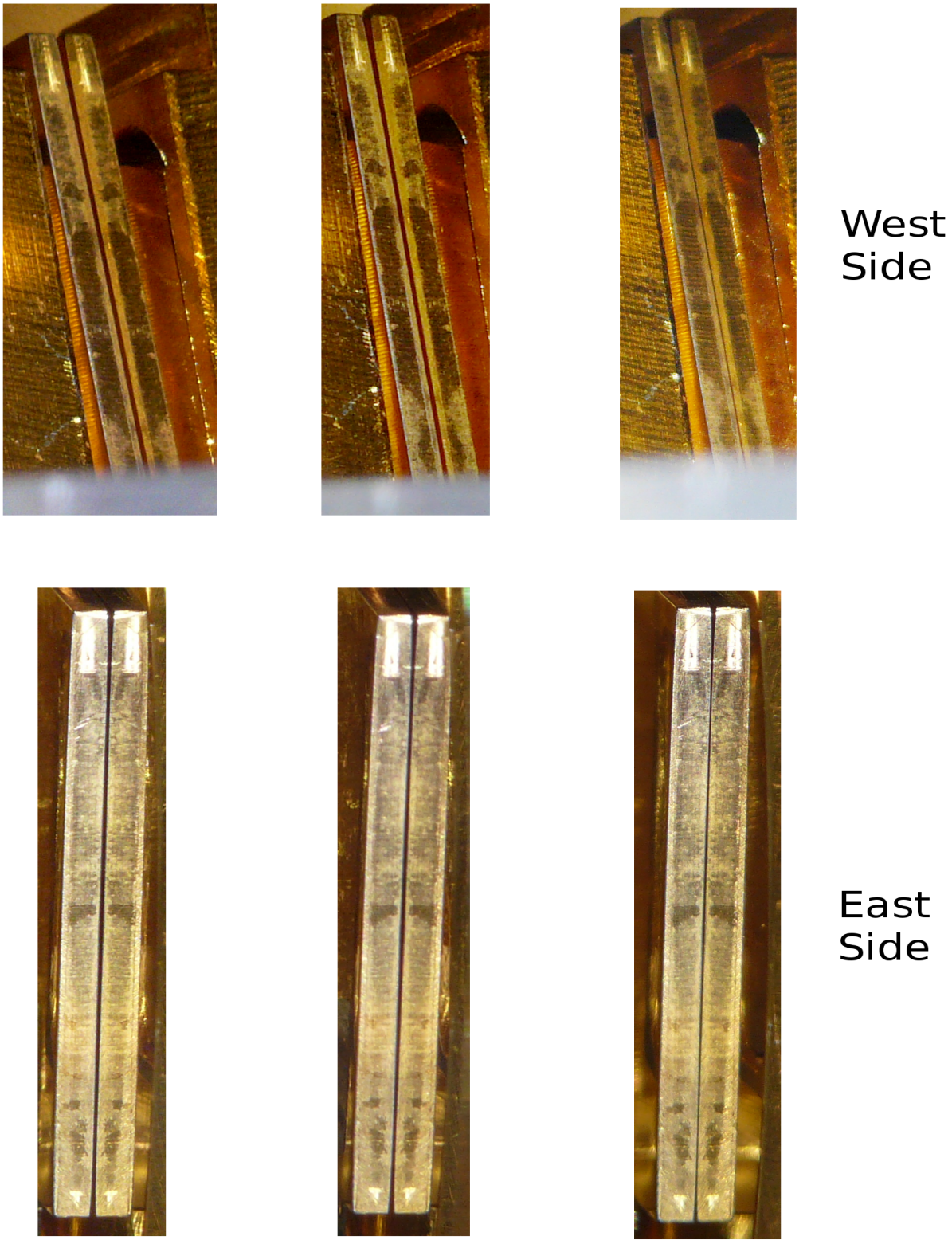

Figure 3.8: Pendulum position photos

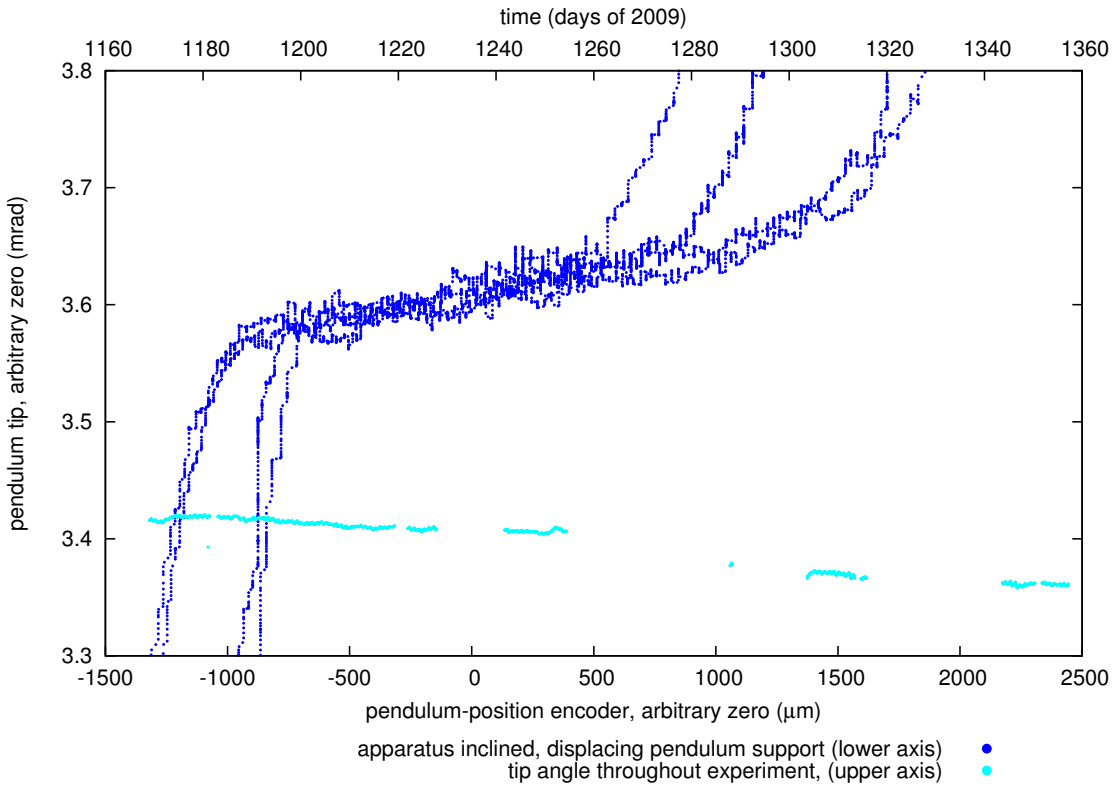

Figure 3.9: The "Plateau Test" to determine the relative angle, about a horizontal axis parallel to the foil, between pendulum and foil. The dispersion at either end of the plateau comes from hysteresis; the pendulum peels away from the foil very slowly when leaving the plateau.

3.4.2.1 Plateau Test

The "Plateau Test" is a quantitative method for determining the angle where the pendulum's surface closest to the foil is parallel to the foil, but is susceptible to systematic error due to pendulum curvature and dust. The entire apparatus is tipped to a significant angle (at least tens of milliradians) about the horizontal axis perpendicular to the attractor symmetry axis. Then, the pendulum displaced toward and away from the foil while the autocollimator reading (both up-down and left-right) is recorded. The angle at which the pendulum sits parallel to the foil appears as a plateau in the plot of measured angle as a function of pendulum-support displacement. As seen in Figure 3.9, we claim a difference of (200 ± 71) μrad between the science position and the maximally-parallel position.

3.4.2.2 Photos

As with the pendulum-distance and twist angle, we can use the pendulum photos to extract a tip-angle measurement. For each photograph of the pendulum, we used the pendulum-thickness-calibrated measurements of the apparent gap (again correcting for the factor of two from reflection) to determine the relative slope

of the pendulum and its reflection in the foil. The average difference between the slopes (in units of gap pixels per pixel along the pendulum edge) of the science photographs and the inclined/on-foil photographs was taken as the pendulum tip angle of (-80 ± 758) μrad.

3.5 Foil Thickness

The thickness of the foil was determined by making many micrometer measurements of foil from an adjacent portion of the same roll. Touching the foil, particularly with a hard/sharp-cornered object like a micrometer, can easily damage the foil by denting it. Even a piece of dust, compressed between a micrometer jaw and the foil, can be sufficient to dimple the foil and increase its effective thickness. For this reason, we have not directly measured the actual foil's thickness. If the foil thickness measurement were to become of increased concern, we could destructively test it.

As measured the foil is (13.9 ± 0.5) μm thick.

3.6 Assignment of attractor/foil separation

The distance between the pendulum and the attractor is determined to be the sum of the pendulum-foil distance, the foil thickness, the attractor-foil distance, and the pendulum-angle correction. Both the statistical and systematic errors are summed in quadrature, as the systematic uncertainties quoted here are truly uncertain. We have chosen the central values for those measurements that have meaningful systematic uncertainty to reflect any bias we expect from these measurements.

The flatness measurements quoted in the table are taken using conservative estimates drawn from Figures 2.10, 2.11, and 2.13.

Quantity	Measurement Method	Measured Value	Contribution (μm)	Statistical Error (μm)	Systematic Error (μm)
Pendulum					
Twist	Bounce Test	(0.8 ± 0.2) mrad	17.4	0.0	4.3
	Photos	(829 ± 814) μrad			
Tip	Plateau Test	(200 ± 71) μrad	3.2	0.2	1.1
	Photos	(-80 ± 758) μrad			
Flatness	Laser Scan	7.5 μm	7.5	1.0	7.5
Separation	Bounce Test	(68 ± 5) μm	68.4	0.0	5.0
	Photos	(66 ± 17) μm			
Attractor					
Tip	Interferometer	(20 ± 5) μm	20.0	2.0	5.0
Distance	Interferometer	20 μm	20.0	0.0	0.0
Flatness	Laser Scan	4.5 μm	4.5	1.0	4.5
Foil					
Thickness	Micrometer	(13.9 ± 0.5) μm	13.9	0.5	1.0
Flatness	Laser Scan	2 μm	4.0	2.0	4.0
Total:			158.9	3.2	12.8

Table 3.1: Uncertainty table for the minimum effective attractor-pendulum separation

Chapter 4

Strategic choices and operation

This chapter describes important strategies and choices we made in the operation of the instrument: selecting the signal frequency, selecting the pendulum-angle signal as our science signal, random attractor locations, and the overall distribution of attractor locations.

4.1 Frequency-space

Each PlateWash measurement operates in a frequency regime spanning 100 μHz to 1 Hz. This section describes the way in which important experimental frequencies were selected. The most-important frequency, the frequency with which the attractor changes position, is pushed by several concerns to the highest-feasible frequency. The limit of feasibility is set by both the limited speed with which our attractor drive can operate and by the speed with which the pendulum feedback control can react without adding extra noise.

4.1.1 Spectral Character of Noise

Our torsion balance is an excellent harmonic oscillator. For quality torsion fibers operated at amplitudes much smaller than a radian, the anharmonic terms are negligible[1]. The torque-to-angle transfer function of an internally-damped torsion oscillator with resonant angular frequency ω_0, torsion spring constant κ, moment of inertia I, and quality factor Q is [105]

$$T(\omega) = \left|\frac{\theta(\omega)}{\tau(\omega)}\right| = \frac{1/\kappa}{\sqrt{(1-\omega^2/\omega_0^2)^2 + 1/Q^2}}; \Rightarrow \text{If } \omega \gg \omega_0, \left|\frac{\theta(\omega)}{\tau(\omega)}\right| \rightarrow \frac{1}{I\omega^2}$$

[1] In a precise measurement in beryllium-copper at 77K on a 20 μm diameter fiber with $\kappa \sim 10^{-7}$ N-m/rad, with a parametrization of $\kappa(\theta) = \kappa_1 + \kappa_2\theta + \kappa_3\theta^2$, Bantel and Newman found $\kappa_2/\kappa < 4 \times 10^{-7}$ rad^{-1} and $\kappa_3/\kappa < 5 \times 10^{-8}$ rad^{-2} [104]. For this experiment, with $\kappa \sim 10^{-8}$ N-m/rad, maximum angles of < 1 mrad, and a torque goal of 10^{-16} N-m, a κ_2/κ ratio as large as 10^{-2} rad^{-1} would yield a sufficiently linear response.

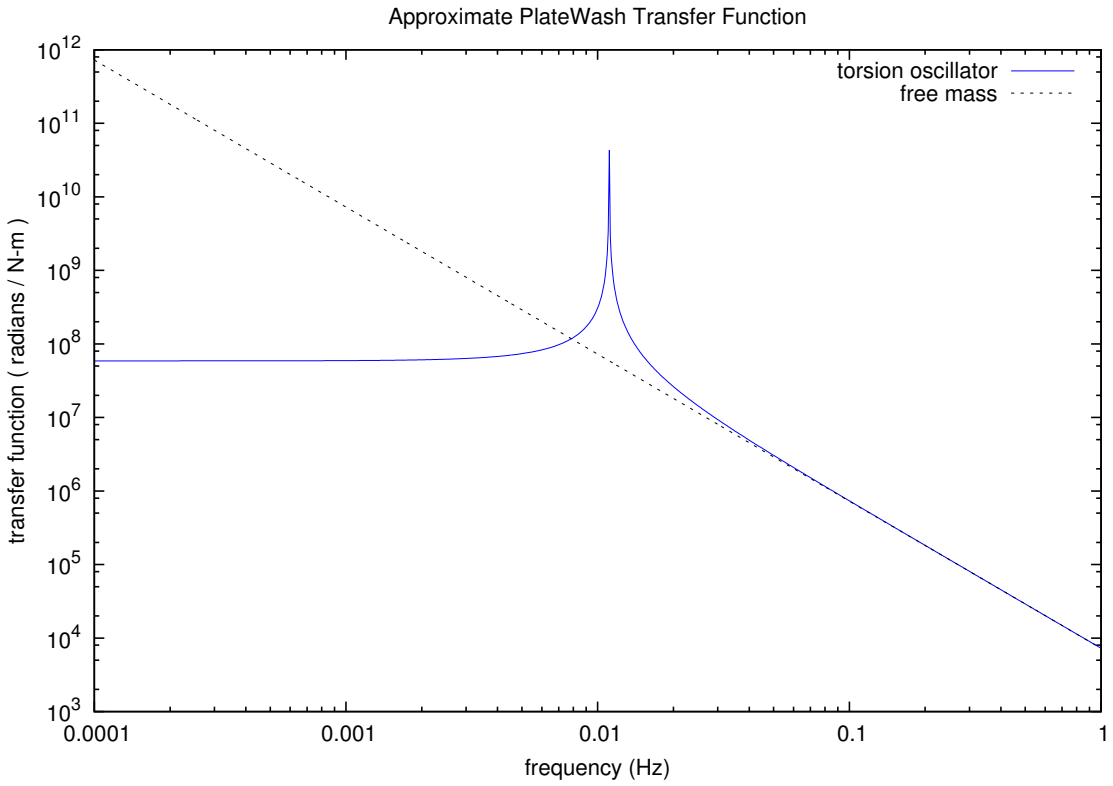

Figure 4.1: Approximate torque-to-angle transfer function

For $\omega \gg \omega_0$, the pendulum response is independent of the torsion spring. This is the torque equivalent of $F = ma = m\ddot{x} \rightarrow \tilde{\tau} = I(-\omega^2)\tilde{\theta}$; at these frequencies, the pendulum acts like a free mass. When $\omega \ll \omega_0$, the pendulum response is frequency independent.

4.1.1.1 Thermal Noise

As described in Section 2.4.1.1, the torque noise power of a thermally-limited torsion balance is

$$\tau^2(f) = 4k_B T \frac{\kappa}{2\pi f Q}.$$

This drives any measurement toward higher frequencies for decreased noise. In power spectral amplitude, the improvement only comes as $1/\sqrt{f}$, implying that the benefit of increased frequency is limited.

Convolving this noise with the torque-to-angle transfer function (Figure 4.1), we recover the angle noise spectrum of a thermally-limited torsion oscillator.

4.1.1.2 Readout Noise

No physical measurement has arbitrary precision; we must consider the effects of finite-precision angle readout on our ability to measure torque. For simplicity, let our angle noise at all frequencies be $\delta\theta$ (a practical approximation, see Figure 2.14). Our minimum resolvable torque signal at any given frequency is

$$\delta\tau(\omega) = \frac{\delta\theta}{T(\omega)} = \kappa\sqrt{\left(1-\frac{\omega^2}{\omega_0^2}\right)^2 + \frac{1}{Q^2}}\,\delta\theta \approx \kappa\left|1-\frac{\omega^2}{\omega_0^2}\right|\delta\theta$$

At frequencies above the pendulum resonance, torque noise rises with frequency.

4.1.1.3 Proximity Noise

As described in Section 2.4.1, the noise in the apparatus increases as the pendulum approaches the foil. In our frequency range of interest, it appears as relatively frequency-independent noise. The level of this noise is time-dependent, correlated with wind, human activity, and both local and worldwide seismic activity.

4.1.1.4 Experimentally determined noise curves

As shown in Figure 4.2, the minimum noise is near ~ 25 mHz, equivalent to a 40 s period. As will be seen below, the ~ 20 s response times of both the pendulum feedback loop and the attractor drive prohibit useful operation at such a high frequency. Near the attractor frequency of 3.906 mHz, the torque noise is generally at or below 20 fN-m/$\sqrt{\text{Hz}}$ during quiet periods.

4.1.2 Statistics

The higher the attractor frequency f, the more individual measurements we can make. In principle, given frequency-independent noise, the choice of signal frequency is arbitrary. If the noise power-spectral-density $S(\omega) = S_0$, then the resulting uncertainty of N measurements each of duration $t_m = 1/f$ is $\sigma_{all} = \sqrt{\frac{S_0}{Nt_m}}$. If the measurements are performed in a fixed time $T = Nt$, then $\sigma_{all} = \sqrt{\frac{S_0}{T}}$, independent of t_m and N.

Practically, however, noise is "non-Gaussian", i.e. $S(\omega)$ is both time and frequency-dependent. Sudden/prominent events including earthquakes, human activity, attractor/foil interactions, and computer crashes happen. They are often short in duration, but overwhelming in amplitude. Measurements taken at these times are unusable. In this experiment, such transients happened at most a few times in a day. It will be shown in Section 5.4.2 that for any large transient, measurements for at least $30t_m = 7680$ s $= 2.1$ hr must be discarded. If t_m were smaller, the experiment's fractional dead-time would be smaller.

A second benefit to small t_m is the increased number of independent measurements. This experiment relies

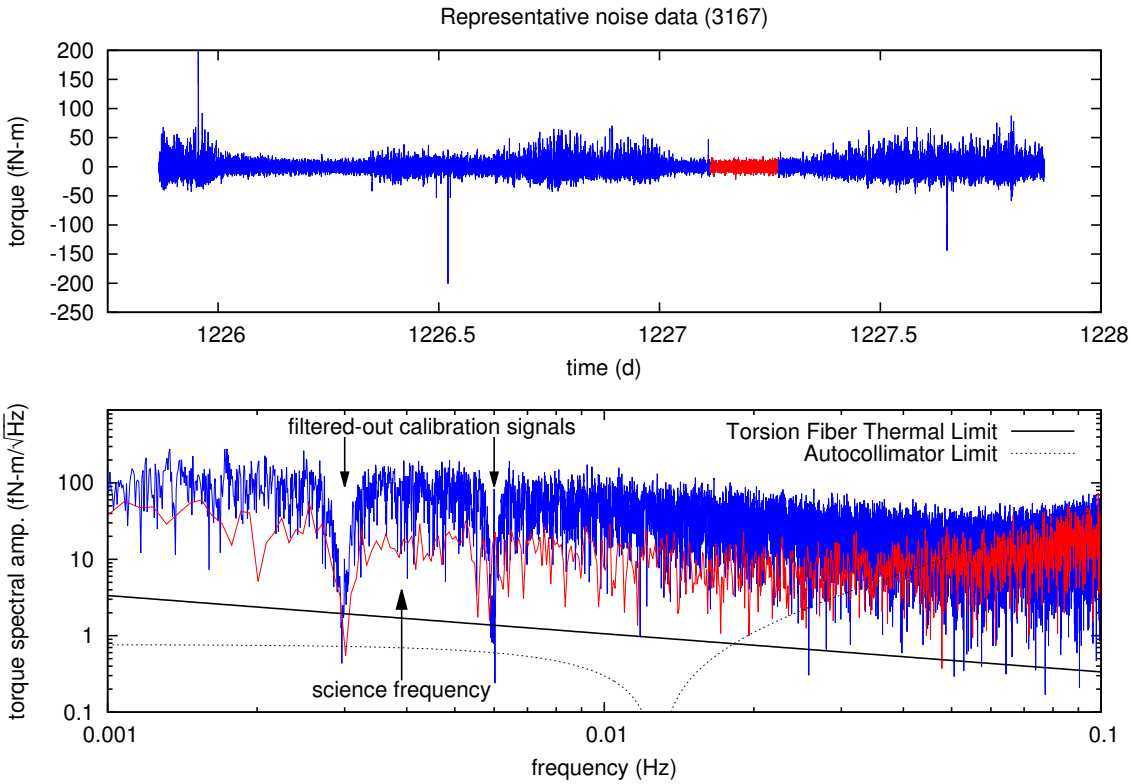

Figure 4.2: Experimentally-measured torque noise, from systematic test 3167. Attractor held stationary. Data in blue cover the entire time-series shown, while data in red are taken from the quiet period early in the morning of day 1227. The power spectrum prominently shows the effects of the calibration-signal notch filters discussed in Section 5.4.2. Our lowest noise reliably occurs for several hours early each morning.

Harmonic (m)	1	3	5	7	9	11	13	15	17
Fourier coefficient	1.27	0.42	0.25	0.18	0.14	0.12	0.10	0.08	0.07
Combined signal-to-noise ratio (1 to m)	1.27	1.34	1.36	1.39	1.39	1.40	1.40	1.40	1.40

Table 4.1: Square wave Fourier coefficients and combined signal to noise ratio up to $n = 17$

upon Monte-Carlo-like measurements (random attractor locations) to constrain both systematic uncertainties and possible violations of the inverse-square law. Frequent individual measurements allow faster exploration of parameter space (exploring a 6×6 grid of measurements requires a day with the current choice of t_m). Measurements that are twice as frequent allow twice as many questions to be answered, albeit with precision degraded by a factor of $\sqrt{2}$. When noise is frequency-independent, there is no penalty for this approach.

Faster switching is better, much more than $\sqrt{\text{better}}$, as increased statistics allow greater immunity to transient sources of noise and improved study of experimental parameter space.

4.1.3 Harmonics

In order to avoid any attractor-velocity-dependent systematic uncertainty, the attractor must move in a square wave, with measurements taken only when the attractor is at rest. The Fourier coefficients of a square wave are $A_n = 4/\pi n$ for all odd n ($A_n = 0$ for even n) [76].

For measurements with equal uncertainty σ_h of each harmonic, the uncertainty in a combined measurement of the first m harmonics is $\sigma_{combined} = \sigma_h / \sqrt{\sum_{j=1}^{m} A_m^2}$. In Table 4.1, the signal to noise ratio is computed for $\sigma_h = 1$. The first harmonic is the only one that matters.

It is important that amplitude of the first harmonic of a unity-amplitude square wave is greater than one. Without resolving the square shape of the signal through its harmonics, we reap the increased sensitivity one expects from an attractor path that spends far more time at maximum amplitude.

4.1.4 Dead Time

The first 65 s after a position change and 35 s preceding a position change are excluded to conservatively reject any position-change related effects (see Section 5.4.4 for detail). As a consequence, the highest available switch frequency is <5.0 mHz. At that frequency, the accepted measurements would have zero duty cycle. The duty cycle available for measurement is $D = (1 - 2Lf)$, where L is the total time excluded from an individual step and f is the attractor frequency. The SNR attenuation is roughly $\sqrt{D}$, plotted in Figure 4.3.

The combination of the large conservative dead-times with the 10 s low-pass torque filter should render the measurement insensitive to small variations in the specific amounts of time to be dropped. The blind analysis makes it impossible for us to fine-tune any choice of dead-time for a desired result.

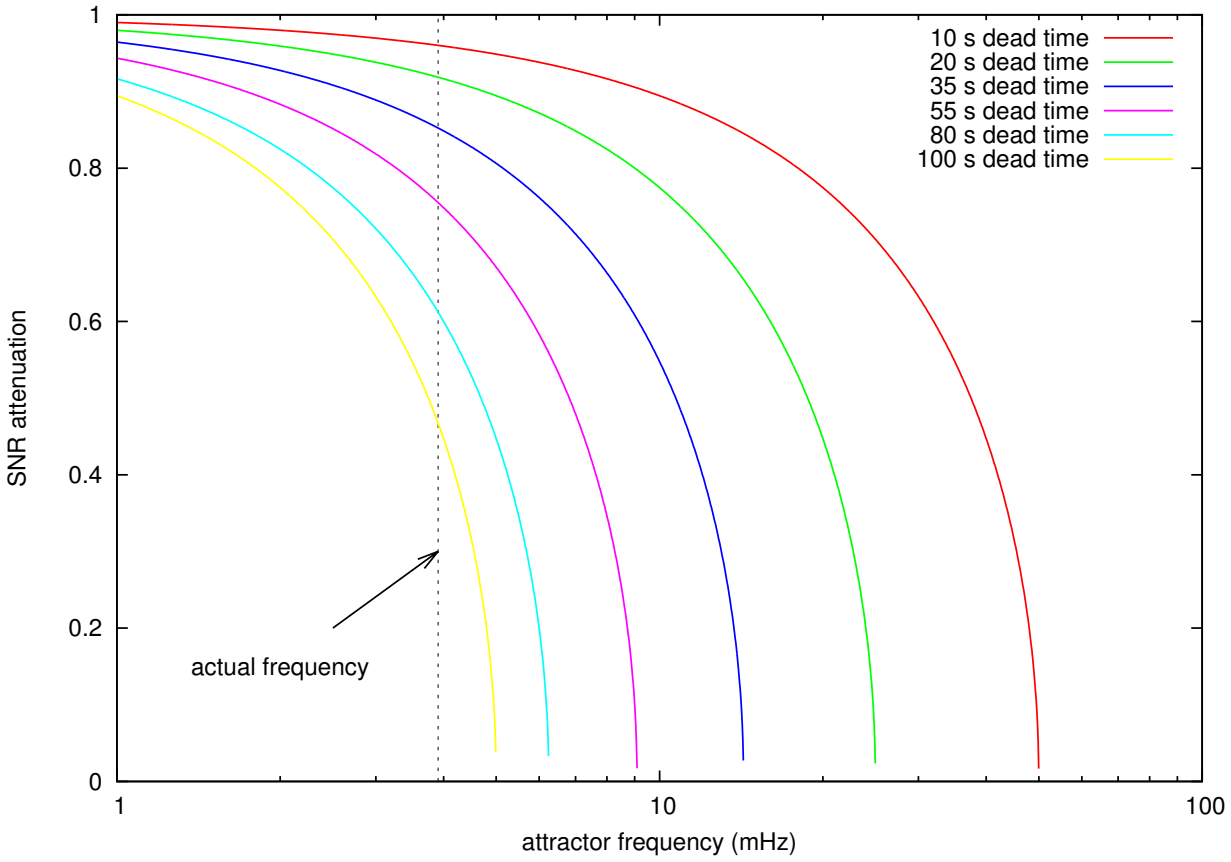

Figure 4.3: Signal-to-noise attenuation as a function of attractor frequency (horizontal axis) and of total dead time duration (individual curves). As expected, once the per-step dead time is smaller than 75% of the step duration, half of the SNR is available for use.

4.1.5 Choosing a frequency

As the first harmonic is the only signal of importance (4.1.3), we should place the first harmonic at the location of lowest noise (4.1.1.4). Furthermore, we are driven by statistics (4.1.2) to operate at the highest practical frequency. The required use of dead times to eliminate systematic uncertainties places a soft upper bound on usable frequencies (4.1.4).

The operating frequency was chosen well in advance of the science run reported here, before the instrument was fully commissioned. We selected a frequency that we expected would give low noise and was sufficiently slow to allow transients associated with attractor motion to dissipate. We selected 3.90625 mHz, or 1/256 s. 256 s is an integer (320) multiple of the 800 ms sample time, making each attractor period an integer number of samples. This choice improves the performance of discrete Fourier analysis and minimizes aliasing between signal and sample frequencies.

4.1.6 Switches per cut

The number of attractor position changes per 'cut', a group of identical measurements, was chosen to be large-enough to allow a sufficient understanding of the switch-to-switch variation in the torque at each point. We settled on 10 switches per cut in an ad-hoc fashion. It provides a reasonable understanding of the switch-to-switch variation, and is still short-enough to address the speed concerns described in Section 4.1.2.

4.2 Choosing a science signal

As the pendulum is locked in feedback, there are multiple junctures in the feedback loop where a science signal could be extracted. The pendulum twist angle, the feedback signal itself, and the proportional, integral, and differential terms of the feedback loop are written to disk. The pendulum twist angle and the proportional term of the loop are equivalent. In practice, the twist signal and proportional term were anti-aliased differently before being written to disk. The proportional term had a 5 s filter applied (a factor of 3 below the Nyquist frequency), making it the signal of choice for a reliable angle measurement. The loop was operated with zero integral-term to prevent improper wind-up in the challenging near-foil environment, so the integral channel contains no signal. The derivative-term response is frequency-dependent, growing roughly linearly with ω, making it unsuitable for low-frequency signal extraction.

So, there are two reasonable science channels: the feedback signal applied to the pendulum, and the proportional term. The feedback signal has units of torque, the proportional term has units of angle. Even though the pendulum is held fixed, the torque-to-angle transfer function matters. At frequencies below

the resonance of the pendulum, the two channels are proportional. At frequencies above resonance, the frequency-independent angle noise of the autocollimator appears in the applied feedback signal, as torque noise that rises like frequency squared. As shown in Section 4.1.3 and Table 4.1, the higher harmonics of the attractor motion contain no interesting signal, so it is in our interest to low-pass them away.

By choosing the angle/proportional-term signal instead of the feedback signal as the science signal, we use the inertia of the pendulum as an analog low-pass filter. While, in principle, both signals contain the same information, the science signal is more-readily extracted from the angle signal. Furthermore, as the feedback signal is the sum of the proportional and derived differential terms, if the science were extracted from the feedback signal there would be greater opportunity for software or digitization errors to introduce excess noise. Finally, we note that selecting among signal sources is straightforward with a continuous torque calibration signal operating at a nearby frequency.

4.3 "Random" attractor motion

One of this experiment's important departures from past parallel-plate-style gravitational experiments [62, 18] is the use of random attractor locations. This approach trades away some signal-to-noise in return for improved characterization of systematic effects. Instead of focusing exclusively upon a small set of start and end locations for the attractor, we considered a large distribution of them. This method has the potential to expose important systematic curiosities, like the one described in Section 8.2, and has the added advantage of allowing a more-complete measurement of the distance-dependence of any measured force. The drawbacks of this approach, primarily additional complexity, are discussed in Sections 13.1.1.1 and 13.2.2.1. A representative sample of attractor data is shown in Figure 4.4.

The random locations were generated by random number generator and written into configuration files for the attractor computer to execute. This extra step allows systematic tests that repeat the attractor positions used in science runs. The distribution of random attractor locations used was the simplest, and perhaps least efficient, uniform distributions of start and end positions. This spends perhaps more time than is most-efficient exploring shorter attractor-position differences.

4.4 Attractor Positions

The attractor position was varied through both "random" and square-wave patterns. More than half of the science data were taken using random patterns. Figure 4.5 shows all attractor positions used in this analysis.

The attractor positions were chosen by estimating the expected sensitivity of the aggregate science mea-

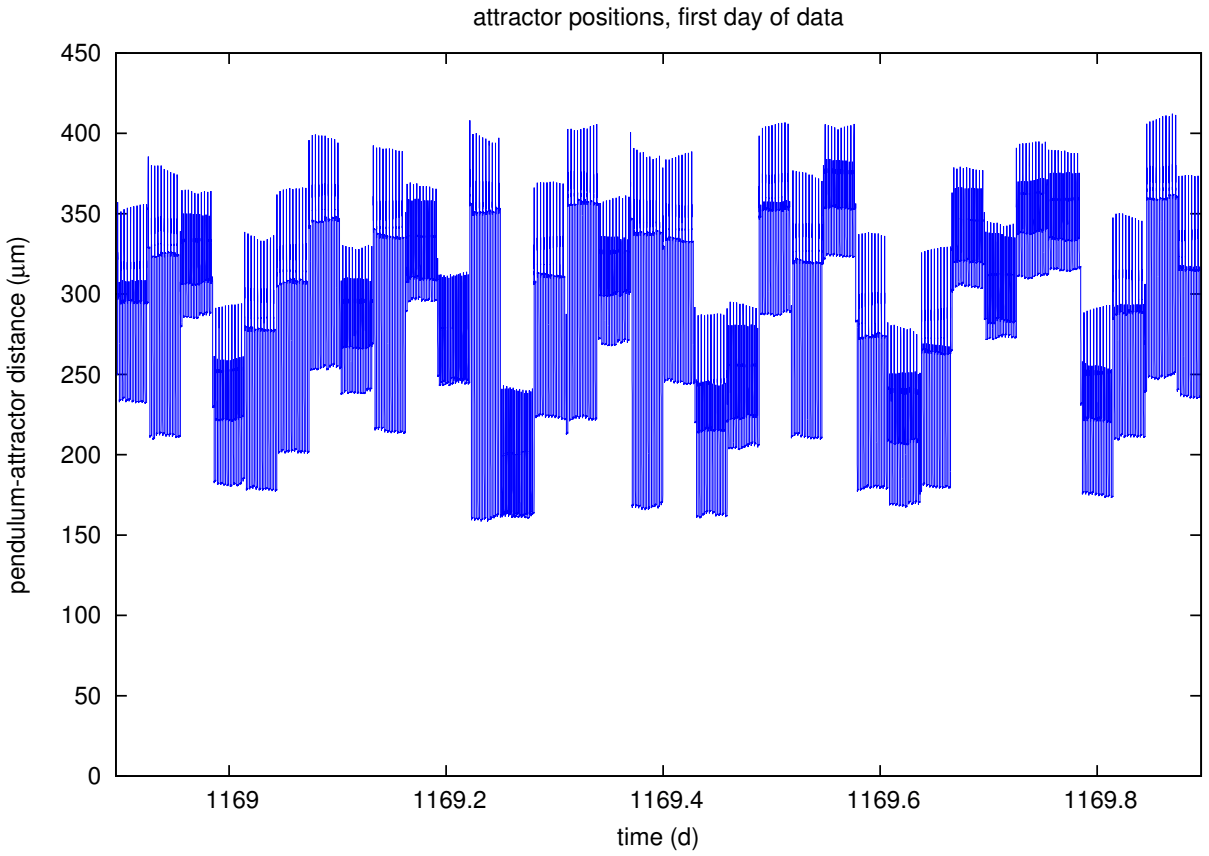

Figure 4.4: First day of attractor positions. Random attractor positions sample the available parameter space in an unbiased way. The short-duration upward-going spikes are an artifact of the pneumatic system discussed in Section 2.3.6.1.

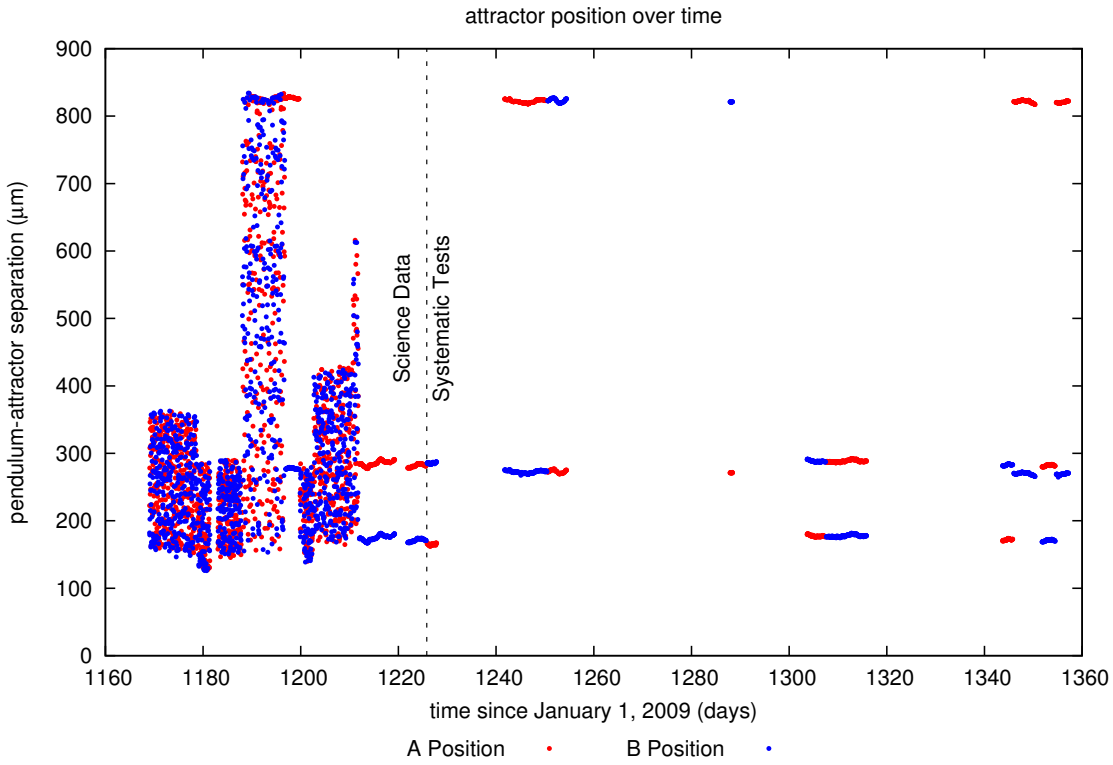

Figure 4.5: Attractor positions used in this analysis.

surement. As the statistical noise fluctuates significantly on a day-to-day basis, it was impossible to predict precisely the performance of the apparatus. A few reliable measurements during conditions of comparatively low noise can carry more statistical power than a day's worth of measurements in noisy conditions. As we used our blinded sensitivity for guidance, we are confident that we have not biased our experimental outcome by tuning into any desired result. In practice, this guidance led us to switch from predominantly "random" runs to a focused square-wave approach late in the science run in an effort to improve/decrease the shortest range at which we could reach gravitational strength. All of the systematic tests were done with "short" and "long" square-wave measurements to allow more-straightforward intercomparison.

4.5 Campaign description

The measurements began with science data, comprising runs 3146/7 to 3166. Systematics tests followed; bellows valve-off in 3167–3172; resistor tests 3174–3183; magnetic tests from 3184–3205; heating/thermistor tests from 3206–3222; magnetic signal injection 3223–3225; attractor electrical bias 3234. Some runs preceding 3234 had the interferometer laser turned off; more archaeology awaits to determine if any torque sensitivity to interferometer-light-heating can be extracted.

In the future, interleaving systematics tests with science runs would be more satisfactory. With the exception of thermal tests, which have long time-constants, it's possible to do a science measurement, magnetic tests, interferometer on/off tests, and electrical bias tests programatically with each attractor setting, merging them into a single super-block of measurements that comprise a self-contained ISL test for one attractor-position pair. The ability to make incremental/aggregated ISL tests may be worth the added programming complexity.

Chapter 5

Data Analysis

5.1 Flow

This section, Section 5.1, is an executive summary for the details in the remainder of the chapter, with the flowchart in Figure 5.1 as a companion. Figure 5.1 diagrams the flow of data through the data analysis code. This section provides a chronological narrative. The data analysis that precedes new-physics fits is functionally run twice, once using only the pendulum and attractor data streams, and once including the interferometer. As detailed in Section 2.3.9.1, the interferometer data are only useful over parts of the run, requiring model-building in order to make a torque correction during times that the interferometer was unusable.

5.1.1 Source data

The data used in the analysis come from two sources, those that we determined externally and hand-entered and those sourced directly from the data acquisition systems.

5.1.1.1 External inputs

Externally-applied systematic injections and environmental variables (temperature, magnetic field, etc.), were determined by hand (Section 9.2) or in separate analyses (Section 9.3) on a run-by-run basis.

In addition, all experimental parameters (alignment, dimensions, densities, analysis parameters, etc.) are stored in global configuration files that are read by each analysis program that requires them.

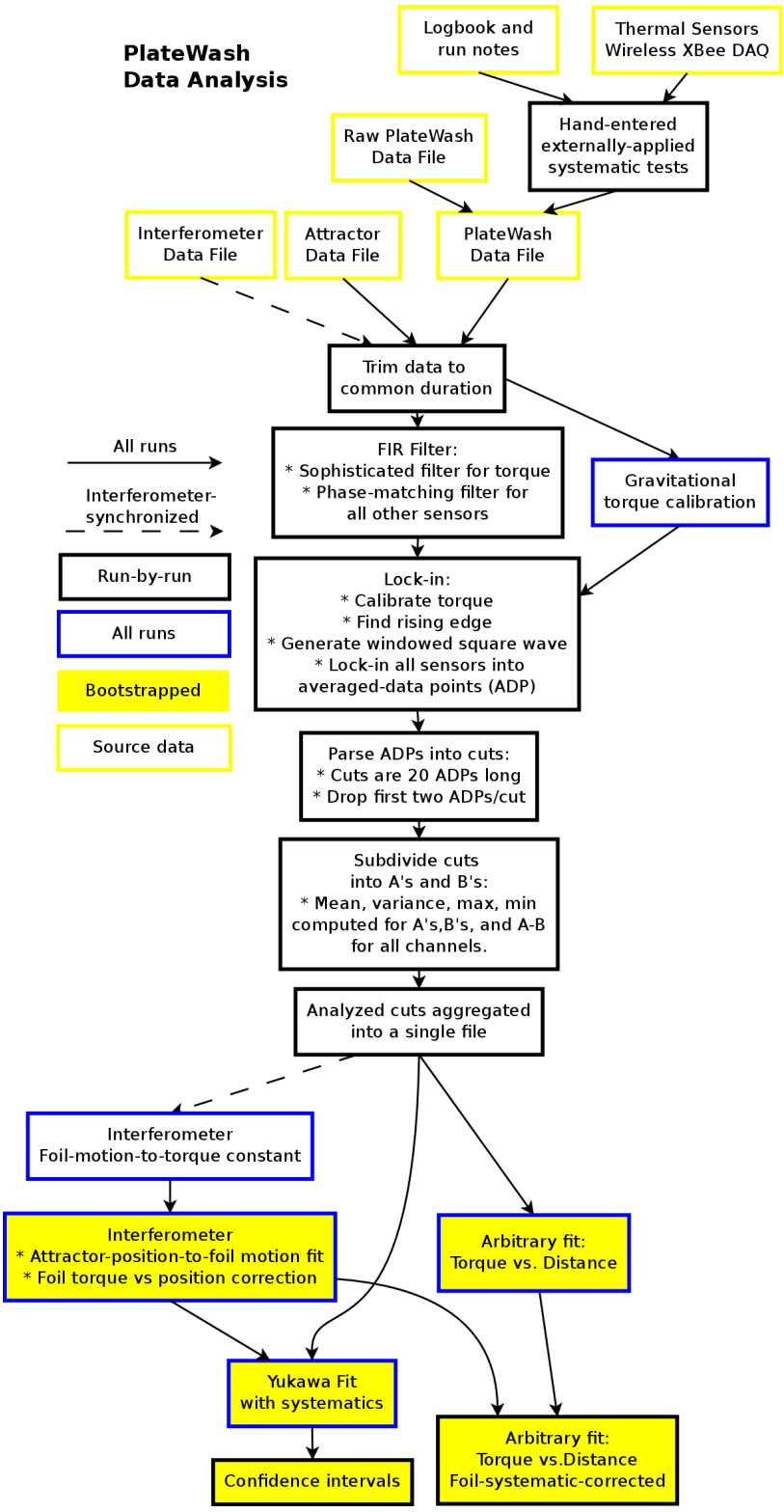

Figure 5.1: Flowchart diagramming the flow of data analysis. After configuration of experimental and analysis parameters, timing cuts, and the hand-input of some systematic test parameters, it is automated and replicable.

5.1.1.2 Data files

Each data-acquisition computer generates its own data files. When the pendulum computer's data are read, externally-measured data are joined to the raw output from the pendulum DAQ. The timing data are then referenced to a common datum (00:00:00 January 1, 2009 PST).

5.1.2 Run-by-run analysis

This section describes the analysis done for individual "run" of the experiment. A "run" refers to a length of coincident operation of two or more DAQ systems specifically delineated by a run-configuration file as having something interesting to say. The main routine of the run-by-run analysis looks through the directory of run configuration files and analyzes the data loaded by each one.

5.1.2.1 Trimming

As described in Section 5.3 and Section 5.4.1, the data from each computer do not begin simultaneously and may have offsets between their clocks. We correct the differences in the clock rates and conservatively absorb the differences in clock zeroes into the assigned experimental dead times.

5.1.2.2 Filtering

As described in Section 5.4.2, the torque data are passed through FIR filters, primarily to notch out the gravitational calibration signal and to bandpass away drift in the pendulum. All other channels to be analyzed alongside the torque data are passed through an inconsequential (low-pass at the Nyquist frequency), but equal duration, filter to guarantee phase-matching with the torque signal.

5.1.2.3 Calibration

In a parallel analysis with the filtering step, the torque data are calibrated using the gravitational calibration injection, as described in 5.2. The calibration takes the form of a multiplicative constant, applied to the torque data in the "Lock-in" step.

5.1.2.4 Lock-in

The attractor moves in an $ABABABABABABABABABAB$ fashion. Before measuring the differences between state A and state B, it is necessary to phase-match the analysis to the phase of the attractor motion, as described in Section 5.4.3. The code finds rising edges in the attractor data and uses them to find the initial steps in the data. Using these steps, the code then projects, open-loop, the temporal location of

all the steps in the dataset. The value of each signal channel is then measured during each step, using the dead-times projected in Section 5.4.4.

5.1.2.5 Parsing into "cuts"

The attractor's motion is broken into "cuts" of ten attractor periods, after which the attractor's A and B positions may change. The code parses the data into these cuts, as referenced from the starting location (up to an attractor half-period, runs must begin in-phase with a cut). The first A and B states are dropped from the cut before the aggregate properties of the cut are determined. As seen in Figure 2.35, the positions of the first two states are often inconsistent with the rest of the cut and retain position information from the prior cut. As detailed in Section 5.4.6, all channels, $\bar{A}$, $\bar{B}$, σ_A, σ_B, $\max(A)$, $\min(A)$, $\max(B)$, $\min(B)$, $\bar{\Delta}_{AB}$, and $\sigma_{\Delta_{AB}}$ are measured. Out-of-phase quadrature measures of these quantities are also recorded, but are subjected to no further analysis.

Only these cut-wise data are used for analysis beyond this point.

5.1.2.6 Aggregation

All the cuts from the many runs are aggregated into a single file (one for the pendulum/attractor two-DAQ analysis, and one for the three-DAQ interferometer analysis). These 2929 cuts (two-DAQ, 1313 three-DAQ) comprise the measurement.

5.1.3 Interpretation

Now that the data are measured, what do they say? As the foil systematic is important, we first ascertain the size of any necessary correction due to foil motion. We then fit for the presence of any Yukawa signal. In addition, we perform an "arbitrary fit", a model-free approach to characterizing our measured signal, with less attention to systematic uncertainty.

5.1.3.1 Interferometer branch

If the isolating foil moves in an attractor-correlated way, we need to know how much influence it may have on the pendulum. An optimal approach would have the foil interferometer operating continuously and in-range; it would then be possible to add a joint fit for the foil-motion-to-torque constant to the global fit for new physics. As we only had partial coverage with the interferometer (Section 2.3.9.1), model-building is required.

Interferometer foil-motion to torque constant As detailed in Section 2.3.9.4, we displaced the attractor into contact with the foil and measured the attractor-displacement to torque transfer function.

Attractor-position to foil-motion fit As detailed in Section 10.1, we use the data from times when the interferometer was in-range to generate a piecewise-linear model of the foil displacement as a function of attractor position. The fit is bootstrapped to quantify the uncertainty in the extracted fit parameters.

Foil-driven torque vs foil position model/correction By combining the foil-motion-to-torque calibration with the modeled foil displacement as a function of attractor position, we construct a foil-motion-related torque correction as a function of attractor position, with bootstrapped uncertainties, which can be applied to the entire dataset.

5.1.3.2 Yukawa fit with systematics

As detailed in Section 5.8, the science signal and estimates for the magnitude of systematic effects are fit jointly. This bootstrapped fitting approach yields a distribution of best-fits in $\alpha - \lambda$ space. The many bootstrapped fits of the attractor-driven foil-motion torque correction are individually applied to these fits, allowing direct propagation of the correction uncertainty into the final fit results.

5.1.3.3 Confidence intervals

There are many ways to draw two-dimensional confidence intervals from a distribution of best-fit points. We extract confidence intervals using an algorithm very similar in behavior to the "pick a lambda, fit for alpha" method that is traditional for inverse-square law tests, as seen in Section 5.9.

5.1.3.4 "Arbitrary fit"

Just as for the foil displacement as a function of attractor position, we make a model-free piecewise-linear fit for the pendulum torque as a function of attractor position. Furthermore, by subtracting the foil-motion torque correction from this fit, we are able to compensate for any foil-motion systematic. This is perhaps the most intuitive and general result generated by the experiment.

5.2 Torque Calibration

The torque scale of the instrument is set gravitationally. Throughout the operation of the experiment the revolving gravitational calibration masses (Section 2.3.7) inject a pendulum torque of amplitude 17 fN-m.

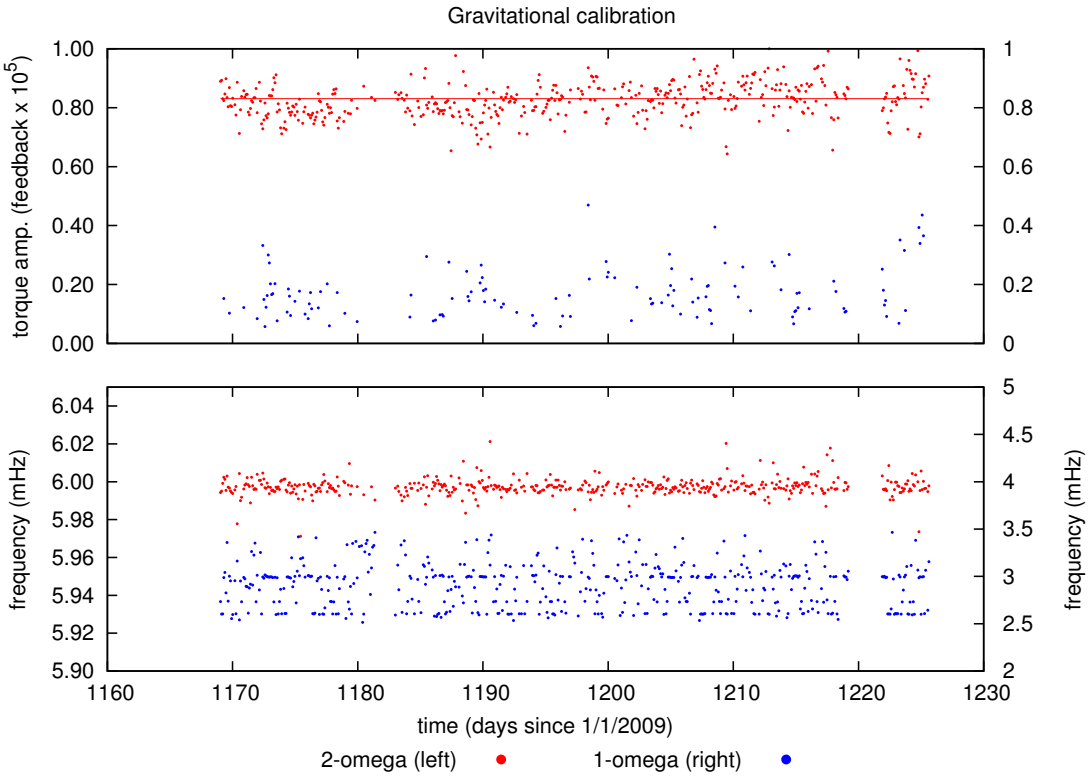

Figure 5.2: Gravitational calibration.

The data are broken into 9600 s chunks and fit for a sinusoidal signal. The signal occurs at twice the quadrupolar gravitational calibrator's rotational frequency (2ω); a fit is also incorporated for the 1ω signal. Both frequencies are allowed to float. The fit is an ordinary least-squares fit of $A\cos(\omega t) + B\sin(\omega t) + C$ to each chunk. The frequency is varied through a downhill binary search to minimize χ^2. The 1ω and 2ω signals are fit independently.

The mean amplitude (solid line in Figure 5.2) $\sqrt{A^2 + B^2}$ throughout the experiment is used to compute the fixed measured-torque to actual-torque conversion ratio. The fitted amplitude is (8.31 ± 0.03) μN-m/AFU (arbitrary feedback unit). The uncertainty stated here is statistical; the systematic uncertainty is $\sim 10\%$. This may be converted to an approximate effective spring constant for the feedback loop of (52.7 ± 0.2) nN-m/rad, which should be compared with the fiber κ of 20 nN-m. The time variation of the calibration signal is most-probably understood as slow drifts in the pendulum/foil/electrode spacing/alignment. It is also possible that some of the electronics involved in the feedback loop have had small gain drifts over time.

5.2.1 Calibration Systematics

As originally-designed, the calibration system had a magnetic-relay "cycle-mark", which would allow us to constrain the phase of the rotating attractor's motion. Both because the cycle-mark had reliability trouble and because the rotating magnet and associated electrical signals could inject a false signal of its own, the cycle-mark system was disabled, and the rotating magnet removed.

The largest possible systematic signals (tilt and direct magnetic coupling to the pendulum magnetic moment) should appear at the rotating attractor's 1-ω frequency, not the critical 2-ω gravitational signal frequency. While a resolved 1-ω signal would occasionally appear, it was inconsistent and often absent; the 2-ω calibration signal was far more consistent in fractional amplitude (see Figure 5.2). While an unintended 2-ω signal is possible, perhaps through modulation of the background magnetic field, the acoustic environment (an acoustic modulation from the gravitational calibration system was seen during intentional driven-loudspeaker studies of the foil interferometer), or the thermal environment, any of these effects should be small.

Most-importantly, as this is a null experiment, with no absolute gravitational subtraction required, the calibration line is used primarily to set the torque scale and to show that the torque scale remains approximately stable throughout the experiment. For this reason, systematics on the calibration signal are not of grave concern.

5.3 Timing

Timing is an important weakness to this particular measurement.

In a major effort to guarantee that no electronic systematic coupling could occur between the attractor drive system and the pendulum and interferometer readout systems, the computers were electrically isolated. They share a common transformer-isolated power source, and a common electrical ground through a starred grounding arrangement (Section 2.3.11), but no other connections. Of particular concern was the isolation of the analog-to-digital converting boards on each data-acquisition system; no shared signals were allowed between them. This arrangement appears to have worked very well (see Section 9.4). The crux of this arrangement is understanding the synchrony of the three systems' timing.

In retrospect, it is clear that an investment of the ∼week of work required to implement and debug a functional optically-isolated time distribution system would have paid great dividends. Instead, we chose to rely upon the gold-standard Network Time Protocol (NTP) time synchronization protocol [106] supported (we thought) by every major computer operating system. As designed, the three computers would resynchronize their clocks with an external reference (either time.nist.gov, or a common local time server referenced

to time.nist.gov) every 3141 seconds. The computer clocks' relative drift rates are less than 10 s/day, so continuous resynchronization would yield overall synchronization at better than the 1 s level at all times. Had this worked correctly, there would be no important timing concerns. In practice, the Windows Time service on at least one computer intermittently failed, forcing us to pay much closer attention to timing issues. With retrospective research we note that Microsoft states [107]:

> *"Windows XP and later versions of the Windows operating system implement NTP. We do not guarantee and we do not support the accuracy of the W32Time service between nodes on a network. ... The W32Time service cannot reliably maintain sync time to the range of one to two seconds. Such tolerances are outside the design specification of the W32Time service."*

Caveat emptor.

There are at least ten clocks relevant to this analysis: the remote NIST-referenced Network Time Protocol (NTP) server, the user-space clocks on each of the three DAQ systems, the system clocks on each of the three DAQ systems, and the hardware clocks on each of the DAQ cards. The torsion-balance and attractor computers are of greatest importance, and share a common software codebase.

The user-space clocks are, in principle, resynchronized with the NTP clock regularly. The user-space clock values are written into the run header files at the beginning of each recorded run. The programs then make a call to the timeGetTime() function in the Windows API, and reference all time measurements to this datum. This function returns the number of milliseconds elapsed since the current session of Windows began. Critically, the returned value is from a user-immutable system clock, and not the user-space clock. No changes due to NTP-related synchronizations (or lack thereof) are evident in the data. Therefore, all runs that began with synchronized clocks share a common datum. At the ends of runs which terminated normally, the user-space clock (which is altered by NTP) is again written into the header file. This allows a cross-check on the duration of the run by cross-comparing the user-space and system-clock durations.

Each DAQ computer has both a system clock and a hardware clock in the DAQ card. The data-taking cadence is defined by the DAQ card. For both the pendulum-computer's internal NI-6259 and the two external USB NI-6211's used on the attractor and interferometer computers, the manufacturer guarantees the clock stability to 50 ppm, or 4 s/day. In Figure 5.3, we use the NTP timing data to characterize the tick-rate stability of the system clocks and then use the stability of the system clocks to characterize the stability of the DAQ hardware clocks. Furthermore, we use the NIST-synchronized NTP data to correct both the pendulum and interferometer computer's Windows clock tickrates to guarantee that they tick at equal rates.

Once we have characterized the clocks' stability, we can synchronize the clock rates. Of note is the fact

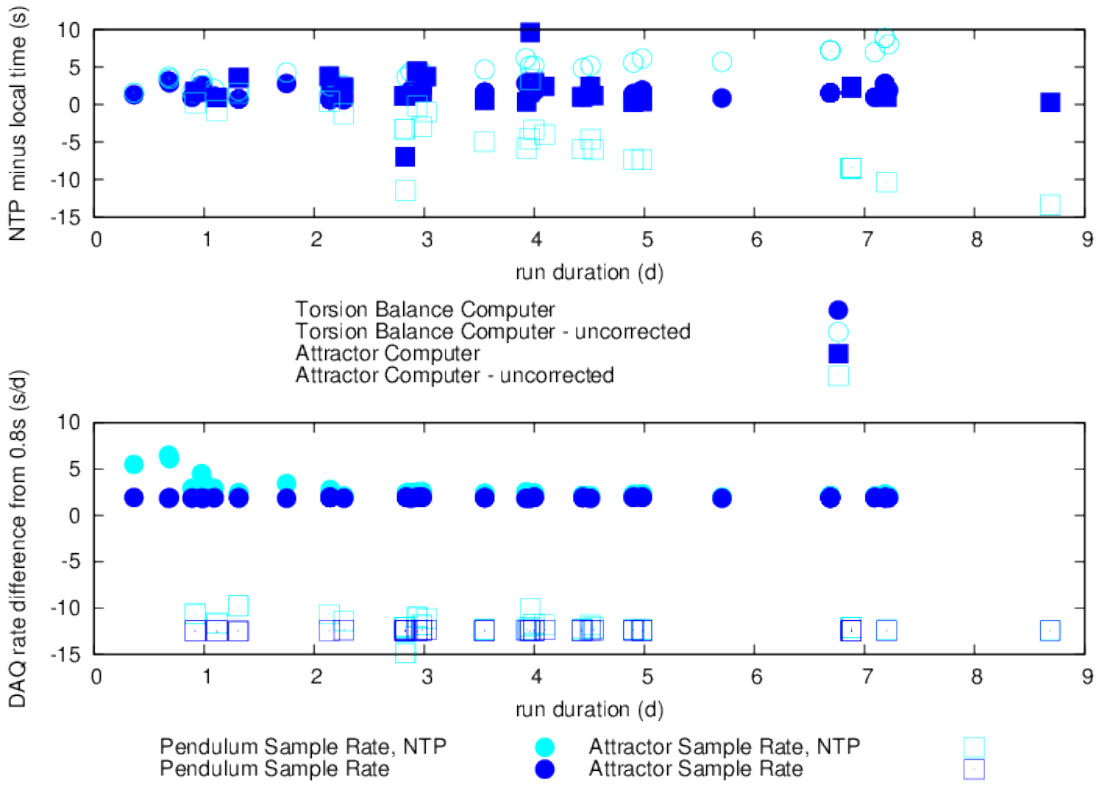

Figure 5.3: Clock rates: corrected clock rates (upper panel) and demonstrating the long-term stability of DAQ sample rates (lower panel)

that the attractor computer's DAQ sample rate differed by 10^{-4} from the other two computers. Fortunately, the attractor drive was timed by the attractor computer's clock, and not the DAQ clock, so no additional clock shifts were necessary, as the lock-in algorithm uses computer-clock time, not sample number.

With the clock rates matched, we must then align the zeroes of each clock at the start of each run. The clocks were manually observed to be synchronous with sub-second precision in almost every run. In a few cases, they were observed to be asynchronous with as much as a 25 s offset. To be conservative, we assume that no run is synchronized to better than 25 s. Dropping 50 s per step, or 100 s per period, is a brutal but acceptable step, as the attractor-motion period is 256 s. The installation of a simple time-distribution system will completely remove this source of uncertainty in the future, making sub-second synchronization straightforward.

A minor concern was the occasional "clock roll". The timekeeping on the attractor and pendulum DAQ systems was referenced to the Windows clock, which reports time as a 32-bit integer number of milliseconds. If the DAQ program runs for more than 49 days without restarting, the 32-bit integer will overflow. These clock rolls are easy to detect and correct, so the analysis software automatically detects and corrects the few occasions on which they occur.

5.4 Signal Extraction

This section describes the way in which more than 5998592 seconds of raw data are converted from three independent raw timeseries into 2929 'cuts' for further analysis.

5.4.1 Synchronization

Data are imported from three separate sources (torsion balance, attractor control, interferometer), and trimmed in length to present only data with complete temporal overlap between relevant sources. All three sources sample with a fixed 0.8 s sample interval, so any extra timing mismatch (beyond the discussion in Section 5.3) is at most 1.6 s.

5.4.2 Post-acquisition filtering

In this field, it is traditional to fit for and subtract the gravitational calibration tone, remove low frequency drifts on a per-cut basis through linear fits and average away unnecessary high-frequency information (often in hardware). In combination, these actions amount to the construction of a complicated filter, with interesting frequency dependence and phase behavior. The low-frequency linear fit is of particular interest, as a linear

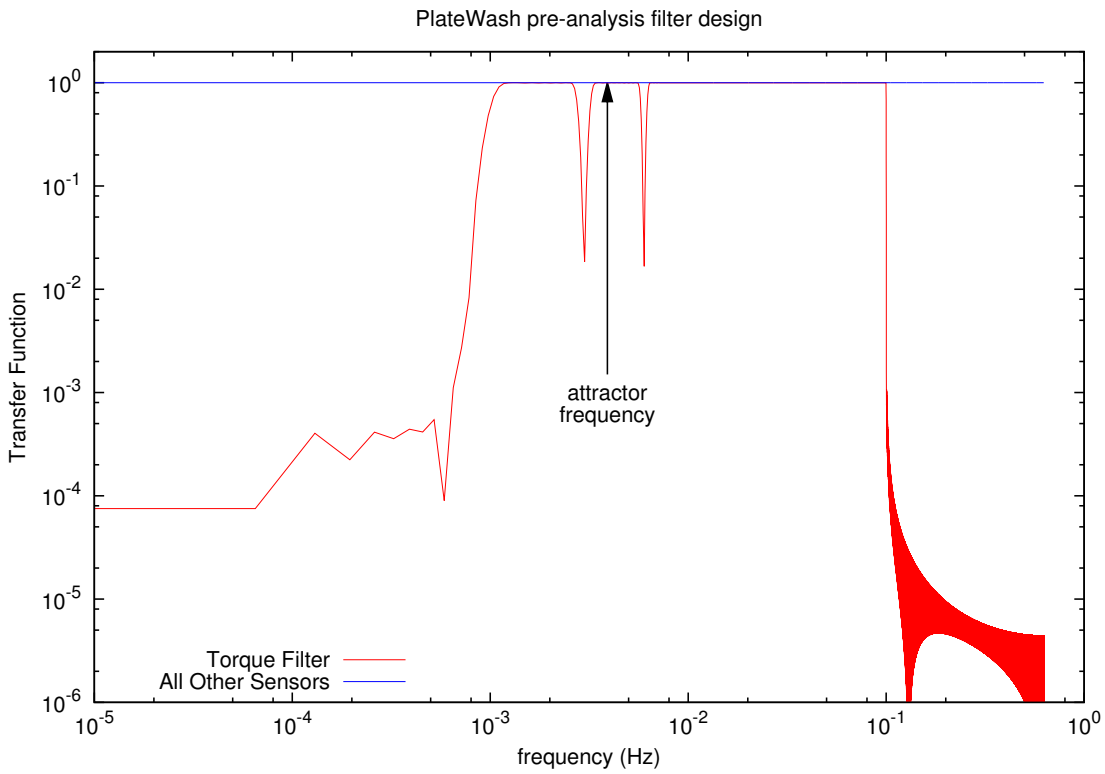

Figure 5.4: Transfer functions of software low-pass filters.

fit is not orthogonal to a Fourier series. We trade the complex combination of filter-like fits for a single non-trivial filter.

Data are fed through FIR filters of length 9600 samples (10 samples × 256-second attractor period × 3 cuts ÷ 0.8 s per sample). For a finite-impulse-response filter (FIR) of this duration, the band-stopping power can be larger than 1000 (60 dB). The filters are designed by the `octave-forge` function '`fir1`' [108]. The function designs 'windowed filters' with linear phase response. Filters with linear phase response have frequency-independent phase delays [109, 110]. The torsion balance science signal is sent through a filter with passband from 1-100 mHz with notch filters at the calibration tone's 1-ω and 2-ω frequencies (3 and 6 mHz). All other sensors to be analyzed in-sync with the attractor are fed through a low-pass filter with a corner frequency at 625 mHz, coincident with the Nyquist frequency. This almost-inconsequential filter is applied to guarantee that all locked-in sensors share timebase/phase with the filtered torque signal. The first $1.5 \times$ (filter length) samples are removed from the beginning of each filtered run to remove filter-initialization effects. The filter transfer function is shown in Figure 5.4 and the torque filter's coefficients are plotted in Figure 5.5. Periodograms of the filters applied to experimental data are shown in Figure 5.6.

This procedure correlates three cuts. As the total science dataset consists of 2929 cuts, the effect of cut-to-cut correlation is extremely small in aggregate. In addition, as attractor motion was either random or

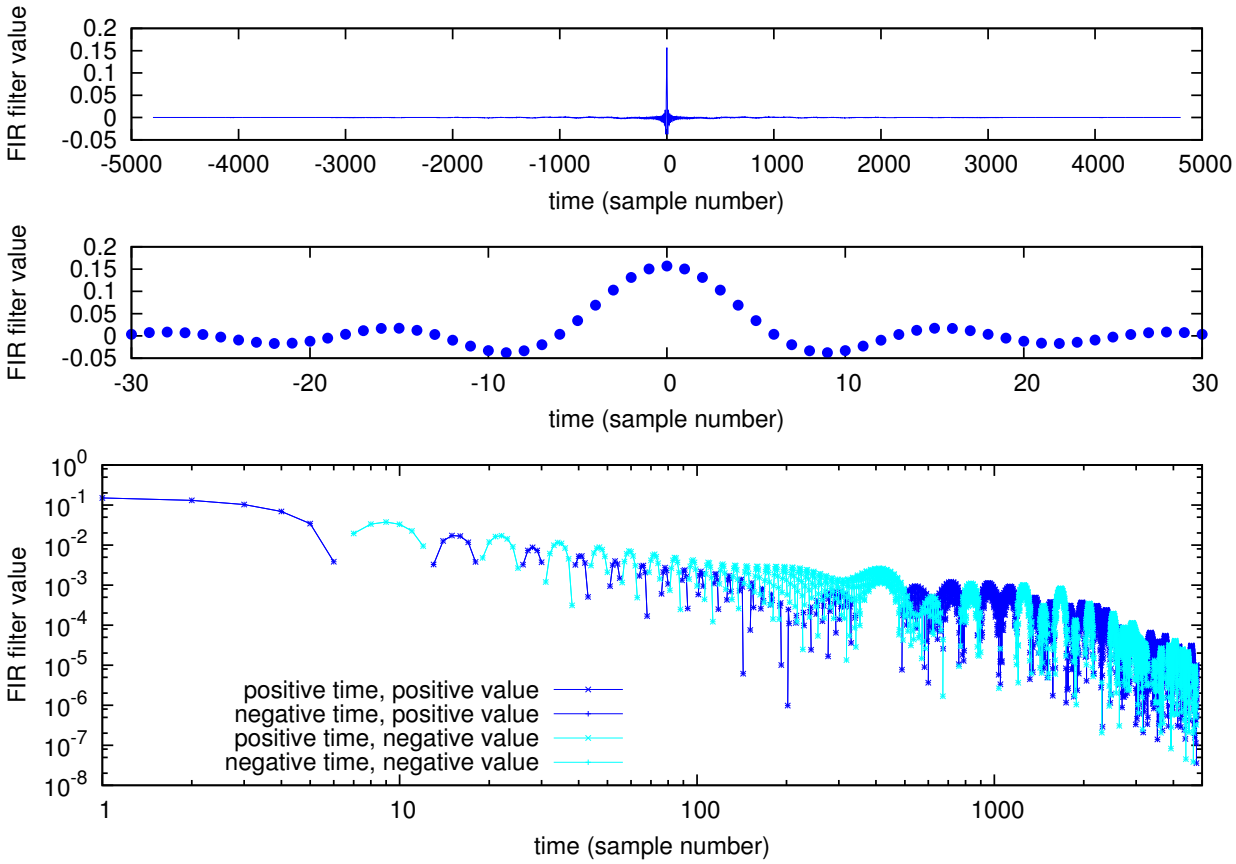

Figure 5.5: Time-domain structure of torque filter coefficients. The prominent "sinc" structure is the low-pass filter. The calibration-line filters and high-pass filters are more subtle and act at periods of ~ 200 samples and below.

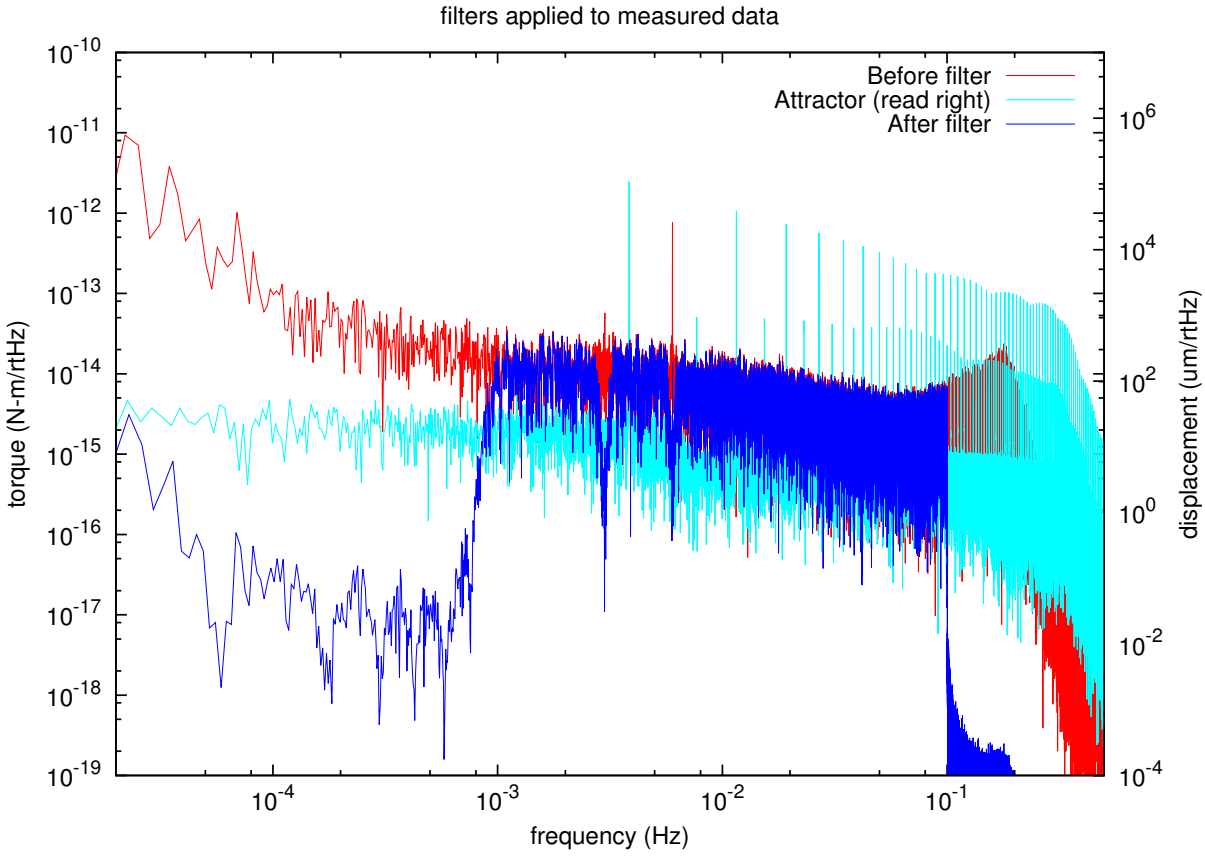

Figure 5.6: Torque data before and after filtering. Attractor included to show relevant frequencies. The attractor-drive system was operational for these data, but the isolating 'apparatus valve' was closed, preventing any attractor motion or gravitational signal.

uniform, it should not introduce any systematic uncertainty. The 3-cut length of the FIR filter was chosen by balancing the need to sufficiently attenuate and frequency-select the injected calibration signal (pushes to longer-duration filters) with the need to reject the effects of spikes/kicks (earthquakes and attractor-foil contact). A spike, with amplitude orders of magnitude larger than the calibration tone, contaminates the entire FIR filter length with ringing. With a 3-cut filter, a single spike makes more than two hours data unusable; a longer filter (5-cut is the next symmetric choice) would be even less forgiving.

5.4.3 Edge finding

We must synchronize the data analysis cuts with the attractor motion. The first step in synchronizing is to find the square-wave edges. This is done by passing a simple edge-detection algorithm along the filtered data, looking for the first time the attractor's position passes through the mean position of the data set. Once this edge is found, the edge is projected backward in time, modulo the attractor step frequency, to find the first step. While crude, it is simple, and has proven reliable in synchronizing the analysis cuts with the data.

5.4.4 Dead Times

In this section, we determine the times during the experiment during which we believe the torque signal might be contaminated with information associated with the act of changing attractor position. During these times, which we deem to be 'dead', no data are used for scientific analysis.

5.4.4.1 Important durations/times

With each attractor position change, it takes a finite amount of time for the pneumatic drive to change position ($\sim$ 27 s, including the DAQ filter, see Figure 5.7). Furthermore, the attractor control system begins to actuate the modulator valve 10 s before the position change. In addition, the pendulum feedback has a finite response time of about 20 s (See Figure 5.8). The pneumatic post-acquisition filter low-pass has a corner frequency at 625 mHz. The pneumatic DAQ filter has a lowpass at 333 mHz. The synchronization uncertainty among the three acquisition computers is at most 1.6 s (two sample-times). To account for clock drifts and offsets, we add $\pm$25 s. The autocollimator (and on-line feedback) DAQ filter corner was 1 Hz. The pendulum feedback control loop had an internal low-pass filter at 200 mHz. The feedback proportional-term was anti-aliased at 333 mHz (for recorded data only). The acausal torque passband FIR filter 100 mHz cutoff frequency corresponds to a 5 s dead time before and after each switch.

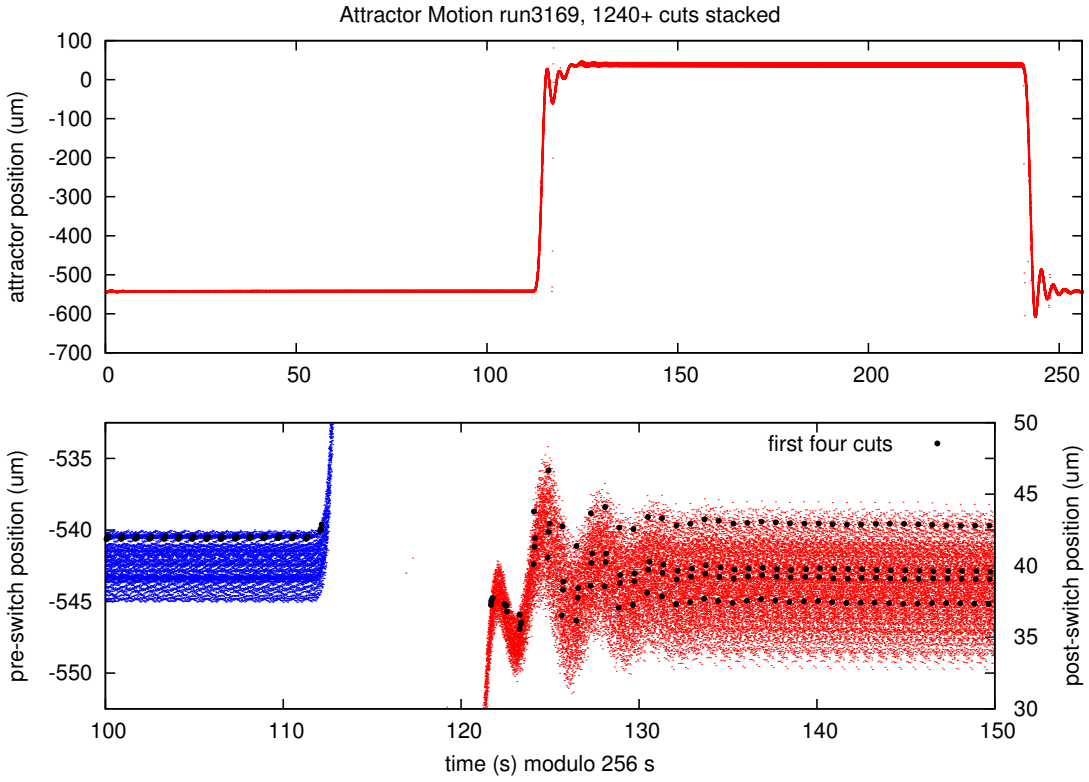

Figure 5.7: Attractor motion, convolved with DAQ IIR low-pass response. Scatter in the lower frame is from ambient air pressure variation (affects both blue and red), and variation in pneumatic-drive pulse-width modulation phase (red only).

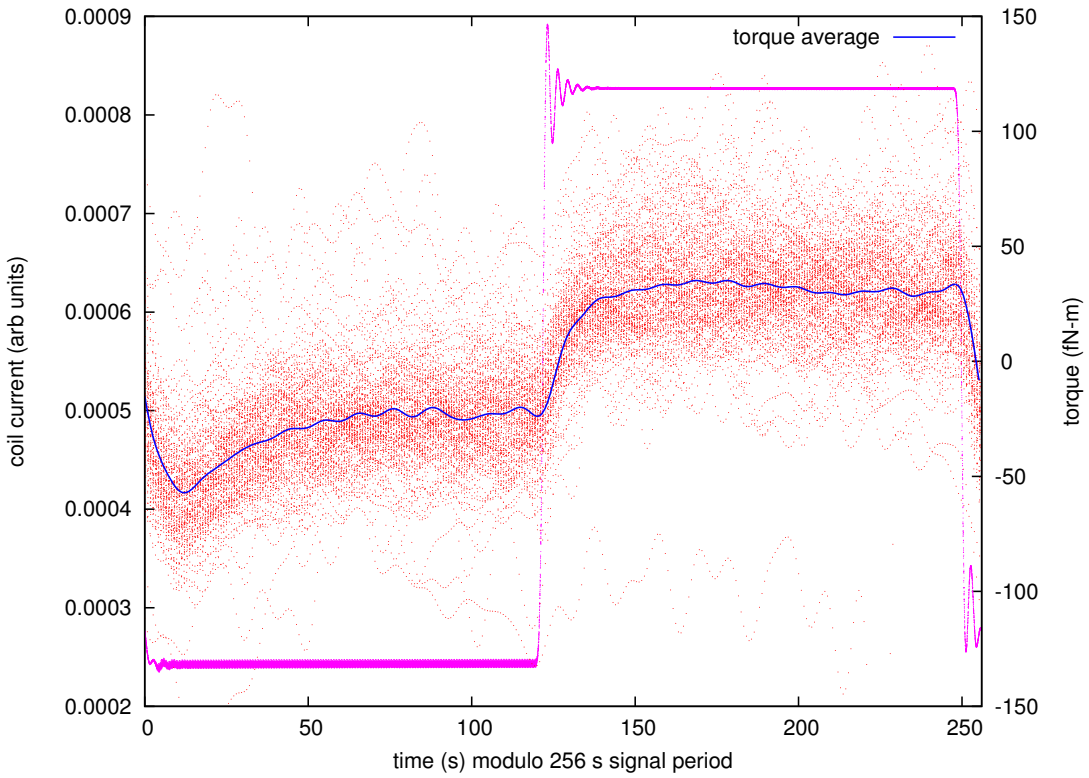

Figure 5.8: Injected torques for dead-time determination after all filters. The apparently-solid pink line is the injected coil current. The scattered red dots are an entire torque dataset plotted with horizontal-axis modulo the injected square-wave period. The blue curve is the binned average of the experimentally-measured points.

5.4.4.2 "End-to-End" test

The general problem of filter combination is simple in frequency-space, but difficult in the time domain.

The experiment itself provides a solution to the problem of determining the settling-time of the overall torque feedback system. The pendulum has a magnetic moment (see Section 2.4.2.5). Using a coil placed around the apparatus for systematic tests, it is possible to produce arbitrary torques on the pendulum. The attractor position is measured using an analog electronic pressure gauge (see Section 2.3.6). By combining the gauge with the coil and an amplifier, it is possible to make an exact and magnified simulation of a gravitational signal that is linear in attractor position. See Figure 5.8.

5.4.4.3 Delay times

Using Figure 5.8 and the timing-offset concerns enumerated above, we conclude that it is conservative to choose the following dead times: The "delay time" from the beginning of a position change to the beginning of accepted data is set to 65 s. The "end-pad time" from the end of accepted data to the end of the position change is set to 35 s.

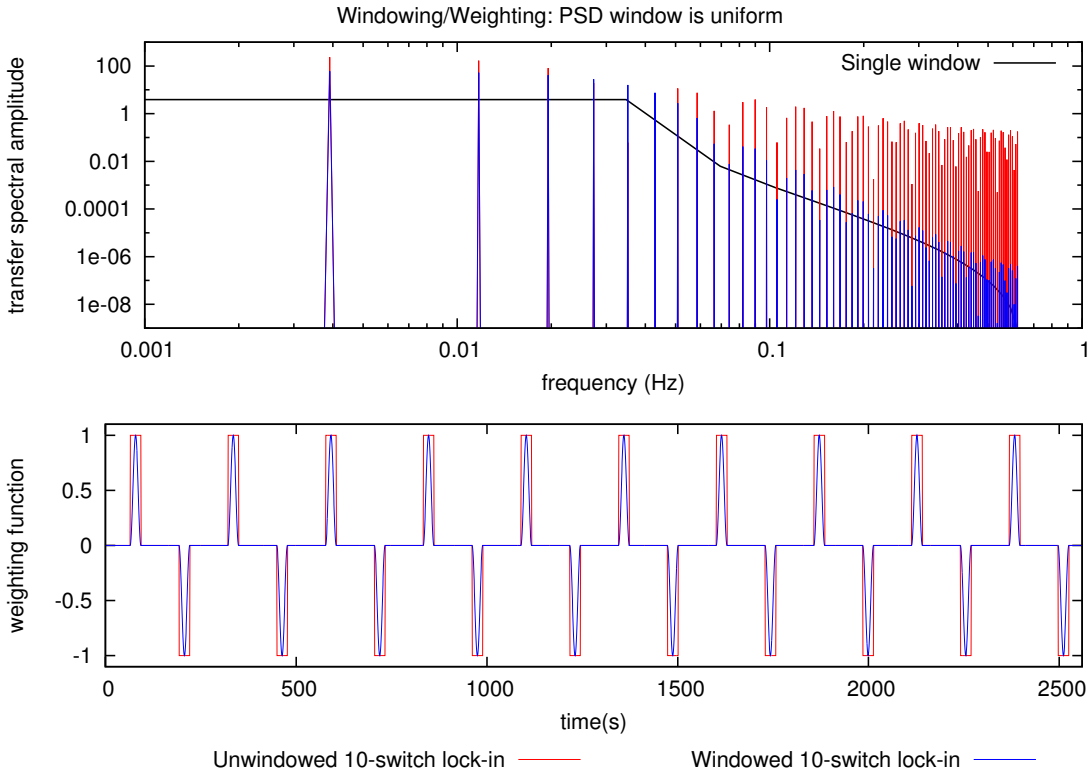

Figure 5.9: The effect of windowing in a lock-in measurement. The solid black line is the periodogram of a single window. The periodogram of the windowed lock-in signal is approximately that of the convolution of the window with a square-wave signal.

5.4.5 Windowing

Windowing is familiar to many as a method for reducing the magnitude of spectral leakage in Fourier transforms. It is also useful in the time-domain as a low-pass filter, reducing the sensitivity of an averaged value to high frequency fluctuations at a modest expense in statistical resolving power. In this analysis, a Hann window is used [111]. Just as in periodogram estimation, the choice to use a window is much more important than the choice of a particular window. Several windowing schemes were explored, without significant improvement. The final weighting of data used for our analysis is shown in Figure 5.9.

5.4.6 Lock-in

Once the dead times are chosen, a window selected, and a leading edge found, locking in is straightforward. Beginning at a time prescribed by a leading edge, the analysis program walks through each data stream, multiplies the data found in each 'step period' (1/2 the attractor motion period) by the properly-delayed window function, and computes the weighted average. The resulting average forms a single measurement. There are 20 measurements per cut.

The first measurement, and every-other subsequent measurement, is regarded as an 'A' measurement. The second measurement, and every-other subsequent measurement (the complement of the As), is regarded as a 'B' measurement. Because the pneumatic drive takes about one attractor period to come into equilibrium (if coming from a different A/B position pair in the previous cut), the first A and B measurements are dropped from quantitative analysis.

The average of the A-state measurements and the average of the B-state measurements are stored for each cut. In addition, the standard deviation, maximum, and minimum of each set of As and Bs are stored.

The 'lock-in' is made by taking the difference between the average of the As and the average of the Bs. The difference's mean, standard deviation, maximum, and minimum are all recorded. For the torque channel, these torque differences, in concert with the A and B positions, comprise the science measurement.

5.5 "Triangle Plot"

A clear method of visualization is important to understanding the output of any experiment. For PlateWash, the natural plot is three-dimensional. The experiment measures the difference in pendulum torque between two different attractor positions. So, if we plot $\Delta\tau(\text{start}, \text{end})$, we can plot every torque-difference measurement in a single plot, shown in the middle panels of Figure 5.10.

We refer to these plots as "triangle plots" because, if all is well experimentally, the plot will be reflection anti-symmetric about the ascending diagonal ("$y = x$"), as $\Delta\tau(x, y) = -\Delta\tau(y, x)$ for all y and x, with the corollary that $\Delta\tau(x, x) = 0$ for all x. Furthermore, any conservative interaction will be associative: $\Delta\tau(a, c) = \Delta\tau(a, b) + \Delta\tau(b, c)$. Lines parallel to "$y = x$" are lines of constant position-difference. Vertical lines share constant A position, and horizontal lines share constant B position.

In the lowest panel, we plot the projection of the scatter plot along the diagonal, which allows an alternate and useful view.

In this work, these will be presented only as scatter plots, but if the out-of-page scatter of the points is relatively uniform, it is possible to bin the points locally and plot circular error-bars to add more clarity to the plots. As the PlateWash data have considerable variation in variance over time, it is extremely challenging to plot.

5.6 Excised cuts

In this section, we describe ways in which data expected to be anomalous were excised from the dataset, commonly called "cuts". To prevent a vocabulary collision with our previously-defined "cut", meaning a single

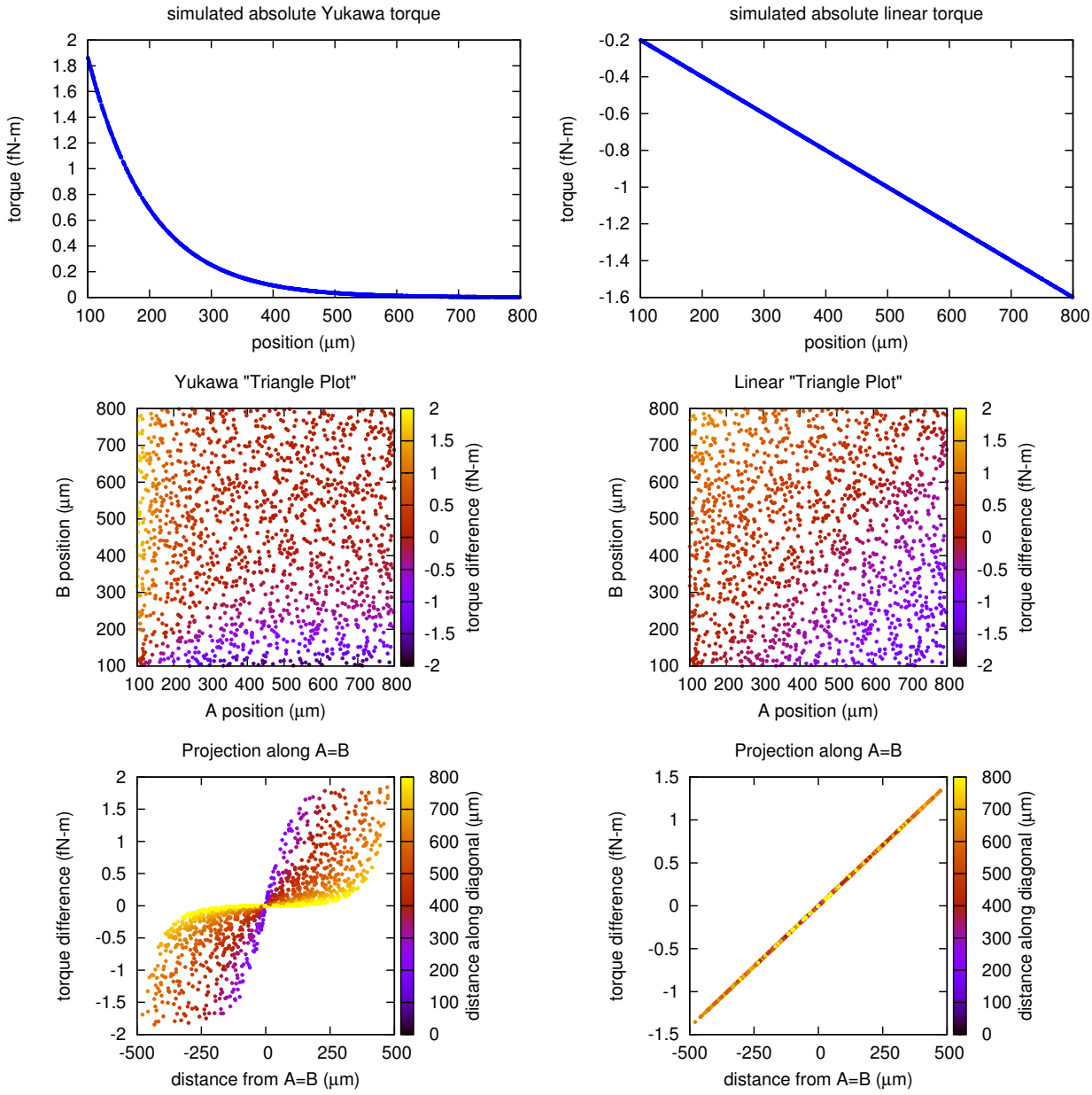

Figure 5.10: 'Triangle plots', for both Yukawa-like and linear torques. The upper frames show the simulated torque-as-a-function of distance curves. The middle frames show the 'triangle plots' themselves; note the difference between them in the upper right corners. The bottom frames show a projection of the torque-difference data along the diagonals ($A = B$) of the triangle plots; this is a helpful view that may make the difference between the plots more clear.

measurement, we refer here only to "excisions".

- To reject a short period of data where the pendulum autocollimator was disabled, causing the recorded torque to read 0 for a continuous stretch of data, we impose a minimum threshold on the torque *uncertainty* of 0.001 fN-m, far below our expected per-cut uncertainty.

- To reject truly non-Gaussian transients from the data (e.g. earthquakes, foil strikes, etc.), we reject cuts whose torque *central value* or *uncertainty* is greater than five times the standard deviation of the torque *central values* of the entire dataset. With this approach, we will not accidentally reject a real signal that persists for more than a few cuts, as might happen with an excision determined from the distribution of torque-cut uncertainties. While this excision seems like an oddity, there are few ways that one can reliably determine outliers while blind. This approach can be viewed as a conservative analog to Chauvenet's Criterion [112], as we only have thousands of measurements in total and expect that our science signal will be buried within the single-measurement statistical noise..

- To prevent any attractor-foil touching signal from entering the dataset, we excise away any cut with even a single attractor point within 20 μm of the expected attractor-foil touching point.

Post-unblinding comment: the order of the latter two excisions should have been exchanged, but were not. We believe that this has an inconsequential effect on the result, and would tend to improve the measurement's sensitivity slightly.

5.7 Bootstrap

The analysis presented here makes extensive use of a statistical method called 'the bootstrap' or 'bootstrapping'. Its primary merit in this context is that it allows straightforward and reasonably robust estimation of confidence intervals. Bootstrapping relies upon the premise that the best estimator for the unknown distribution from which a data set is drawn is the data set itself. The idea is to 'resample' the data to create synthetic data sets which may be passed through any analysis routine developed for the original data set. With repetition, this process generates distributions of important output quantities (best-fit points in the Yukawa $\alpha - \lambda$ space, for example) without recourse to complex error estimation/propagation.

The resampling scheme used in this analysis was to generate synthetic data sets of the same size as the original set by randomly selecting measurements (there are always duplicate picks, except in the vanishingly rare case where the original data are recovered, out of order) from the initial data set. There is a related class of methods called 'the jackknife' that generate statistics by considering complete sets of subsets of data, perhaps by dividing the initial data into subsets, fitting each subset, and drawing statistics from the

fit outputs. The jackknife may be more readily accepted by those uncomfortable with the bootstrap; in the case of this measurement, there are too few measurements to allow meaningful statistics to be drawn from jackknifed fit outputs. Halving or quartering the dataset could be possible, but cutting it into ten pieces would result in too-little coverage in terms of varied attractor displacements.

5.7.1 Implementation

In all instances where the bootstrap was used, this simple Octave function was called:

Listing 5.1: bootstrapData.m

```
% Given an m x n matrix "data", creates (samples with replacement)
% a new n x m matrix "out" from n randomly-selected rows of "data"
function out = bootstrapData(data)
        % generates a column vector of uniform random numbers
        % spanning 1 to n
        r = floor( rand( rows( data ) , 1 ) * rows( data ) ) + 1;
        % uses r to select rows of "data" and copy them into "out"
        out = data(r,:);
endfunction
```

5.8 Model Comparison

The bootstrap makes comparison with any proposed model straightforward. No matter how exotic the fitting function and how sensitive it might be to individual datapoints or non-Gaussianity in the experimental data, it is straightforward to bootstrap the data, fit, bootstrap again, fit, repeating as needed. The distribution of fit outcomes can always be studied, and qualitative understanding is immediate. Settling on an algorithm for generating confidence intervals is, however, more difficult, especially in multi-parameter spaces.

5.8.1 Fitting, χ^2

A weakness of bootstrapped analysis is its reliance upon computational power to produce useful results. If the dimension of a problem gets too high, it is difficult to ensure that fitting parameter spaces are well-explored. It is possible to include systematic effects into bootstrapped confidence intervals by modeling each systematic and injecting the expected amount of random uncertainty into the data to be fitted. If little is known about the systematic effect, however, this approach is challenging.

It is more-reliable to include poorly-known systematics into the fit, fitting over all the data, including both the science runs and the systematic tests. We do this by ascribing a new fictitious force to each systematic

effect, parametrized conservatively as a Yukawa torque, with its own strength and lengthscale. Furthermore, if we're able to compute χ^2 directly, there is a well-accepted approach to the determination of confidence intervals and exclusion limits.

5.8.1.1 Fitting approach – first half

We proceed down a path well-trodden by past experiments.

- Let N be the number of measurements
- Following the torque calculations from Section 2.2, we assume that the torque on the pendulum takes the form

$$\begin{aligned} F(x, B_{1\ldots n}|\alpha_{1\ldots n}, \lambda_{1\ldots n}) &= \pi G h \left(\frac{w^2}{4} - j^2\right)(\rho_h - \rho_l)\,\rho_a \sum_i \alpha_i B_i \lambda_i^2 \\ &\quad \times \left(1 - e^{-(t_p - t_I)/\lambda_i}\right)\left(1 - e^{-t_a/\lambda_i}\right)\left(1 - e^{-t_I/\lambda_i}\right) e^{-x/\lambda_i} + Cx \end{aligned} \tag{5.1}$$

  where $n = 1$ is taken to be the gravity/science signal ($B_1 = 1$, always) and the remaining B_i correspond to externally-modified systematic parameters, each with individual associated α_i and λ_i . The Cx term subsumes all torques that are linear in the attractor stroke.

- To simplify notation, we subsume the finite-thickness corrections into

$$T(\lambda_i) = \rho_a \left(1 - e^{-(t_p - t_I)/\lambda_i}\right)\left(1 - e^{-t_a/\lambda_i}\right)\left(1 - e^{-t_I/\lambda_i}\right)$$

  and constant factors into

$$Q = \pi G h \left(\frac{w^2}{4} - j^2\right)(\rho_h - \rho_l)\,,$$

  yielding

$$F(x, B_{1\ldots n}|\alpha_{1\ldots n}, \lambda_{1\ldots n}) = Q \sum_i \alpha_i B_i \lambda_i^2 T(\lambda_i) e^{-x/\lambda_i} + Cx$$

- As we measure the difference in torque between two different attractor positions, we define :

$$G(x_A, x_B \ldots) = F(x_A, \ldots) - F(x_B, \ldots) \tag{5.2}$$

- As a reminder, we measure a torque difference between positions x_A and x_B, and all external systematic

parameters $B_2, \ldots, B_n$, with uncertainties $\sigma_{x_A}, \sigma_{x_B}, \sigma_{B_1}, \ldots \sigma_{B_n}$ as:

$$\Delta\tau(x_A, x_B, B_1, \ldots, B_n, \sigma_{x_A}, \sigma_{x_B}, \sigma_{B_1}, \ldots, \sigma_{B_n})$$

- We treat all measured uncertainties as Gaussian, distributed as $P(a, \bar{a}, \sigma_a) = \frac{1}{\sigma_a\sqrt{2\pi}} e^{-(a-\bar{a})^2/2\sigma_a^2}$

We pause on our path to confront nonlinearity.

5.8.1.2 Linearity and limitations

It is traditional in this field [16, 113] to assume that the effective uncertainty $\tilde{\sigma}_G$ for a model quantity $G(a_1, \ldots, a_n)$, with measured uncertainties $\sigma_{a_1}, \ldots, \sigma_{a_n}$ can be estimated (for the j-th measurement) by

$$\tilde{\sigma}^2_{G_j} = \sum_{i=1}^{n} \left(\sigma_{a_{i_j}} \left. \frac{\partial G}{\partial a_i} \right|_{\{a_1,\ldots,a_n\}=\{a_{1_j},\ldots a_{n_j}\}} \right)^2 \tag{5.3}$$

The χ^2 is then de-weighted by

$$\chi^2 = \sum_{j=1}^{N} \frac{(\Delta\tau_j - \bar{G}_j)^2}{\sigma^2_{\Delta\tau_j} + \tilde{\sigma}^2_{G_j}} \tag{5.4}$$

This approach is appropriate when the linearization of the function G is appropriate.

In the case of PlateWash, it is clear from inspection of Equation 5.1 that when the effective Yukawa interaction range λ becomes comparable to, or smaller than, the precision with which the attractor position is determined, the experimental sensitivity should be degraded. In the linearization approach, this corresponds to a diverging $\tilde{\sigma}^2_f$. Furthermore, linearizing $e^{-x/\lambda}$ when $\sigma_x \gtrsim \lambda$ is a recipe for trouble. With aid and inspiration from Jason Detwiler, we computed the full non-linear variance of Equation 5.2, shown in full detail in Appendix A (the "Detwiler Trick" is worth a look). While we succeeded, this entire approach has a fatal flaw; the flaw is in Equation 5.3.

5.8.1.3 The Flaw

The flaw is easy to see qualitatively. In Equation 5.4, if σ^2_G diverges, χ^2 is driven to zero. Any χ^2-minimizing fitter will, given the opportunity, exploit this to favor locations in which σ^2_G is huge. Thus, fits with smaller λ are incorrectly preferred over perfectly reasonable fits with larger λ. Nikolai Tolich was the first of those with whom we conferred to reach a quantitative understanding of the problem.

Explanation: χ^2 is not Likelihood If we take, as above, all our uncertainties to be distributed according to $P(a, \bar{a}, \sigma_a)$, then the probability distribution for our torque-difference measurements should be

$$P(\lambda_1,\dots\lambda_n,\alpha_1,\dots\alpha_n) = \int\prod_{j=1}^{N}\Big(e^{(G(x_A,x_B,B_1,\dots B_n,\lambda_1,\dots\lambda_n,\alpha_1,\dots\alpha_n)-\Delta\tau_j)^2/\sigma^2_{\Delta\tau_j}} \times P(x_A,\bar{x}_{A_j},\sigma_{x_{A_j}})P(x_B,\bar{x}_{B_j},\sigma_{x_{B_j}})\prod_{i=1}^{n}P(B_i,\bar{B}_{i_j},\sigma_{B_{i_j}})\Big)\,dx_A dx_B dB_1\dots dB_n,$$

if you're Bayesian. If you're a Frequentist, and you prefer profile likelihoods,

$$L(\lambda_1,\dots\lambda_n,\alpha_1,\dots\alpha_n) = \max_{x_{A_j},x_{B_j},B_{2_j},\dots,B_{n_j}\forall j}\left(\log\left[\prod_{j=1}^{N}\Big(e^{(G(x_{A_j},x_{B_j},B_1,\dots B_{n_j},\lambda_1,\dots\lambda_n,\alpha_1,\dots\alpha_n)-\Delta\tau_j)^2/\sigma^2_{\Delta\tau_j}} \times P(x_{A_j},\bar{x}_{A_j},\sigma_{x_{A_j}})P(x_{B_j},\bar{x}_{B_j},\sigma_{x_{B_j}})\prod_{i=1}^{n}P(B_{i_j},\bar{B}_{i_j},\sigma_{B_{i_j}})\Big)\right]\right)$$

$$L(\lambda_1,\dots\lambda_n,\alpha_1,\dots\alpha_n) = \max_{x_{A_j},x_{B_j},B_{2_j},\dots,B_{n_j}\forall j}\left(\sum_{j=1}^{N}\left(\frac{G(x_{A_j},x_{B_j},B_1,\dots B_{n_j},\lambda_1,\dots\lambda_n,\alpha_1,\dots\alpha_n)-\Delta\tau_j)^2}{\sigma^2_{\Delta\tau_j}} + \frac{(x_{A_j}-\bar{x}_{A_j})^2}{\sigma^2_{x_{A_j}}} + \frac{(x_{B_j}-\bar{x}_{B_j})^2}{\sigma^2_{x_{B_j}}} + \sum_{i=1}^{n}\frac{(B_{i_i}-\bar{B}_{i_j})^2}{\sigma^2_{B_{i_j}}}\frac{1}{\sigma_{B_{i_j}}\sqrt{2\pi}}\right)\right)$$

The Frequentist approach oversimplifies the problem by choosing the maximum alone, but it makes clear the resolution to "the flaw"; here, no value of λ_i can cause the likelihood to diverge. σ_x and λ do not appear in any ratios.

The profile-likelihood maximization would be a challenging one, as it requires global maximization over x_{A_j}, x_{B_j}, B_{2_j}, … , B_{n_j} for all j, which numbers in the thousands. The Bayesian approach would be perhaps more tractable if evaluated numerically through a Monte Carlo or Markov-chain approach.

Implication - giving up on quantitative use of χ^2 Without a rigorous way to use χ^2 in a way that bakes in uncertainties of dependent variables, we cannot use χ^2 to pull out meaningful confidence intervals. So, we return to the simple land of Monte Carlo tricks. We'll keep the linearized variance approach for the B_i, as it is appropriate there, but we resort to Monte Carlo to determine the effect of σ_x uncertainties.

To do so, we will bootstrap the torque data, generating a synthetic data set. When we do, we will "fuzz" the measured x_A and x_B positions by a normally-distributed amount determined by each measurement's

σ_{x_A} and σ_{x_B}. This fuzzing process, in concert with the bootstrapping of the torque data, should distribute the best-fit points as expected from a more-sophisticated approach to error analysis. This choice tables until Section 5.9 the important problem of confidence-interval determination.

Advantage - systematic metrology uncertainty The bootstrapped handling of uncertainties allows a trivial inclusion of a traditionally difficult-to-handle uncertainty, colloquially known as "horizontal systematic error-bars". As the determination of the absolute separation of the pendulum and foil, and hence the pendulum and attractor is necessarily uncertain (see Table 3.1), this must be reflected in degraded confidence intervals at small λ. From the perspective of our global fit, this measurement is unchanging, making it a systematic uncertainty (and, indeed, it contains its own systematic uncertainties). The "fuzzing" process introduced above is readily adapted to this concern.

For each iteration of the bootstrapped fit, in addition to individually randomly displacing each position measurement by its individual statistical uncertainty, we displace all positions by a single random amount, distributed by the known metrology uncertainty. This approach naturally incorporates our absolute position uncertainty without degrading our known and meaningful differential position sensitivity.

5.8.1.4 Fitting approach – second half

With our treatise on non-linearity complete, the fitting plan continues (see Appendix A for what happens with non-zero σ_x).

- Compute $\bar{F} = \int_{-\infty}^{\infty} F(x, B_{1\ldots n}, \alpha_{1\ldots n}, \lambda_{1\ldots n}) P(B_i, \bar{B}_i, \sigma_i) dB_i$ (see Section 5.8.1.5)
- Compute $\bar{G} = \bar{F}_{x_A} - \bar{F}_{x_B}$
- Compute $\sigma^2_{F_{x_A}} = \int_{-\infty}^{\infty} F(x, B_{1\ldots n}, \alpha_{1\ldots n}, \lambda_{1\ldots n})^2 \prod_i P(B_i, \bar{B}_i, \sigma_i) dB_i - \bar{F}^2$ and $\sigma^2_{F_{x_B}}$ (see Section 5.8.1.6)
- Compute $\tilde{\sigma}^2_G = \sigma^2_{F_{x_A}} + \sigma^2_{F_{x_B}}$
- Compute χ^2

$$\chi^2 = \sum_{measurements} \frac{(\Delta\tau - \bar{G})^2}{\sigma_\tau^2 + \sigma_G^2}$$

- Minimize χ^2 as a function of C, $\alpha_1, \ldots, \alpha_n$, and $\lambda_1, \ldots, \lambda_n$. The minimization was done via simulated annealing. Downhill simplex, Levenburg-Marquardt, and sequential quadratic programming methods all were susceptible to being trapped in local minima. Simulated annealing is computationally-intensive, but has proven more robust in recovering injected test signals.

- To extract science, consider the distribution of the (α_1, λ_1) best-fit values over many bootstrapped-and-fuzzed fits.
- To extract understanding of systematic uncertainties, study the distribution of all remaining fitted variables.

5.8.1.5 $\bar{F}$

The integrals needed for these sub-subsections are worked out in Appendix A.2.

$\bar{F} = \int_{-\infty}^{\infty} F(x, B_{1\ldots n}, \alpha_{1\ldots n}, \lambda_{1\ldots n}) \prod_i P(B_i, \bar{B}_i, \sigma_i) dB_i$

$\bar{F} = Q \sum_i \alpha_i \bar{B}_i \lambda_i^2 T(\lambda_i) e^{-\bar{x}/\lambda_i} + C\bar{x}$

5.8.1.6 σ_G^2

$\sigma_G^2 = \sigma_{F_{x_A}}^2 + \sigma_{F_{x_B}}^2$

$\sigma_{F_x}^2 = \int_{-\infty}^{\infty} F(x, B_{1\ldots n}, \alpha_{1\ldots n}, \lambda_{1\ldots n})^2 \prod_i P(B_i, \bar{B}_i, \sigma_i) dB_i - \bar{F}_x^2$

Let $L_{ij} = 1/\lambda_i + 1/\lambda_j$.

$$\begin{aligned}
\sigma_{F_x}^2 &= Q^2 \sum_i \sum_{j \neq i} \alpha_i \alpha_j \bar{B}_i \bar{B}_j \lambda_i^2 \lambda_j^2 T(\lambda_i) T(\lambda_j) e^{-\bar{x} L_{ij}} \\
&\quad + Q^2 \sum_i \alpha_i^2 (\bar{B}_i^2 + \sigma_{B_i}^2) \lambda_i^4 T(\lambda_i)^2 e^{-2\bar{x}/\lambda_i} \\
&\quad + 2CQ \sum_i \alpha_i \bar{B}_i \lambda_i^2 T(\lambda_i) e^{-\bar{x}/\lambda_i} (\bar{x}) \\
&\quad + C^2 \bar{x}^2 - \bar{F}_x^2
\end{aligned}$$

This works, but speed, stability, and simplicity urge us farther.

$$\sigma_{F_x}^2 = Q^2 \sum_i \alpha_i^2 \lambda_i^4 T(\lambda_i)^2 \sigma_{B_i}^2 e^{-2\bar{x}/\lambda_i}$$

5.9 Confidence Interval Algorithm

There is no unique method for drawing two-dimensional confidence intervals. The method we selected was chosen for its clarity after exploring more sophisticated methods (kernel density estimators among them). The most important lesson from our investigation of differing methods is the notion that, *a priori*, the Yukawa α and λ have nothing to do with one another. Any attempt to constrain them jointly makes a prior

assumption about their relationship. The long-standing tradition in this field is the "pick a lambda and fit for alpha" approach, which sidesteps any presupposition of covariance. What we do is in that spirit.

The outcome of the bootstrapped Yukawa fits is a distribution of best-fit points in the $\alpha - \lambda$ plane. The final form of our algorithm was suggested by Matt Turner, earning the moniker, "Turner Smoothing". Working from small λ to large λ, we compute the mean m_α and standard deviation σ_α of the αs of the subset S composed of the first N (for this thesis, 22) best-fit points. N is chosen as the square-root of the number of bootstrapped points in order to limit over-zealous fluctuations of the confidence interval curves. We then plot four points, $(\lambda_{min}, m_\alpha \pm 2\sigma_\alpha)$ and $(\lambda_{max}, m_\alpha \pm 2\sigma_\alpha)$, where λ_{min} and λ_{max} are the smallest and largest λs of points in the subset. We then remove the point corresponding to λ_{min} from S, add the next point with λ larger than λ_{max} to S, and repeat the procedure.

The two envelopes defined by the four lines that emerge define the 95% confidence intervals for our best fits. Regions that are excluded by both envelopes are certainly excluded at 95% confidence. Regions that are included by both are included at 95% confidence. Regions excluded by only one envelope are indeterminate; we've roughly quantified the uncertainty on our confidence interval. The volume of the indeterminate region can be made arbitrarily small by computing greater numbers of best-fit points. In regions where α is a strong function of λ, convergence is necessarily slow. The final confidence-interval plots can be found in Section 12.2.1.

5.10 "Force law" inversion / "Arbitrary Fit"

In addition to extracting model parameters from the data, we can reconstruct the function $\tau(d)$, the pendulum torque as a function of attractor/pendulum separation. The entire data set consists of differential measurements of $\tau(d)$, where the experiment measures a $\Delta\tau = \tau(d_1) - \tau(d_2)$. Recovering $\tau(d)$ is then similar to recovering a function using knowledge of its derivative. Just as in indefinite integration, there is a "$+C$", a free constant offset to which the experiment is insensitive.

The method used to recover $\tau(d)$ is simple – attempt to find a best-fit function which minimizes the difference between the measured torques $\Delta\tau_{meas}$ and those computed using the fitting function $\Delta\tau_{fit}$. To provide confidence intervals on the fit parameters, we bootstrap the measured data many times; the distribution of the extracted fit parameters provides the necessary foundation.

In order to generate results that are independent of any physics-model, we choose to fit a piecewise-linear function to the data. Each piece is 10 μm long for $d \in [160\ \mu\text{m}, 300\ \mu\text{m})$ and 50 μm long for $d \in [300\ \mu\text{m}, 900\ \mu\text{m}]$.

5.11 Blindness

The most accurate experiments are free of bias. It is possible to perform an unbiased experiment without a blind, but it is difficult to guarantee that all preconcieved notions of experimental results and implications are suppressed. Critics of blind experiments claim that, "no experiment is truly blind," in reference to the current reality that experimenters unblind their results privately prior to public dissemination of the result. This is true, but it hides the power of the blind – forcing complete and equal attention to each *known* systematic effect.

PlateWash is not an entirely blind experiment. It is a precursor to a blind experiment. More than half of the science data were taken without a blind, largely because we had not discovered a way to monitor the health of the experiment without access to data which included the central values of each measurement. During that time, it became clear that monitoring the distribution of the error-bars of each cut was sufficient to ensure nothing was going wrong within the instrument. Once that determination was made, we blinded ourselves to the central values of all torque measurements. All of the systematic investigations were performed blind. In this first iteration of PlateWash, our understanding of the instrument's metrology was in continuous refinement, which blocked any attempt to inject a fake signal into the data as they were acquired. Furthermore, the intentional physical separation between pendulum and attractor DAQ systems would make injecting a fake signal extremely difficult.

The data analysis was performed blind. Critically, the foil-systematic coupling and cuts were not determined until many months after the data had been acquired. Even if the experiment had been tuned during the commissioning of the instrument, our blind systematic tests and subsequent analysis have the power to expose several classes of tuning.

5.11.1 Blinding procedure

During the analysis, we blinded ourselves to the central values of all measurements by avoiding looking at the torque data directly. This is, of course, a challenge when operating a torsion balance, as the best measure of proper function of is the character of the torque data. Furthermore, important steps, such as calibration, must necessarily operate upon unblinded data. Therefore, we used a combination of data-analysis design and an honor code to minimize any chance of seeing the experimental results before the analysis was finalized.

To make it difficult to accidentally see any analyzed data, immediately following the lock-in step (Section 5.4.6), the central values of all torque-related data in each cut were set to zero. All testing of post lock-in data analysis was done by injecting synthetic signals and noise into those central values. This allowed us to ensure that the data-analysis system worked correctly without accidentally viewing blinded data. Unblinding

the experiment is as simple as disabling this blinding mechanism.

For the approximate systematic tests detailed in Chapter 9, we separately saved the unblinded results of the lock-in procedure. These measurements were then compared only as blind-differences, where measurements with an external parameter changed were subtracted from measurements with the external parameter nulled. This procedure allows us to determine the experimental susceptibility to external parameters without having any access to the central values of either measurement.

5.11.2 Blinding failures

All systems fail, and this fairly fragile blinding system was no exception. Early in the blind analysis, we viewed the unblinded outcome of the lock-in twice. The more-important occasion was during the development of the blinding/unblinding code. A bug in the blinding code led to one of us (Charlie) seeing a Yukawa fit that resolved a signal with $\lambda < 20$ μm, $|\alpha| \sim 10^{3\pm2}$, and with indeterminate sign. This predated our determination of systematic cuts, particularly those related to the foil; the indeterminacy of the sign of the observed signal is particularly important. We don't know whether our later determinations added to, subtracted from, or eliminated this signal.

This type of failure of the blind is a weakness to software-blinds in null measurements. Often, the best estimate of an instrument's proper operation is its output. In a software-blinded null measurement, the experimenter is forced to perpetually dance around any accidental observation of the result. A reliable hardware-implemented blind, such as a clock skew, a hard-coded unknown offset, or an unknown stable experimental parameter is much more resistant to innocent mistakes.

We look forward to using what we've learned on the next iteration of this experiment, as at least the known pitfalls of this blinding procedure can be avoided.

Chapter 6

Software

The software used to compute these results and programmatically generate this document is all freely available under the GNU Public License (GPL) or more-permissive licenses. This thesis and all supporting code necessary to go from the original raw data to every plot and datapoint expressed in this document are housed within a single git repository. Once the build environment is configured (with '`git pull`') and linked to the experimental data, this document can be generated simply by typing 'make'.

6.1 Software stack

The software used in this analysis may be unfamiliar to some physicists, so it is partially detailed here. There are many tutorials available across the web to sate any depth of interest.

6.1.1 Git

Git [114] is an industry-standard "source code management" utility designed to provide fast and easy tracking of changes. The entire development process is tracked, so changes can be followed and selectively reverted at any time. Furthermore, code can be maintained seamlessly in multiple branches, preventing speculative exploration from stepping on main-line production work. Git was designed from the start to be distributed. Anyone working on the project has a complete copy of the project source; it's effectively a distributed backup. This project is hosted on github, in addition to several machines at CENPA.

6.1.2 GNU Make

Make [115] is a program designed to specify the recipe by which a project is built. When used correctly, it forms a "directed acyclic graph" of the project's interdependencies and allows any part of the project to be rebuilt without requiring a complete rebuild. This thesis makes partial use of those utilities. The dependency handling is primarily used to ensure that dependencies are properly built before being used. Without a defined and programmatic way to go from raw data to final results, there's no way for anyone, including the original experimenters, to vet the process used to generate a final result. Make resolves this concern, making it possible to perform the entire analysis with a single auditable command.

In addition, Make natively handles parallel processing. To speed project builds, this thesis is normally built across multiple cores in parallel.

6.1.2.1 Sample

As a sample, here is the root makefile for the project.

Listing 6.1: Makefile

```
include Makefile.inc
CURRENTDIR := $(shell pwd)
PATHINJECT := tmp/pathinject
dissertation: gitlog.log debianPackages.tex
        $(if $(shell ls data), ,$(shell ln -s ~/PWData/ data))
        $(shell bin/countXXXs.sh thesis/thesis.lyx >> data/xxxCount.dat)
        $(shell bin/commitPlot.sh > data/gitCommitTimes.dat)
        $(shell sed -i  "s|HOMEDIR := .*|HOMEDIR := $(CURRENTDIR)/|" Makefile.inc)
        $(MAKE) $(PARALLEL) -C mlib
        $(MAKE) $(PARALLEL) -C NewtonianSimulation
        $(MAKE) $(PARALLEL) -C calibration
        $(MAKE) $(PARALLEL) -C autocollimator
        $(MAKE) $(PARALLEL) -C runAnalysis
        $(MAKE) $(PARALLEL) -C metrology
        $(MAKE) $(PARALLEL) -C ifo
        $(MAKE) $(PARALLEL) -C bootstrap
        $(MAKE) $(PARALLEL) -C conclusion
        $(MAKE) -C thesis
gitlog.log:
```

6.1.3 GNU Octave

Octave [108] is an open source clone of the commercial general-purpose scientific computing environment MATLAB. It implements the entire basic MATLAB command set, is comparably fast in many applications, and is free to anyone. The octave-forge repositories provide similar function to many of the established

MATLAB toolboxes. The readable source has been directly important to this project; when a library (`fir2.m`) upon which the PlateWash analysis depends changed and incorporated a bug, it only took a few hours to identify and solve the problem.

6.1.4 Gnuplot

Gnuplot [116] is a powerful plotting package. Almost every plot in this thesis was dynamically generated by Gnuplot. In addition to Gnuplot's plotting capabilities, it provides limited but convenient Levenburg-Marquardt fitting routines. For some non-critical fits, Gnuplot was used.

6.1.5 Assorted GNU/Unix tools

The benevolent computer engineers of yore have left us many gifts. This thesis relies heavily upon `bash`, `awk`, `sed`, `grep`, and many other tools of the standard UNIX command set. While these tools may appear cryptic or antiquated to a young physicist, they're powerful. All the problems smart people had in the 1970s are the same problems smart people have now. The smart people in the past have given us their tools.

Bash is the scripting language for a standard UNIX shell (and the name of the shell itself). Awk is designed for handling tabular arrays of data of all kinds. It is fast and flexible. Sed, the 'string editor', is a tool designed for rapidly traversing text files and operating on them; think of it as find-and-replace on steroids.

6.1.6 TEX, LATEX, LYX, BibTEX, JabRef

TEX is the finest document-processing software on the planet. It's also hard to use. This document is written using LYX, a graphical interface to LATEX. The bibliography is handled using BibTEX and JabRef.

Most numbers in this text are automatically generated from data after processing in Octave. Numbers intended for publication are saved, after accounting for significant digits and the peculiarities of the SI unit system, into text files. After a script wraps them in TEX's desired $ symbols, the file is included as a child document into the LYX/LATEX document. Voilà, data analysis flows directly into text!

6.2 Daily Builds

For the last year of this analysis, the data analysis and thesis were rebuilt from scratch hourly or daily. This enables quick catching of integration bugs and allows collaborators to contribute to any part of the work at any time. Furthermore, frequent builds provide a clear picture of the deliverable state of the project. For

a blinded experiment with an automated unblinding procedure, this is essential. A primitive bug-tracking system allowed quantitative tracking of progress.

6.3 Testing

Unit tests are used to verify that each individual function in a project delivers the expected output. If individual program units have failings, it is impossible for an assembly of those program units not to contain failures.

Testing is perhaps best explained by example. In our field, it is customary for everyone to regard the normalization of any FFT program as suspect until proven otherwise. One of the standard checks is to verify that the integrated real-space power and the integrated Fourier-space power are equal. People usually try a few inputs until they're approximately satisfied that the results behave as expected. What if that were programmatic? If, before you made serious use of a library, you could hit 'test', and the software could run a predefined set of such tests? It is *much* easier to check whether a test is written correctly than it is to verify that code is correctly-written. Testing is a requisite part of quality software development, and data analysis is software.

It is further possible to do "integration testing", where the outputs of functions of functions are checked for sanity. There is very little integration test coverage in this project. As a partial check, many of the plots rendered throughout this thesis will allow the reader to verify that the analysis is approximately correct.

Octave has a straightforward syntax for integrating unit tests into functions. Before the PlateWash analysis begins, the entire set of homespun Octave libraries is automatically tested. At present, this software project has insufficient formal testing coverage. This is improving, but our first results must be delivered before testing coverage is complete. As the testing coverage grows, the integrity of the project, too, will grow.

The code used in our Newtonian gravitational calculations (Chapter 7) has good testing coverage. That library's code is an excellent place to glimpse the possibilities unit-testing makes available to the discerning physicist.

6.4 Advantages

The advantages of having a dynamic ability to alter the data analysis at any point in the entire project and be certain that correctly-processed numbers, analysis, and plots will appear in the final text cannot be understated. This makes collaboration with others straightforward, as there's no fear that large portions of an

analysis will have to be "re-done" in response to changes early in the analysis process; it gets automatically "re-done" every night. In addition, this ability is critical to unblinding a sophisticated analysis; complex processes proven to work on simulated data can be applied directly to real data when the time comes. Finally, and of greatest importance, the entire process is replicable; anyone studying this experimental result can know exactly what was done.

Chapter 7

Newtonian Simulation

While the "infinite plate" geometry affords us many luxuries of alignment, where many corrections to gravitational imperfections in the basic setup are second-order, it is essential to have the ability to simulate the forces and torques between arbitrary mass configurations. Furthermore, the calculations must be sufficiently simple that they may be easily audited.

While it is less computationally-efficient, and can have numerical rounding concerns, we have chosen to simulate the Newtonian signals in PlateWash by breaking field and test masses into aggregations of many small point masses and computing the simple sums of forces and torques between them. Done this way, there is only the inverse square law to implement correctly. We make the unorthodox inclusion of code into the text precisely to demonstrate its simplicity. This is the *whole* library!

7.1 Data Format

The data structure used to describe a point mass in this framework is simple:

$$[m, x, y, z]$$

Collections of masses are simple aggregations of point masses:

$$\begin{bmatrix} m_1 & x_1 & y_1 & z_1 \\ m_2 & x_2 & y_2 & z_2 \\ m_3 & x_3 & y_3 & z_3 \end{bmatrix}$$

7.2 Forces and torques

The force $\vec{F}$ and torques $\vec{\tau}$ between mass collections M_1 and M_2 are computed as follows. The torques on M_1 are measured around $(0, 0, 0)$.

Listing 7.1: pointMatrixGravity.m (additional unit tests suppressed for brevity)

```
%Returns the three axis force and torque on array1 by array2
%array entries of the form [m, x, y, z]
function [force, torque]=pointMatrixGravity(array1,array2)

        force  = [0 0 0];

        torque = [0 0 0];

        for i = 1:rows(array1)

                        iforce  = Gmmr2Array( array1( i , : ) , array2 );

                        itorque = cross( array1( i , 2:4 ) , iforce' , 2 );

                        torque += itorque;
                        force  += iforce';
        end
end

%!test
%! 'pointMatrix force test'
%! fundamentalConstants
%! m1 = [ 1 0 0 0 ];
%! m2 = [ 1 1 0 0 ];
%! [ F T ] = pointMatrixGravity( m1 , m2 );
%! Fg = G;
%! assert( abs( F( 1 ) - Fg ) < 2 * eps )
%! assert( abs( F( 2:3 )    ) < 2 * eps )
%! assert( abs( T )           < 2 * eps )

%!test
%! 'pointMatrix torque test'
%! fundamentalConstants
%! m1 = [1 0 1 0];
%! m2 = [1 1 1 0];
%! [ F T ] = pointMatrixGravity( m1 , m2 );
%! Fg = G;
%! assert( abs( F( 1) - Fg ) < 2 * eps )
%! assert( abs( F( 2:3 )   ) < 2 * eps )
%! assert( abs( T( 1:2 )   ) < 2 * eps )
%! assert( abs( T( 3 )+ Fg ) < 2 * eps )
```

Where the inverse square law is evaluated by perhaps the simplest possible code:

Listing 7.2: Gmmr2Array.m

```
%Returns the three axis force on mass1 by mass2 (vector)
function force=Gmmr2Array(mass1, mass2)

        fundamentalConstants

        %Which way does the force act?
        rvec = mass2( : , 2:4 ) - ones( rows( mass2 ) , 1 ) * mass1( 2:4 );

        %Pythagorean theorem to determine |r|
        r = sqrt( sum( (rvec .* rvec) , 2 ) );

        %The inverse square law!
        force = rvec' * ( G * mass1( 1 ) * mass2( : , 1 ) ./ r .^ 3 );

end
```

7.3 Movement

It is convenient to programatically translate and rotate mass distributions. These functions allow the arbitrary assembly of simple components into more complicated ones. Furthermore, when patterning rotation- or translation-symmetric bodies, copies of the repeated element are trivially made and positioned.

7.3.1 Translation

Translation is defined by a three-vector. Translating is as simple as adding the desired three-vector to the spatial components of a mass distribution.

Listing 7.3: translatePMArray.m (some unit tests omitted for brevity)

```
%Translates a point mass array by transVec (a three vector)

function transArray = translatePMArray( array, transVec )

        array( : , 2:4 ) = array(:, 2:4) + ones( rows( array ) ,1) * transVec;

        transArray=array;
end

%!test
%! %The simplest possible test
%! o = translatePMArray( [1 0 0 0] , [1 1 1] );
%! assert(o == [1 1 1 1]);
```

7.3.2 Rotation

Rotation is more complicated. The rotation scheme used here is to define a direction in space, as viewed from the origin, then specify a rotation angle about that axis. Each sub-component point mass of a distribution is individually rotated through multiplication by a rotation matrix [117].

Listing 7.4: rotatePMArray.m (some unit tests omitted for brevity)

```
%Rotates a point mass array by angle radians counter-clockwise
%(as viewed from the origin) around rotVec

function rotArray = rotatePMArray( array, angle, rotVec )

        %Scale rotVec to unit length
        modulus = sqrt( rotVec * rotVec' );
        x = rotVec(1) / modulus; y = rotVec(2) / modulus; z = rotVec(3) / modulus;

        %Form rotation matrix, using:
        %## http://www.gamedev.net/reference/articles/article1199.asp
        c = cos(angle); s = sin(angle); t = 1.0 - c;

        R = ...
        [ t * x * x + c,        t * x * y + s * z,      t * x * z - s * y;...
          t * x * y - s * z,    t * y * y + c,          t * y * z + s * x;...
          t * x * z + s * y,    t * y * z - s * x,      t * z * z + c     ];

        % isolate and properly-transpose point-mass positions.
        points = array(:,2:4)';

        % copy array to preserve masses and allocate correctly
        rotArray = array;

        % rotate!
        rotArray(:,2:4) = ( R * points )';

end

%!test
%! % Simple check
%! m = [ 1 1 0 0 ];
%! o = rotatePMArray( m , pi/2.0 , [ 0 0 1 ] );
%! assert(  sum( o - [1 0 1 0]) < 4 * eps)
```

7.4 Shapes

It is convenient to define a small library of basic shapes from which larger objects may be created. In combination with translation and rotation, a few shapes can be used to represent almost anything. The simple mass-distribution representation also allows the easy construction of more ornate combinations.

Central to PlateWash design are flat plates. These plates come in two types, round and rectangular.

7.4.1 Right Annular Prism

It is convenient to develop a function for generating both round sheets and rings at the same time, as a round sheet is a ring with zero inner radius. This function places points on a Cartesian grid, with points accepted if they're internal to the specified mass. Grid spacing along the in-plane directions and axial direction may be independently selected.

Listing 7.5: genPointMassAnnlSheet.m

```
function array= genPointMassAnnlSheet(mass, iRadius, oRadius, thickness, ...
                                        nxpoints, nypoints)

        xgrid = thickness     / nxpoints;
        ygrid = oRadius * 2.0 / nypoints;
        zgrid = ygrid;
        nzpoints= nypoints;

        density = mass / ( pi * ( oRadius^2 - iRadius^2 ) * thickness );

        pointMass = density * xgrid * ygrid * zgrid;

        array = [];

        for i = 1:nxpoints
                for j = 1:nypoints
                        for k = 1:nzpoints
                                newMass=[ pointMass, ...
                                        ( i - (nxpoints + 1) / 2) * xgrid,...
                                        ( j - (nypoints + 1) / 2) * ygrid,...
                                        ( k - (nzpoints + 1) / 2) * zgrid];

                                pRadius = sqrt( newMass(3)^2 + newMass(4)^2 );

                                if(pRadius <= oRadius && pRadius >= iRadius)
                                        array = [ array ; newMass ];
                                end
                        end
                end
        end

        %Correcting any mass discrepancy
        MassDiscrepancyRatio = sum( array(:,1) ) / mass;
        array(:,1) = array(:,1) / MassDiscrepancyRatio;
        MassDiscrepancyRatio = sum( array(:,1) ) / mass;
end
```

7.4.2 Right Rectangular Prism

The other major shape used in PlateWash modeling is the right rectangular prism. All dimensions may be specified, and grid spacings along each dimension may be independently specified.

Listing 7.6: genPointMassRect.m

```
function array= genPointMassRect(mass, thickness, width, height, ...
                                        nxpoints,nypoints,nzpoints)

        xgrid   = thickness / nxpoints;
        ygrid   = width     / nypoints;
        zgrid   = height    / nzpoints;

        npoints = nxpoints * nypoints * nzpoints;

        if( mod( npoints , 1 ) != 0 )
                error('npoints IS NOT AN INTEGER!');
        end

        pointMass = mass / npoints;

        array = [];

        for i = 1:nxpoints
                for j = 1:nypoints
                        for k = 1:nzpoints
                                newMass=[ pointMass, ...
                                        ( i - (nxpoints+1) / 2) * xgrid,...
                                        ( j - (nypoints+1) / 2) * ygrid,...
                                        ( k - (nzpoints+1) / 2) * zgrid];

                                        array = [ array ; newMass ];
                        end
                end
        end

        if( rows( array ) != npoints )
                ## counting problem in XSheet
                array = 0;
        end
end
```

7.5 Display

When modeling a system, it is essential to have a visual cross-check on its configuration. Our matrix format lends itself to simple plotting of mass configurations. When mass configurations are plotted, many kinds of errors are trivially apparent. Plotting is as simple as

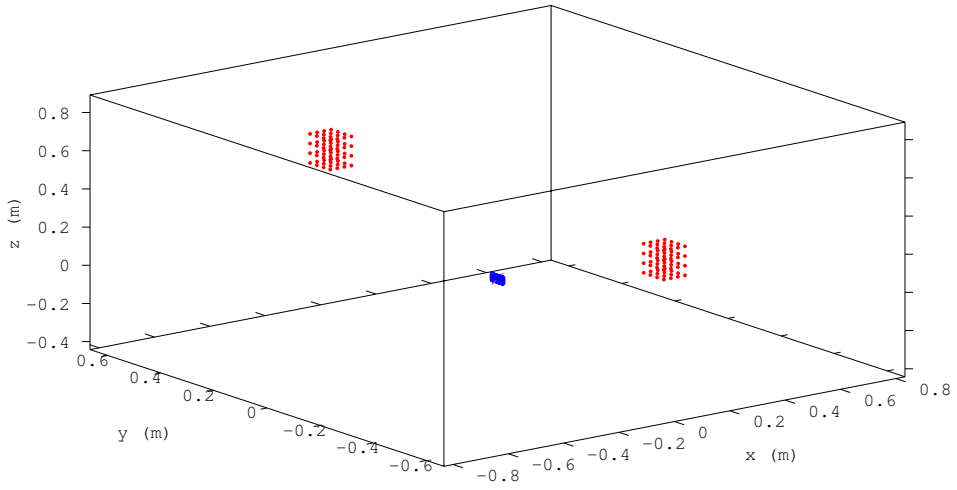

Figure 7.1: DisplayPoints output. Test mass in blue, field mass in red. Forces and torques are computed about the origin.

Listing 7.7: displayPoints.m

```
%Companion function to pointMassGravity
%Displays test mass (array1) in blue, field mass (array2) in red
function displayPoints(array1, array2)

        plot3(0,0,0,'x4', ...
                array1(:,2) , array1(:,3) , array1(:,4) , '.3' ,...
                array2(:,2) , array2(:,3) , array2(:,4) , '.1' );
        xlabel( 'x (m)' );
        ylabel( 'y (m)' );
        zlabel( 'z (m)' );
        axis( "equal" );

end
```

With these plots, an image like Figure 7.1 can be panned, zoomed, and rotated in three dimensions, giving easy intuitive access to the model.

7.6 Parts

For PlateWash-specific modeling, it is convenient to define a few assemblies.

7.6.1 Pendulum

The pendulum is assembled from three components, a rectangular-prism body, and two rectangular-prism inlays. The density of the inlays are chosen to yield the true inlay density when summed with the body's density. The dimensions are drawn from a configuration file. The pendulum body is modeled with 10 points through its thickness, 28 points across its width, and 15 points vertically. The inlays are modeled with two points through their thickness, and 15 points across their width and height.

7.6.2 Attractor

The attractor is rendered from three components. Two planar sheets to generate the effect of the cylindrical attractor, and one annular ring for the "compensation" rim.

7.6.2.1 Superposition

The planar sheets utilize a critical principle of superposition. In the region where the pendulum-attractor separation is comparable to the grid spacing, any point wise calculation of forces and torques is almost guaranteed to be erroneous or require careful attention. This difficulty can be bypassed entirely by superposition. The gravitational field from a cylindrical sheet can be computed either directly, or by computing the force from a large annulus that represents the "missing" mass from an infinite sheet that the cylindrical sheet purports to represent and subtracting it from the known-uniform force due to an infinite sheet. As all point-wise separations are large in the latter case, any residual effects of grid overlap are small indeed.

The attractor is composed of three parts: the dense "plate", a lower density "backer", and the over-compensating "rim". The dimensions are drawn, as with the pendulum, from a configuration file. The "plate" and "backer" are annuli of equal inner and outer radius, with 2 and 4 grid points through their thickness and as many as 70 points in a Cartesian grid laterally. As above, these are annuli with inner radius equal to the true attractor radius and are modeled to have negative mass. The positive-mass "rim" has 10 grid points axially and as many as 50 points laterally.

7.6.3 Spindles

The attractor is supported from the flexure by aluminum and Macor "spindles". The gravitational effect of the spindles should be very small, but they are incorporated into the simulation. Each spindle is simulated as a cylindrical column of aluminum with five axial grid points and five lateral grid points.

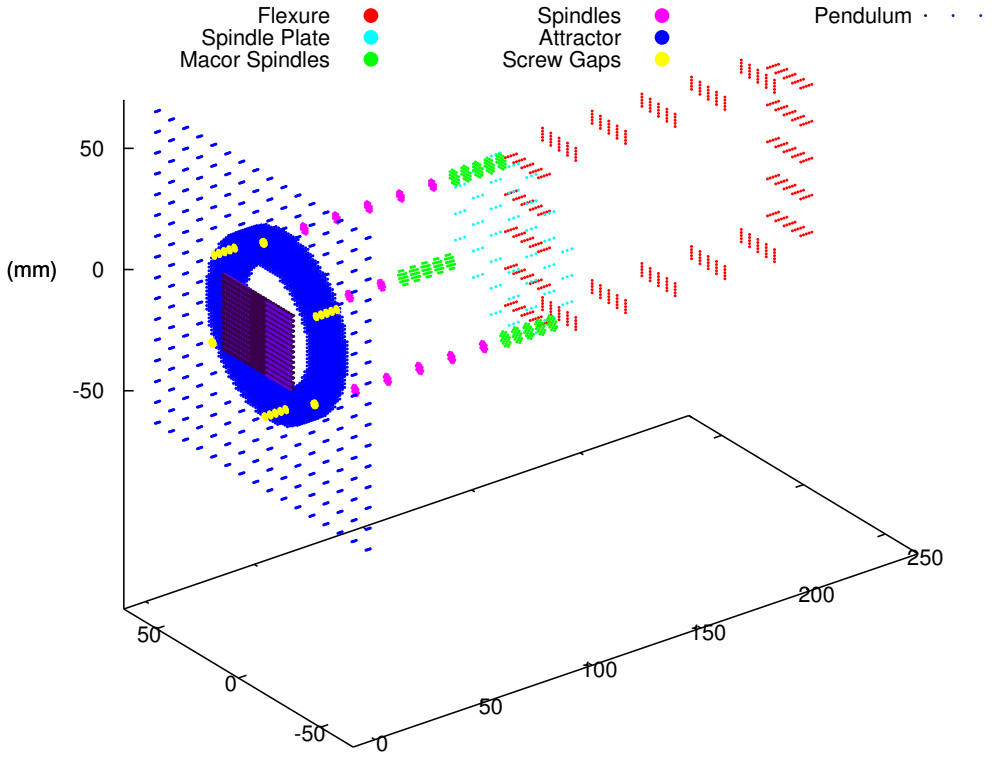

Figure 7.2: Complete Newtonian Assembly. The "infinite" sheet is cropped to make the other components visible. All non-pendulum components move together when the attractor moves.

7.6.4 Bellows

The Newtonian influence of gas flowing into and out of the bellows is easily simulated by computing the gravitational effect of a nearby cylinder of varying density. We model it as a cylinder with 7 axial grid points and 7 lateral grid points. The dimensional change of the bellows is quite small relative to the separation from the pendulum and is near the pendulum's center-line, so the mass-appearance should be the dominant gravitational effect.

7.6.5 Diagram

When all of the parts of the attractor and pendulum assemblies are plotted together, they appear as in Figures 7.2 and 7.3.

7.7 Verification

As discussed in Section 6.3 and described in code in the program listings above, the software libraries are tested with unit-tests before each build of the analysis. Some of these sanity-checks are described below.

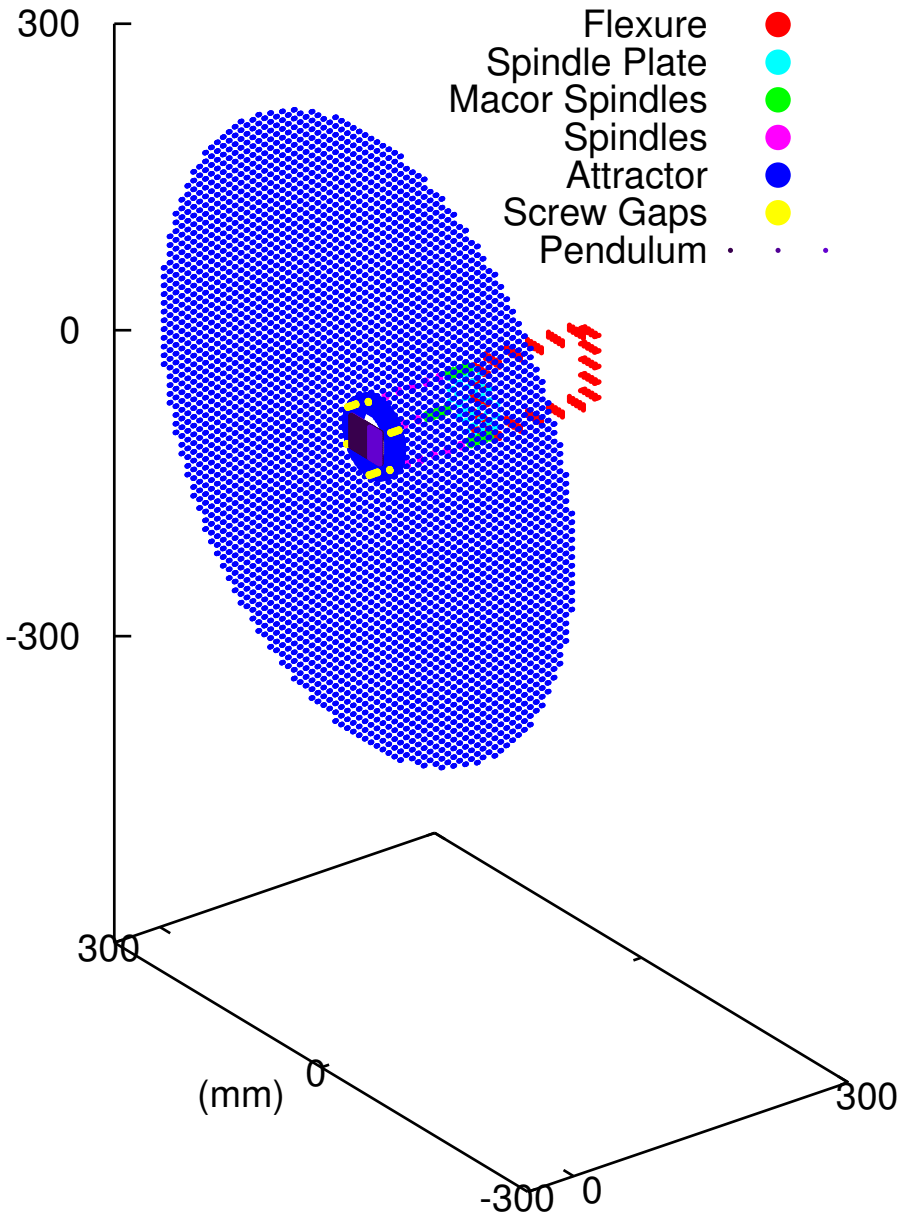

Figure 7.3: Complete Newtonian Assembly, zoomed out. The large thin disk is an effectively-infinite sheet. By cutting a hole in its center, and giving it "negative" mass, we recover the Newtonian gravitational force one should expect from our finite-sized attractor sheet.

Point-point forces make sense: A direct check of the force between two 1 kg point masses with 1 m separation has value G within numerical precision.

Point-point torques make sense: The torque about the origin on a point mass placed 1 m from the origin (at (0 , 1 , 0)) and acted upon by a point mass placed 1 m away along the direction perpendicular to the test mass's displacement from the origin (at (1 , 1 , 0)) has numerical value G, and the correct vector direction.

Forces go to $1/r^2$ at long range: Two cylindrical masses placed at 100 random separations large compared to their individual size follow the inverse square law at better than the 0.1% level.

Infinite Sheet The force on a point mass toward a 140 m radius, 10-cm thick cylindrical sheet is computed for displacements from $\sim$ 0.2-2.2 m away from the sheet. The grid spacing axially was 5 mm, and laterally was 500 mm To test for any grid-sensitivity, the sheet is randomly-translated in-plane with standard deviation of the grid spacing. At distances greater than 1 m, the force is within $< 1\%$ of the expected value.

Newton's shell theorem The force on a test particle placed with standard deviation of 1 m near the center of a randomly-generated 10 m-radius spherical shell with 100,000 component masses is always less than 1% of the force that would be present if the shell's mass were concentrated at a point 10 m away.

Simple quadrupole and point mass The torque between a simple quadrupole (two masses m spaced a distance d in opposite directions from the origin), acted upon by two masses M a distance R from the origin, placed at an angle θ is

$$\tau = 2G\frac{Mmd}{R^2}\sin(\theta)\left(\frac{1}{\left(1-2\frac{d}{R}\cos(\theta)+(d/R)^2\right)^{3/2}}-\frac{1}{\left(1+2\frac{d}{R}\cos(\theta)+(d/R)^2\right)^{3/2}}\right)$$

The numerical simulation reproduces this result at better than the part-per-billion level for R/d spanning 1.1 to 100.

7.7.1 Matt Turner's independent calculation

Prior to the development of this Newtonian modeling package, Matt Turner developed an independent package for computing the PlateWash torques using a semi-analytic numerical approach. His results were found to be in agreement with this simulation package for all pendulum/attractor combinations attempted.

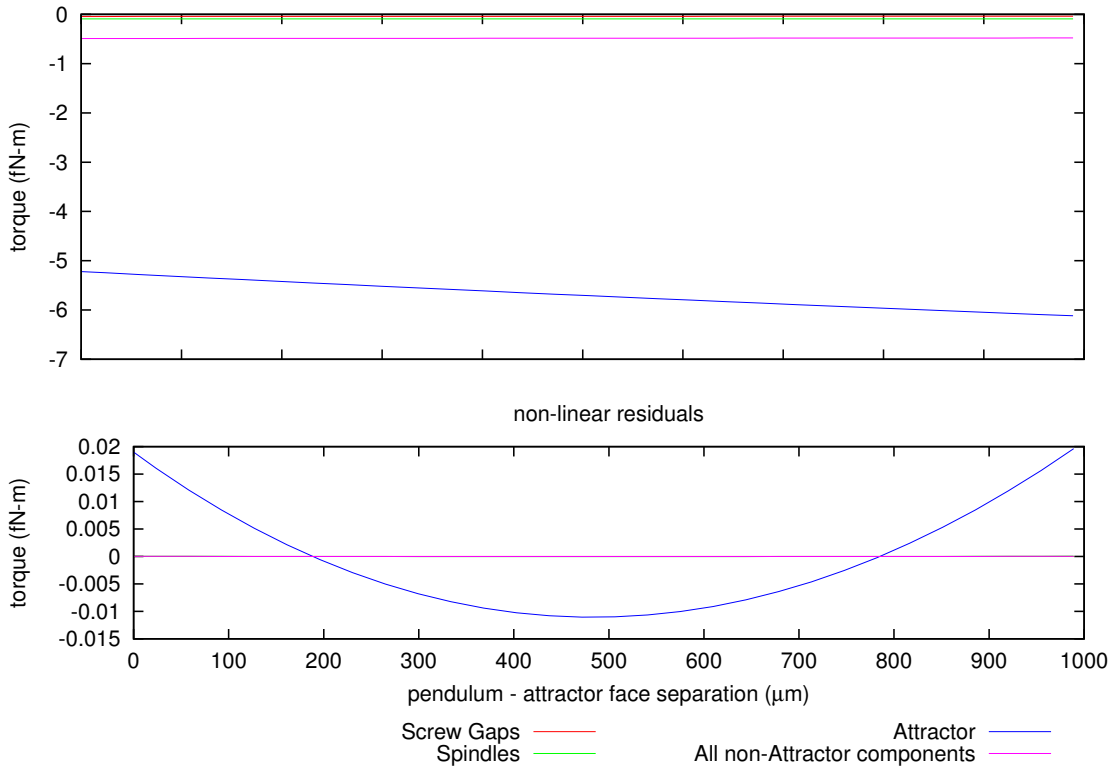

Figure 7.4: Simulation output for a centered pendulum. Non-linearities are below the 0.1 fN-m requirement, especially over the ~ 600 μm stroke range of the attractor.

7.8 Newtonian Simulations of PlateWash

In this section, we use the simulation package to set limits on the sizes of Newtonian systematics for our experiment.

7.8.1 Centered torques

First, we compute the expected torque and departure from linearity with the pendulum and attractor exactly aligned, as seen in Figure 7.4.

7.8.2 Off-center torques

The attractor is offset horizontally by (0.76 ± 0.25) mm (away from the window) and vertically by (0.50 ± 0.30) mm with respect to the pendulum. The torques due to the complete off-center attractor assembly are shown in Figure 7.5.

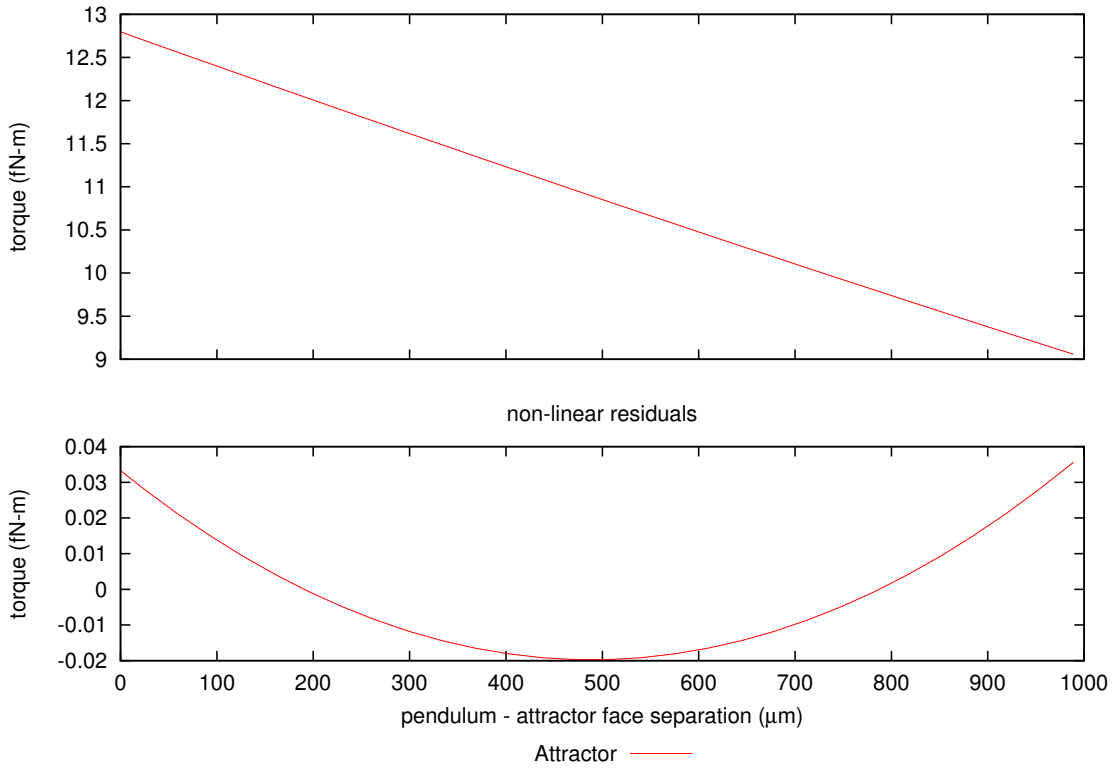

Figure 7.5: Simulation output for the offset pendulum

7.8.3 Torques from surface scans

The most challenging Newtonian task is to compute the effects of irregularities in the attractor and pendulum surfaces. The attractor and pendulum are slightly dome-shaped, as the hand lapping procedure tends to remove more material at the edges than at the center. Furthermore, in the case of the pendulum, the inlay and the pendulum body are made from two different materials, with different responses to lapping. Our SmartScope laser scanner allows us to scan these surfaces with better than 100 μm spatial resolution, and with perhaps a micron's vertical resolution. The maximum variation in surface height is less than 25 μm.

Any quadrature scheme that can resolve the measured surface variation will require

$$N = \frac{L}{g_h}\frac{W}{g_h}\pi\left(\frac{R}{g_h}\right)^2\left(\frac{T}{g_z}\right)^2 \approx 3300 \times \frac{1}{g_h^4 g_z^2}$$

individual calculations, where L and W are the pendulum dimensions, R is the attractor radius, T is the ($\sim$ 25 μm) depth of integration, and g_h and g_z are the horizontal and vertical grid sizes (in mm, for the formula above). In an optimal case, with $g_h = 10$ μm and $g_z = 2.5$ μm, a single calculation would require 5×10^{16} calculations. Assuming perfect optimization, each calculation requires more than seven floating-point operations (FLOPs), or 357 days on a modern quad-core desktop computer (21 hr on the 5

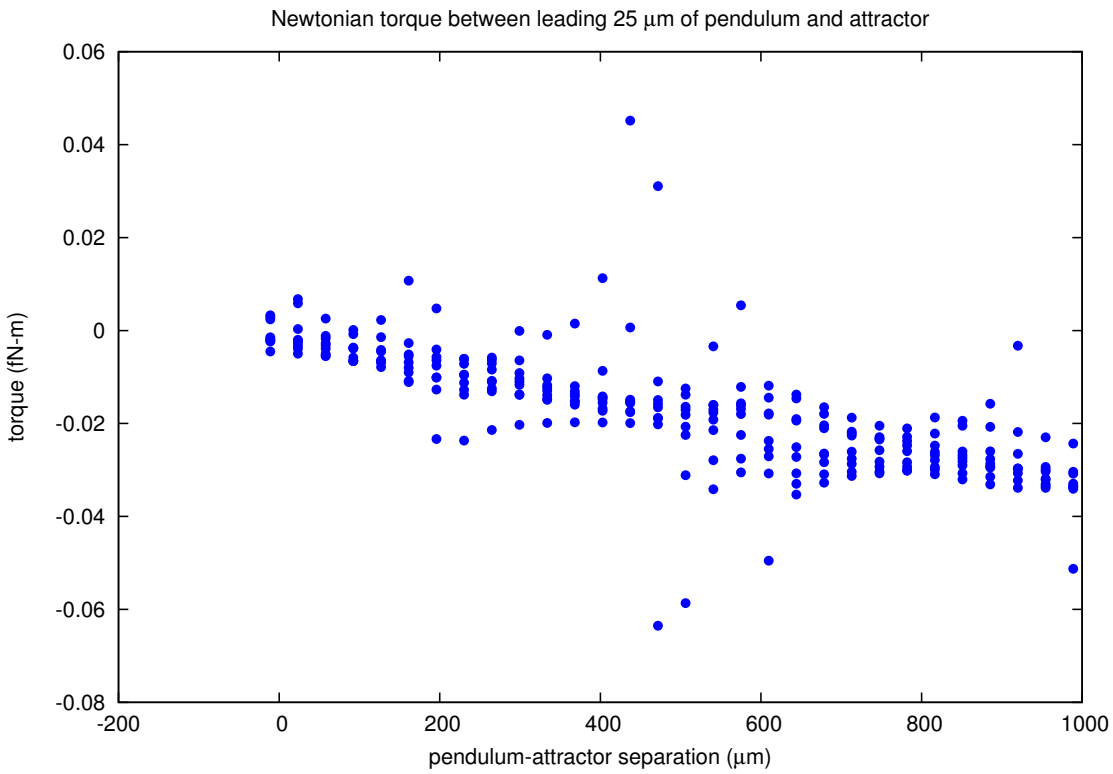

Figure 7.6: Torques due to the leading 25 μm of the attractor and pendulum, computed from the laser-microscope scans of the attractor and pendulum surfaces. Each line of computed points corresponds to an independent grid-scale random-positioning of local component masses (see text) and random rotation of the attractor with respect to the attractor axis.

teraflop CENPA cluster). Clever tricks can help to reduce the necessary volume of computation somewhat, but the detailed spatial resolution required sets the scale. As we would like to make these calculations many times, for perhaps 30 separate positions, and repeated once every hour or two (for hourly builds), we must find a way to tolerate much lower resolution.

We are saved by the fact that we only need the size of the non-flat torques and the associated uncertainties in size and linearity. If they are small compared to our other uncertainties, we're done. To generate speed, we use grid sizes of 2000 μm tangential and 5.0 μm perpendicular to the surface. Given that the pendulum-attractor gap can be as small as 80-90 μm, this is woefully inadequate. To provide an estimate of the effects of the too-large gridsize, the integration was repeated several times, with the point-mass subcomponents of the pendulum and attractor scattered horizontally with Gaussian standard deviation of one grid spacing, with results shown in Figure 7.6.

As we do not know the rotational alignment of the attractor scan, we randomly rotate the attractor scan about its symmetry axis before beginning each attractor-stroke of simulations.

Chapter 8

Narrative

There is insufficient space and time to recount the entire story associated with PlateWash. Several investigations germane to the final result are described in this chapter. A chronological history of the experiment may be found in the CENPA Annual Reports [118, 119, 120, 121, 122, 123, 124].

8.1 Interferometer

We spent a lot of time trying to find clever ways to execute a satisfactory parallel-plate test of the inverse square law without a measurement of the isolating foil's position. We were unable to find a way. For experiments of this type, where the attractor is modulated toward and away from the an isolating foil, estimates of the sensor/foil coupling and the foil motion itself are essential. This is the first measurement of which we are aware to make an on-line measurement of the isolating foil's displacement; an important step forward.

Fiber Fizeau interferometers are common in the atomic-force-microscope world [80, 81]. They are commonly implemented with infrared lasers. New to the field, we first built a visible system at 660 nm. Our Fabry-Perot diode laser suffered from mode-hops and drifts. Careful temperature control, current control, and current selection helped, but could not make the instrument usable for more than tens of minutes. A change to a distributed-feedback 1330 nm laser and the requisite complete exchange of optical fibers, couplers, and sensors yielded a stable instrument. In fact, the laser is so stable against temperature drift that we have been unable to use changes in the laser-head temperature to compensate for other drifts in the instrument.

Another essential ingredient was the micropositioning system (see Figures 2.28 and 2.15) we added to control the fiber-tip/foil separation. Without it, we spent hours attempting to position the fiber by

hand for proper fringe contrast. As the fiber-foil separation is on the order of tens of microns, it is a delicate operation. On one somber occasion, we punctured a foil with the fiber tip, requiring a new foil and complete realignment/cleaning of the experiment. The current system is simple and suffers the backlash and imprecision inherent in servo actuation, but it is sufficient to gain useful control over fringe contrast.

8.2 Encoder

Early in the experiment (2007-8), we uncovered an unexpected systematic effect: The pendulum torque signal was periodic in the attractor stroke, with periodicity of 25 μm. This matches the encoder grating spacing on the Micro-E optical encoder used to monitor the attractor position. We had operated for several months before noticing this effect; it is quite human when hunting for a signal with typical length ~ 50 μm to use strokes and starting positions in increments of 25 μm. It was only when we tried essentially-random stroke lengths that the prominent effect appeared. Powering-off the attractor encoder eliminated the spatially-periodic torque signal.

This effect drove us to make two major changes:

- Pneumatic attractor position metrology: This systematic effect made it clear that electrical pickup within the apparatus could become a concern. By using to the precision pressure gauge as a way of ascertaining the attractor's position, it was possible to do a complete end-to-end test for any electrical pickup associated with the attractor (Section 2.3.6). Once the pressure gauge was shown to be a usable encoder by intercomparison with the position encoder, the position encoder was turned off and disconnected for the duration of the experiment.

- Strong emphasis on random/Monte Carlo-like attractor positioning. For any given lengthscale and interaction type, there is, in general, an optimum square-wave attractor motion to hunt for it. Random positioning helps to elucidate any strange systematic effects, and allows an unbiased search to answer the question "What is the form of the gravitational force law at short distance" through the techniques of Section 5.10.

What we did not do, and should have, was to hunt down the way by which the encoder signal coupled to the experiment. The encoder operates at internal lock-in frequencies from 1-25 MHz, with delicate grounding requirements. We presume that the cross-coupling was through the autocollimator; exposed and unshielded leads extend for several inches within the vacuum chamber between the autocollimator's position-sensitive photodiode outputs and the high-gain transimpedance preamplifiers. This is an obvious target for future upgrades of the experiment (Section 13.2.1.4). It is straightforward to shorten the leads by an order of

magnitude and implement full electrical shielding for the photodiode and preamplifiers.

8.3 Capacitor Systematic

An important and subtle systematic effect concerns the grounding arrangement for the attractor, foil, and pendulum. A simplified grounding schematic is shown in Figure 8.1.

It was decided early in the experiment (∼2007-8) to ground the pendulum directly to the internal grounding block, which defines our experiment's electrical ground (Section 2.3.11). Furthermore, it was observed early on that the addition of 0.1-10 MΩ resistors to the grounding of the attractor and foil gave improved noise and killed off any inexplicable signals we saw. This was very early in our understanding of the instrument. It felt like fine-tuning, but both because we routinely operate torsion balances with 10-100 kΩ resistors in series with the torsion fiber as protection from electric transients and because the equilibration time of even a nanofarad capacitance across a megaohm resistor is $> 10^5$ times shorter than the signal period, we left the resistors in and turned to more-pressing concerns.

Later, the worry that we had fine-tuned the apparatus and the worry that at higher frequencies the foil and attractor were effectively electrically floating resurfaced, and we decided that it would be wise to bypass the resistors with large capacitors. We made that change and proceeded for years to refine the instrument.

In 2012, prior to our science campaign, the experiment was not blind, and an apparent inverse-square-law-violating signal was clearly resolved. The signal was too large to have been an ISL violation, as prior experiments would have trivially resolved it. At a suggestion from Krishna Venkateswara, the wires leading to the capacitors were clipped (at the same time we altered the lock-in frequency and, alas, separately disconnected the contact-potential control system), and the signal vanished, dropping into the noise resolvable overnight. In addition, the instrumental performance improved. Re-introducing the capacitors restored the signal, and shorting out the resistor with a wire also yielded a large signal. With no other reasons to believe that the experiment had a problem, we began the science run reported in this thesis shortly thereafter, with the capacitors removed.

Our systematic tests to show that the science signal appears to be largely independent of the choice of resistor value are described in Section 9.1.

8.3.1 Understanding the capacitor systematic

As hinted by Figure 8.1, our best understanding is that the movement of the attractor tunes an LCR resonance or resonances inside the experiment. The capacitance between the attractor and foil changes with attractor position, and the inductance of the wires connecting the elements together is sufficient to form

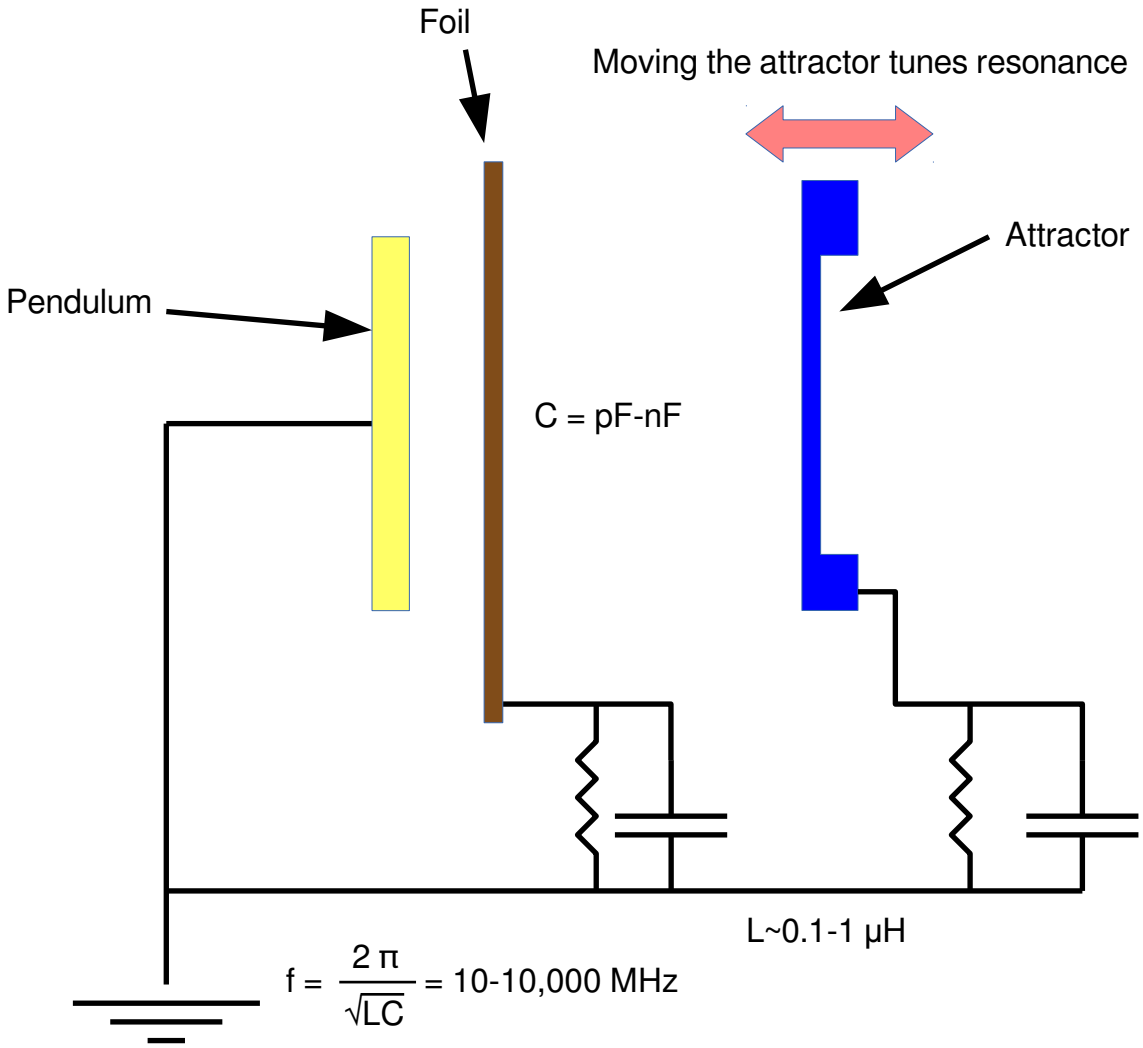

Figure 8.1: The Capacitor Systematic schematic

an LC resonator. Adding sufficient series resistance to spoil the resonator Q will squelch the effect, which we believe the resistor does. Bridging the resistor with a large capacitor provides an alternate path for high-frequency alternating currents, restoring the resonator's Q.

Identifying the resonance gives us a culprit, but not a mechanism. Our best guess as to how a tunable resonance might influence the torque signal follows similar reasoning as in Section 8.2; while the attractor encoder is left unpowered, an encoder monitoring the fiber position perpendicular to the foil is left powered and operating in order to retain datum information. This encoder may be radiating RF power in the $\sim$ 1-25 MHz realm; easily in range of the expected LC resonant frequencies. If the static encoder's RF signal couples to the autocollimator, and that signal is modulated by the presence of a new resonance in the experimental environment, a false apparent torque could emerge.

In addition, the presence of the foil allows a second possible, though less likely, coupling mechanism. Any mechanical resonance or displacement that could be driven by a resonance-enhanced effective attractor-foil or pendulum-foil voltage could also appear as a false torque.

Clipping the capacitors out of the system eliminates any hint of this systematic.

Chapter 9

Systematic Tests

This chapter describes individual systematic tests to give us an approximate picture of this measurement's systematic susceptibilities. As described in Section 5.8, the global impact of important systematics on physics results is determined within a single fit. These smaller analyses are more-interpretable and give insight into the outcome of individual systematic tests.

For most systematic tests, we operated the experiment with square-wave short-stroke and long-stroke attractor positions, seen easily in the right half of Figure 4.5. The short-stroke spanned the ~ 150 μm nearest the foil, and the long stroke spanned the ~ 550 μm farthest from the foil. The analyses in this chapter are blind to gravity; torque measurements are intercomparisons between two square wave measurements, suppressing any common-mode gravitational information. Blindness is preserved by subtracting torque data measured without any externally-applied influence from torque data measured with an externally-applied systematic influence. In the case of the magnetic field systematic, we also considered the difference between data taken with applied fields of opposite sign.

9.1 Resistor values

To assess the effect of changing the attractor/foil resistance, we tried a variety of resistor values. The long-stroke signal was measured (blinded by subtracting the long-stroke signal at 20 MΩ) for resistor values of 10^9, 10000, 50, 1, 0.1, and "0" ohms. The circuit is modeled as an LCR resonator, see Section 8.3. When the circuit Q reaches ~ 1, a major signal appears. This test shows, in Figure 9.1, that the specific choice of the 10 MΩ resistor cannot tune away nor create a signal.

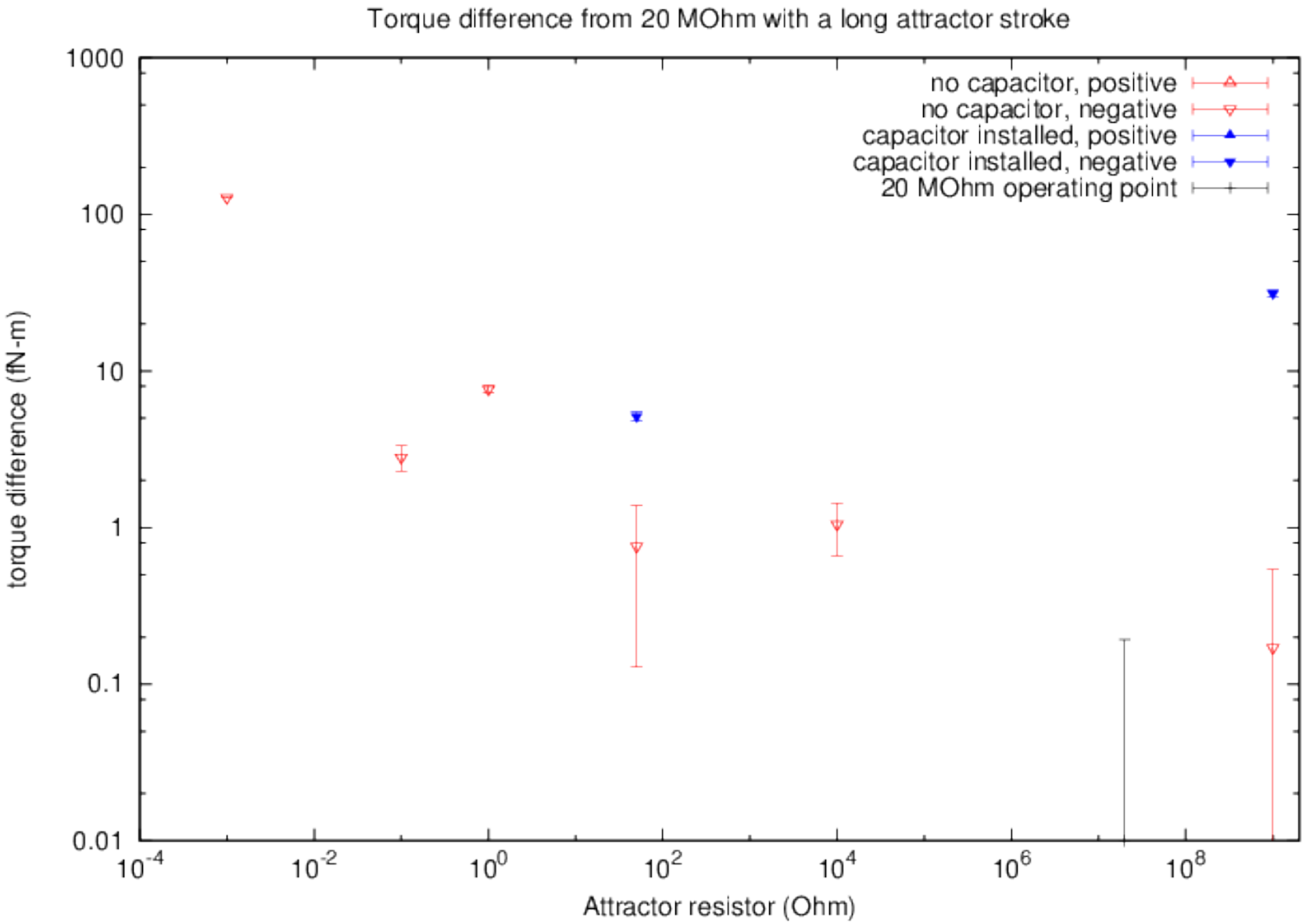

Figure 9.1: Summary of resistor tests to confirm that 20 MΩ was not finely-tuned. Note logarithmic axes.

Figure 9.2: Coil in the on-apparatus position. Current was applied to the coil in both directions, allowing searches for systematics proportional to both $\vec{B}$ and $|\vec{B}|^2$.

Coil Current (A)	Applied Vertical Field (T)	Stroke	Torque Difference
+0.8	500 μT	Long	(-0.39 ± 0.51) fN-m
+0.8	500 μT	Short	(0.90 ± 0.31) fN-m
-0.8	-500 μT	Short	(1.30 ± 0.30) fN-m
+ minus -	2×500 μT	Short	(-0.40 ± 0.41) fN-m

Table 9.1: Magnetic systematic test results. The "+ minus -" is useful as it's an independent blind difference. For scale, the ambient field strength outside the jar is 50 ± 10 μT, as measured by rotating an AK8975 smartphone magnetometer [125].

9.2 Magnetism

A (800.1 ± 6.3) mm diameter coil with 380 turns of wire was placed near and on the apparatus to gauge any change in signal as a function of applied magnetic field. The coil was driven with a constant current of (800 ± 100) mA. Histograms of the long-stroke and short-stroke tests are shown in Figures 9.3 and 9.4. The results are summarized in Table 9.1.

The change in field at the pendulum was estimated to be 500 μT vertical, 200 μT transverse, and 200 μT attractor-axial in the on-apparatus configuration.

9.3 Temperature

9.3.1 Temperature gradient

Using heating resistors glued to the attractor spindles, one leg of the attractor was heated with 20 mW ($(4.5 \text{ V})^2/(1 \text{ k}\Omega)$) of power, both heating the attractor and providing a spatial gradient. This was observed to create a $\sim$ 300 mK rise in attractor spindle temperature and create a 100 mK temperature gradient across

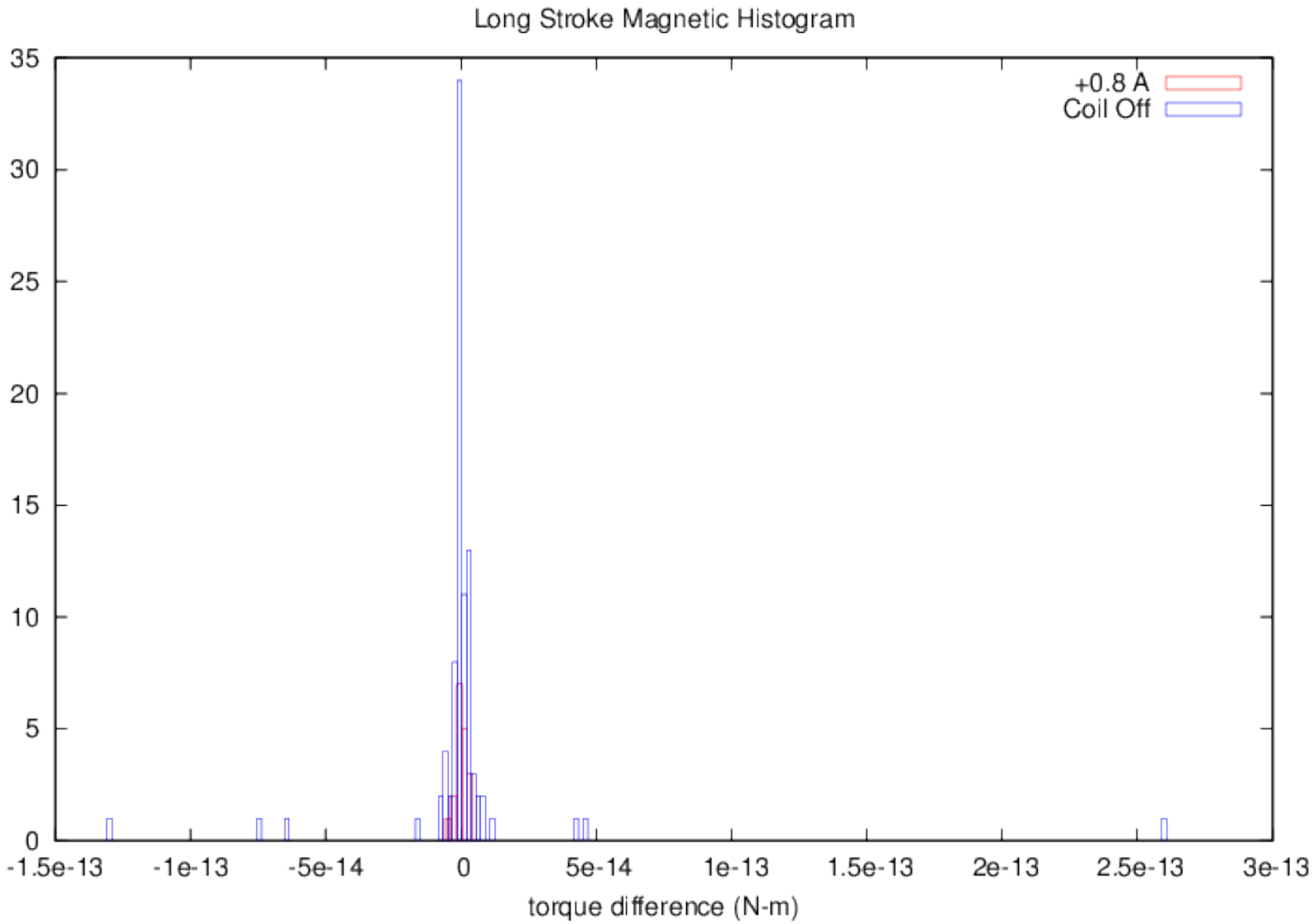

Figure 9.3: Long-stroke Magnetic Tests, a blind comparison of field-on and field-off torque differences.

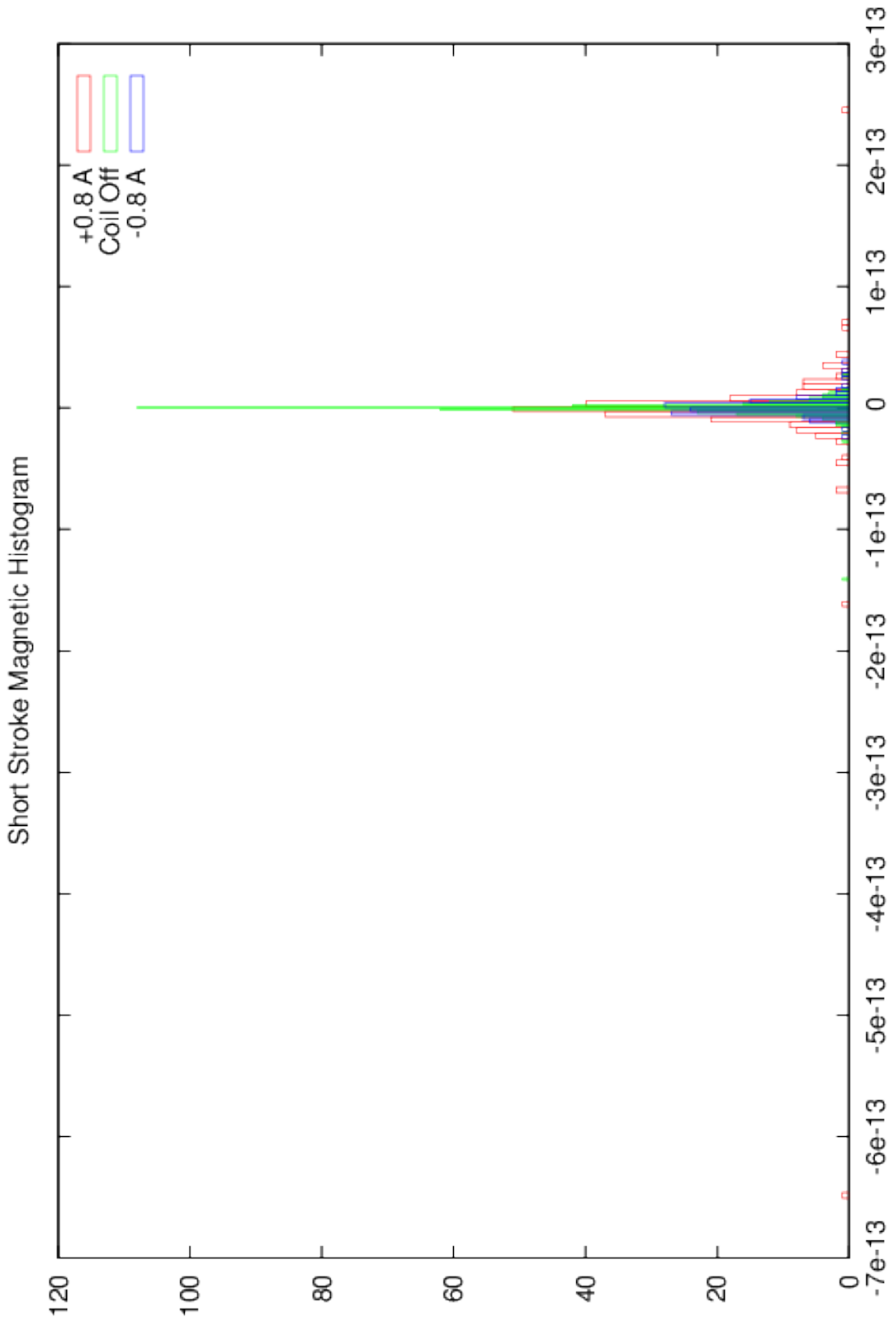

Figure 9.4: Short-stroke Magnetic Tests: A blind comparison of field-on, field-off, and field-flipped torque differences.

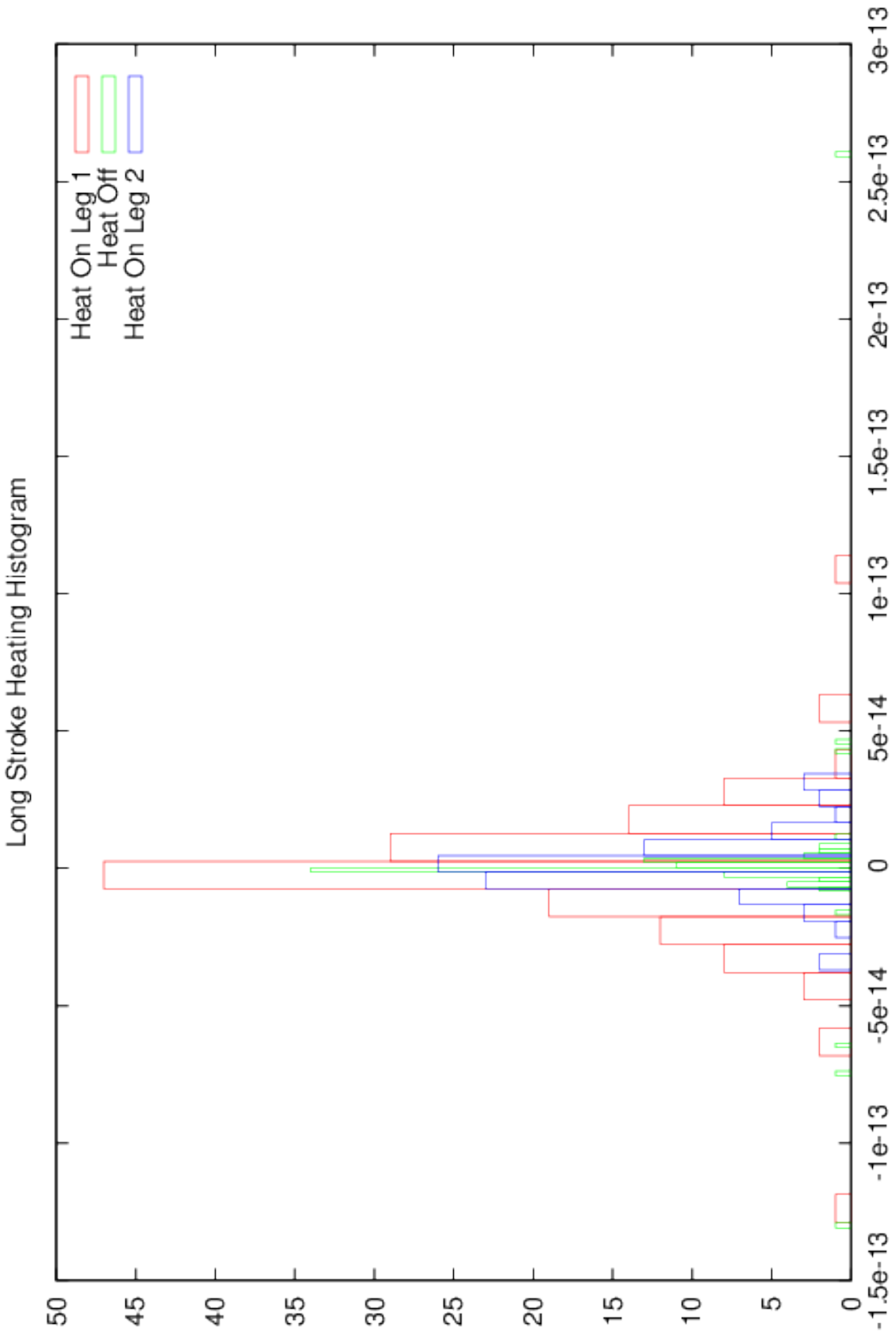

Figure 9.5: Long-stroke Thermal Tests: Blind comparison of torque differences between heat-off and with left and right attractor spindles heated.

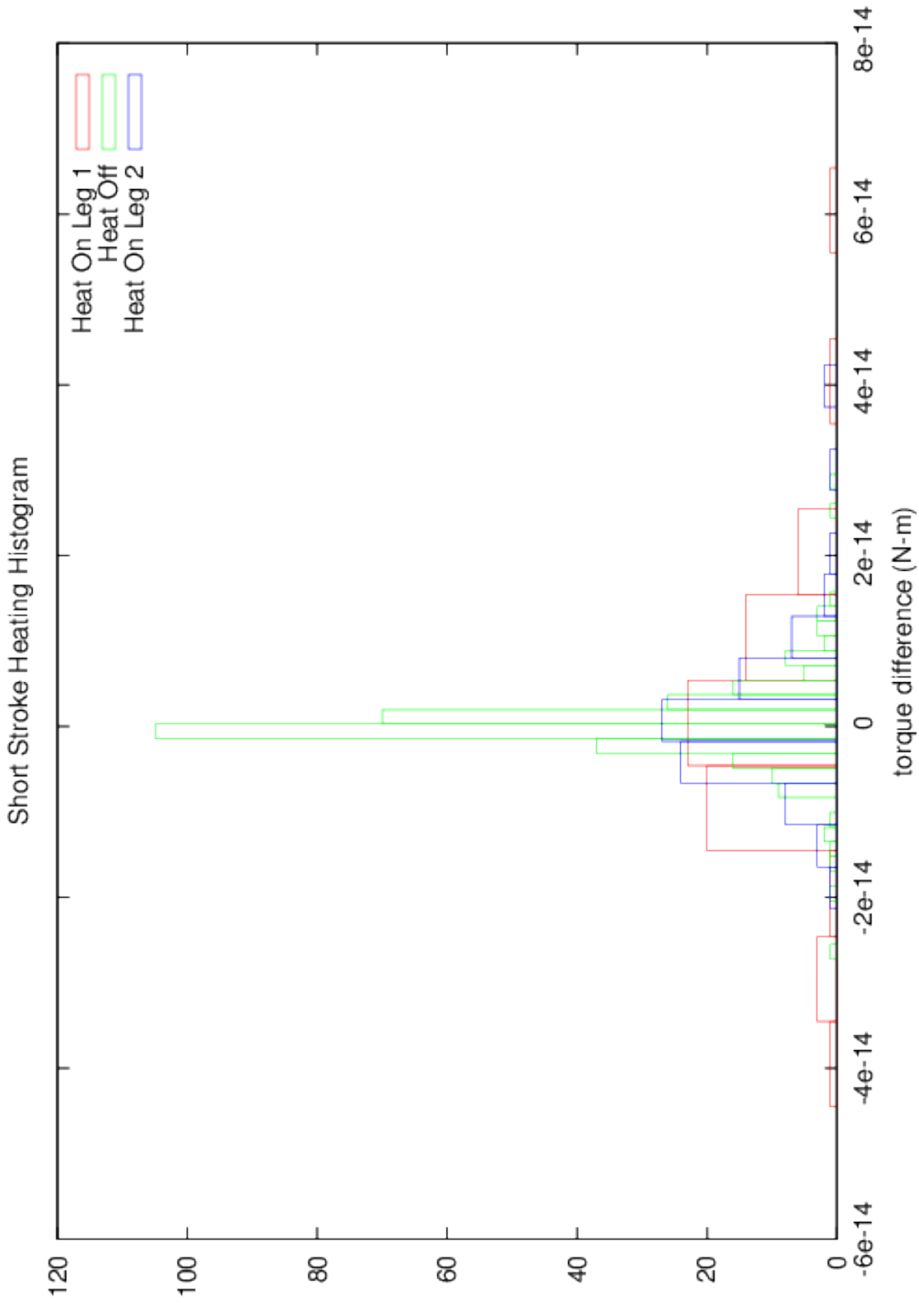

Figure 9.6: Short-stroke Thermal Tests: Blind comparison of torque differences between heat-off and with left and right attractor spindles heated.

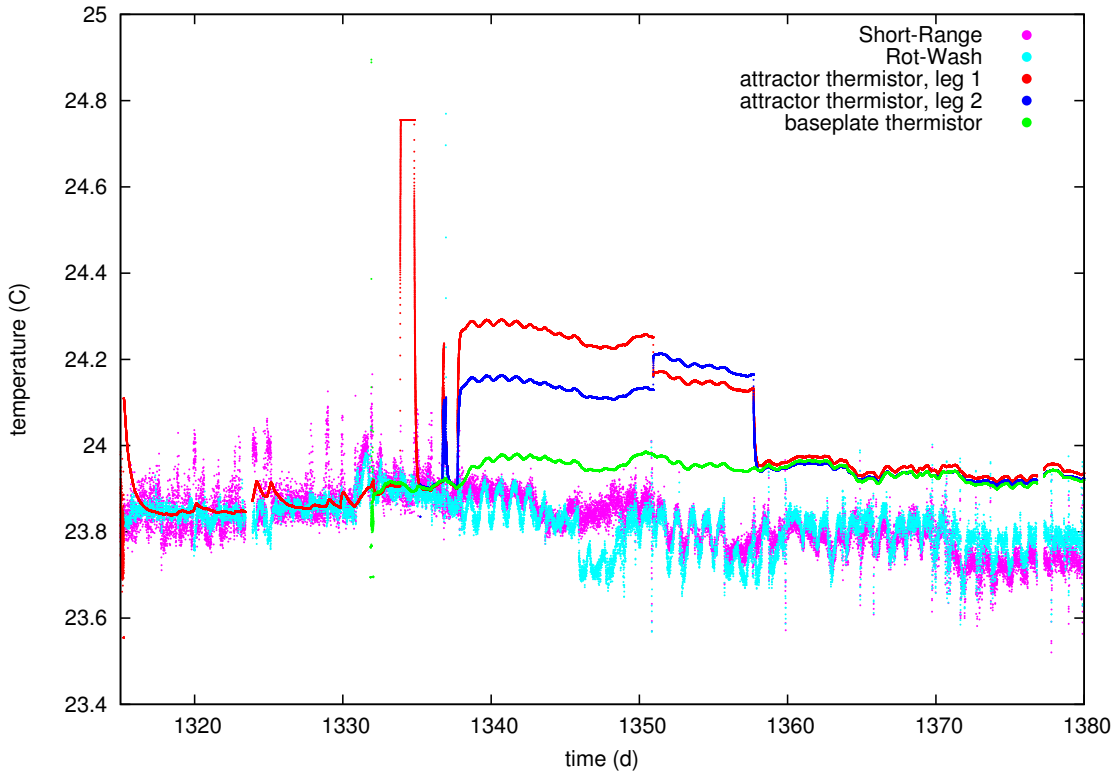

Figure 9.7: Temperatures during thermal testing. "Short-Range" and "Rot-Wash" are temperature sensors nearby in the room. The attractor leg thermistors are on the attractor spindles at left and right (but not top). The baseplate thermistor is screwed into the large aluminum baseplate internal to the apparatus. Spikes prior to day ∼1338 are from heater-verification tests. The large step (red above blue) is when the heater on leg 1 was turned on. The second step (blue above red) has the heater on leg 2 turned on, and the heater on leg 1 turned off. At the end of that sequence (∼day 1358), both heaters were turned off. The sensors are hand calibrated to match each other and the exterior sensors prior to day 1338. Reliability of this calibration is low, but sufficient, as the important scaling quantities are the ratios of the heated temperature and gradient to those measured without heating. The DC discrepancy between internal and external sensors after day 1360 is not understood, but no known instrumental effect accounts for it. It may correspond to a temperature rise within the PlateWash enclosure or to a common-mode ADC drift.

the attractor. The attractor was operated with long and short strokes and the run analyzed blind. For long strokes, the difference in torque-difference between heat-on and heat-off was (-5.55 ± 0.85) fN-m. For short strokes, the difference in torque-difference between heat-on and heat-off was (-0.3 ± 1.2) fN-m. Histograms of the tests are shown in Figures 9.5 and 9.6. The results are summarized in Table 9.2.

9.3.2 Temperature variation at attractor frequency

For runs beginning with ∼ #3217, the experiment was operated with wireless sensors reading out the resistance of the on-attractor thermistors. As the primary purpose of these sensors was to monitor the size of the temperature gradient applied by the heaters, we did not apply sufficient attention to the wireless-

Systematic Measurement	Measured Effect	Applied Stimulus	Ambient Stimulus
Long-stroke thermal	(-5.55 ± 0.85) fN-m	200 mK , Δ50 mK	<20 mK, Δ1 mK
Short-stroke thermal	(-0.3 ± 1.2) fN-m	200 mK, Δ50 mK	<20 mK, Δ1 mK

Table 9.2: Temperature-systematic measurements

thermistor-DAQ's time synchronization to allow attractor-phase-sensitive analysis. The time series of the temperature measurements is still usable for spectral analysis, presented in Figure 9.8.

The resolved signals in the Fourier spectra are commensurate with a long-stroke attractor temperature variation of ~ 30 μK of the spindles at the attractor drive frequency. By virtue of the greater heat capacity of the attractor and the impedance of the spindles, the attractor-face temperature variation should be smaller. Dimensionally, this temperature change is of no consequence; both the attractor/spindle expansion and the pressure change in the bellows driven by a microkelvin temperature change lead to ~ 30 pm displacements of the attractor. These displacements are far smaller than the intrinsic scatter in attractor positions and the slow retreat of the attractor due to leak-back.

If this stroke-dependent periodic variation were to lead to a systematic error, we should see it in an apparent "non-conservative" nature to the measurements, where $\Delta\tau(a,c) \not\approx \Delta\tau(a,b) + \Delta\tau(b,c)$. In this analysis, we do not explicitly search for this effect, but it is possible to do. A weaker signal-for or defense-against a non-conservative torque curve is the automatic intrinsic inflation of the distribution of the bootstrapped best-fit points in searches for new physics; biased sampling of short/long strokes may, however, lead to inaccurate fit results.

9.3.3 Interferometer spot-heating

The light from the foil-monitor interferometer carries energy, some of which may be absorbed by the foil. Less than 1 mW of light is incident upon the foil. At the 1.3 μm wavelength, copper is more than 95% reflective. Much of the light makes only one reflection. While only 4% of the reflected light re-couples into the fiber-optic core, much of it should couple back into the fiber cladding and dissipate elsewhere. Nonetheless, the few microwatts that must be dissipated through conduction in the foil may contribute to a slight foil temperature rise. To test for any related systematic effect, in runs prior to 3234, we operated the attractor with the interferometer laser turned off. Unfortunately, a record-keeping oversight prevented us from knowing during which prior runs the laser was disabled. As noted in our unblinding statement, this test should be done in any future experiment.

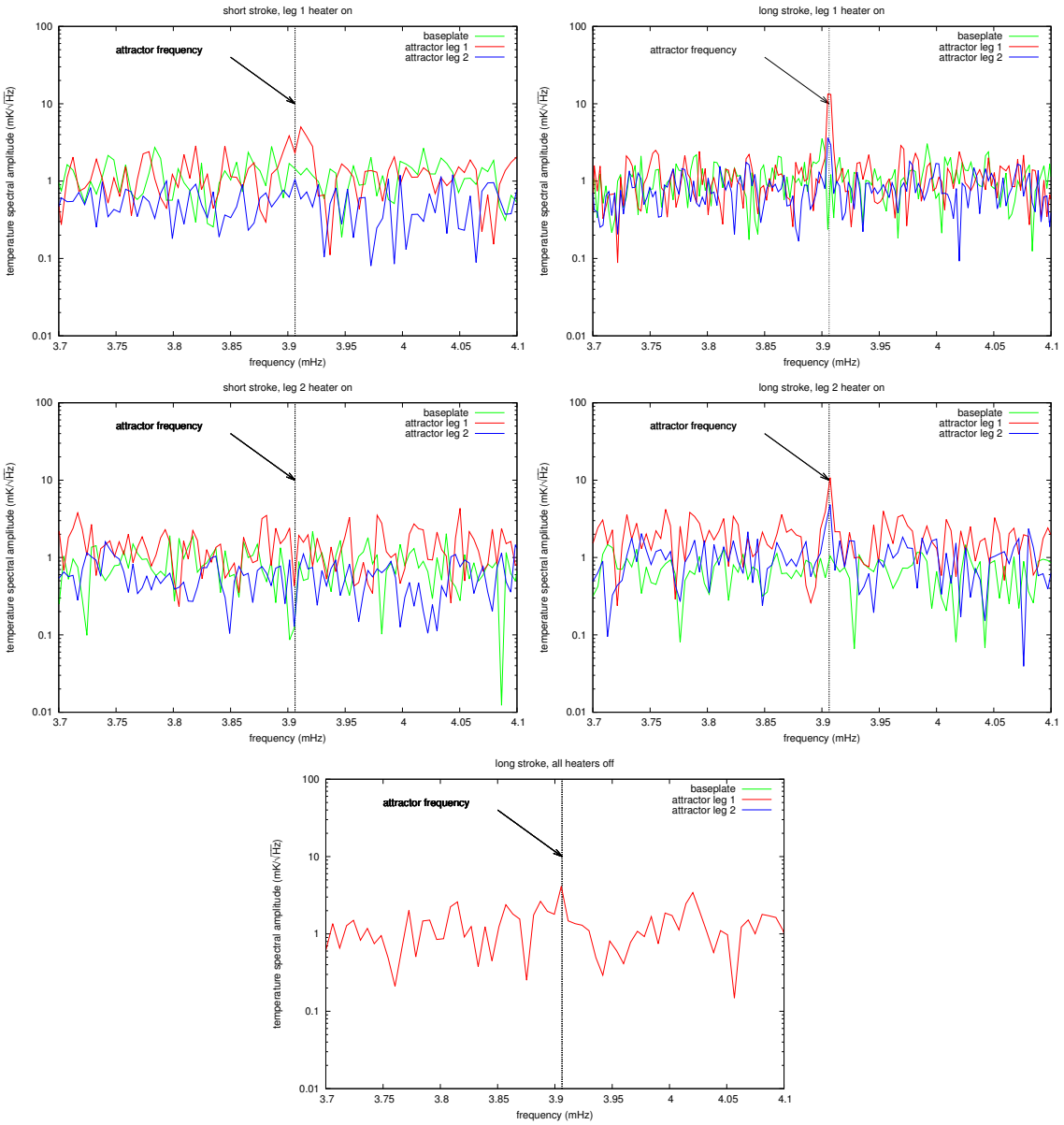

Figure 9.8: Temperature variation at attractor frequency. The top two rows have a heated attractor, the bottom plot is unheated.

9.4 Electrical Crosstalk

The experiment was operated with the pneumatic drive valved-off on the experiment-side of the pressure gauge in runs 3167 and 3168. Electrically, the experiment proceeded *exactly* as if the true experiment were underway, but the attractor did not move. After 168960 s the torque difference over a simulated long stroke was (0.0 ± 0.2) fN-m.

Chapter 10

Foil Systematic

The foil systematic measurement is discussed separately from the other systematic uncertainties for three reasons: the analysis is very different, the outcome is important, and, as a foil displacement is resolved, the effect must be handled with extra care.

10.1 Data/Fit

We used the foil measurements that were in-range as described in Section 2.3.9.1. After the lock-in/cut procedure (Section 5.4) is applied to the interferometer signal, the resulting measurements are plotted in Figure 10.1 as "triangle plots" (see Section 5.5).

As we do not have continuous full-sensitivity coverage of the entire measurement with the interferometer, we must build a model of the foil displacement as a function of attractor position and then apply it (with known uncertainties) via the foil-to-torque transfer function (Section 2.3.9.4) to generate the foil-related torque-difference expected for attractor-position pair. To be conservative, we use the piecewise-linear technique outlined in Section 5.10. To generate an estimate of the uncertainty in the fit, we bootstrap the observed data and generate many fits; the outcome is shown in Figure 10.2.

10.2 Systematic Subtraction and Uncertainty Propagation

Subtracting the foil-related torque is delicate. If we get the sign of the effect wrong, we double the systematic signal instead of subtracting it. For this reason, we will present our final results with the signal subtracted, with no subtraction, and with the signal added, so the importance of the effect will be apparent to the reader.

Furthermore, we must address the uncertainty in the subtraction. To do so, when bootstrapping the

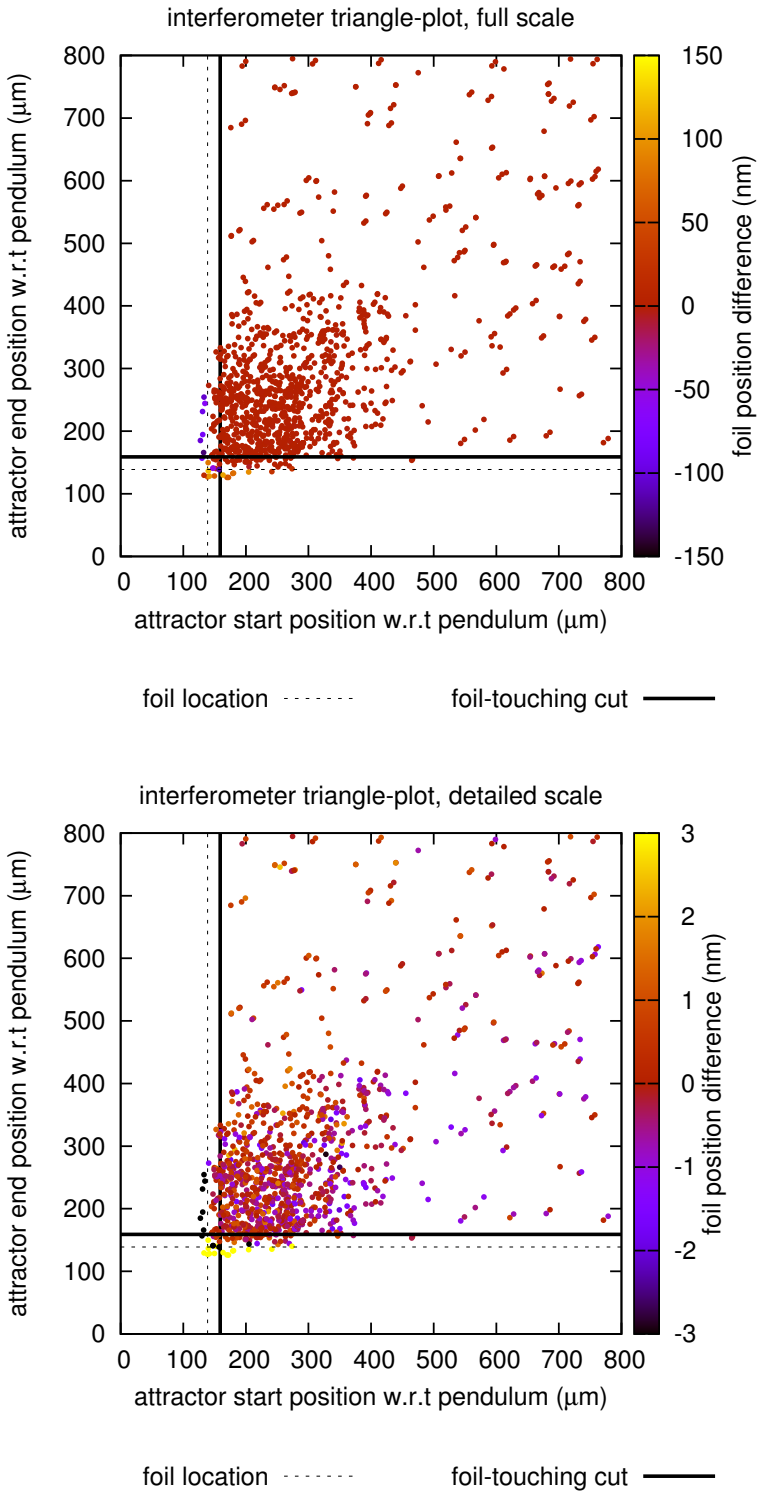

Figure 10.1: Raw foil displacement triangle plot. Upper plot shows the full scale motion, while the lower plot narrows the color scale to show foil motion in normal operation (attractor not in contact with the foil).

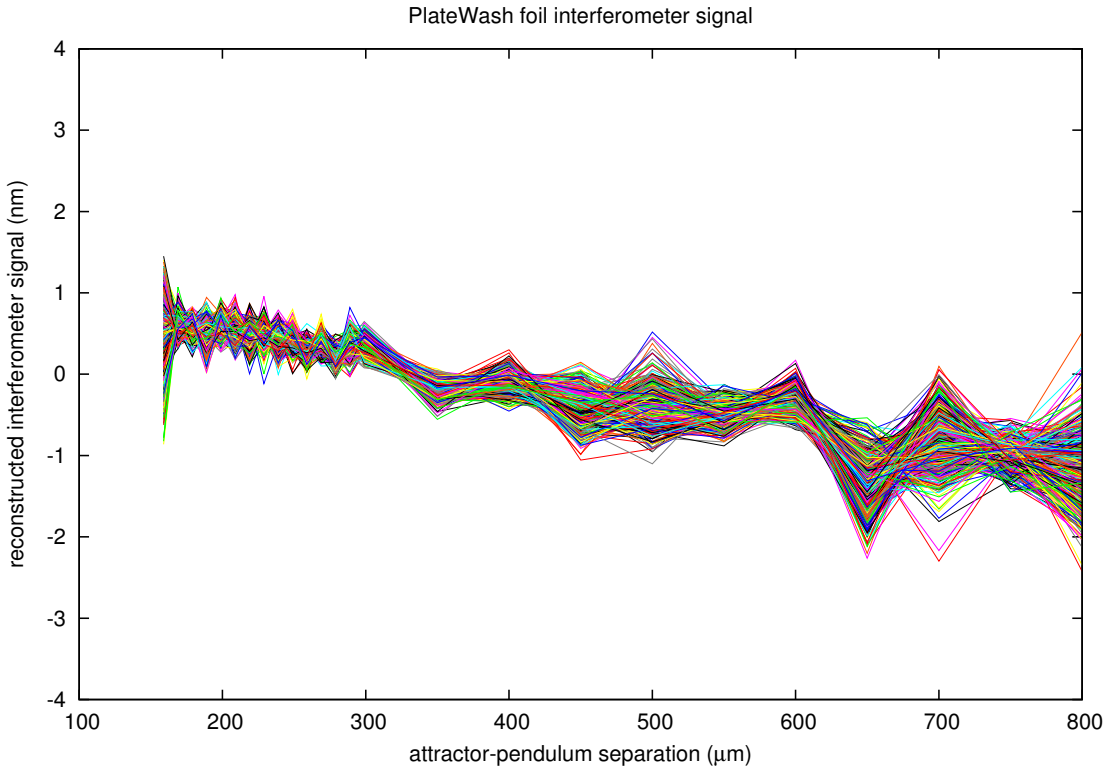

Figure 10.2: Interferometer fit. Each line is a different bootstrapped fit. This fits are used individually for subtraction in the bootstrapped global Yukawa fits. A reanalysis of the data with a more conservative cut may allow us to set tighter limits in the future.

model fit, with each bootstrap, we subtract a different one of the bootstrapped fits from 10.1, thus the scatter in the bootstrapped fits propagates directly into the scatter of our model best-fits.

Chapter 11

Blind Analysis

11.1 Blind commit

We declare that our measurement is unblinded with git commit 3ab13338192cc79cf82697bbcbfeefc29e364be8. Furthermore, the hash of the data directory, as evaluated by `cat $(echo run*.dta run*.hdr run*.dat | sort) | sha1sum` is c68496dd3fefcb5d9c5af912701409ad85695451.

11.2 Unblinding statement

In April of 2015, David Newell asked, "So, what are you going to do after you unblind? Are there extra tests you're going to do? What if you see something? If you do something different if you get a null result than if you see a signal, doesn't that partially defeat the point of having a blind?" Discussion and introspection that followed led to the inclusion of this essential section of the thesis.

Every blind measurement should have a statement written prior to unblinding that explains what the experimenters think the measurement means. Unaddressed sources of systematic uncertainty (i.e. if we see something unexpected, what might it be?) must be enumerated. Possible experiments needed to constrain them should be suggested. Any post-unblinding experimental tests must be enumerated and, in most cases, invariant under whether or not a signal is seen.

The "Unblinding Statement" is a qualitative companion to an error budget. If there are few residual concerns, it can be short.

11.2.1 Unaddressed systematic uncertainties

11.2.1.1 The interferometer

As we moved to unblind, careful consideration of the sign of the reconstructed foil-displacement signal led us to realize that the reconstruction algorithm is more susceptible to the effect of fringe turning points than originally believed. Without understanding the sign-flips at the fringe turning points, there is a danger that the measured foil movement could be underestimated by an aggregate measurement. This statement is being written with about a week to go until unblinding, as we must allow for computation time to analyze the data. More analysis is on the way, and will be presented at the unblinding and in future pre-unblinding commits. Periodic absolute calibration of the sign of the interferometer signal, perhaps by making tiny displacements of the interferometer tip in a known direction, would resolve this ambiguity entirely. *A pre-unblinding note: Subsequent analysis since the beginning of unblinding, and before the result is known, suggests that the degree of suppression should be relatively small. Determining the absolute sign of the foil's displacement in time for the unblinding will elude us. Thus, we will take our limits considering both sides of the interferometer signal. If a resolved signal appears when both the interferometer signal is added and subtracted, it is probably not due to a foil-motion effect.*

11.2.1.2 Pendulum-foil attractor-correlated contact potential variation

As referenced above, a small variation in pendulum-foil contact potential could yield an inverse-square-law violating signal, should the change somehow contain information about the attractor-foil spacing. If it does not have access to spacing information, it would tend to be a smooth, perhaps linear, function of attractor displacement. It is possible to hunt for such an effect by biasing the attractor and foil together with respect to the pendulum, increasing the influence of the small δV voltage as $(V + \delta V)^2 \approx V^2 + V\delta V$. It's a tricky measurement to make, as the pendulum-foil coupling will increase like V^2, but it's not impossible.

11.2.1.3 Magnetism

There are a number of unaddressed small magnetic effects. While we have measured the magnetic fields of our pendulum and attractor to available sensor precision, and the active elements were acid-cleaned, it remains possible that there are magnetic impurities in our materials that could give rise to an inverse-square-law violating effect. These effects should be small, however, as ferromagnetic impurities must be fairly sizable and concentrated before they become important. Our group's rotary magnetic-sensing turntable would allow us to place tighter constraints upon any magnetic effect.

11.2.1.4 Metrology uncertainty (dust)

We have photographic evidence that the pendulum and foil alignment agreed with the "bounce" and "plateau" tests, limiting the size of any possible dust that might have compromised the more-quantitative measurements. Our attractor-foil separation and resolved interferometer signal may provide a hint of dust present in the attractor-foil gap. While we don't expect the dust to substantially affect our overall measurement of attractor-foil spacing (misalignment and dust would yield similar signals), if it did, it would affect only our claimed λ, not the presence or absence of an inverse square law violation.

11.2.1.5 Pendulum/Attractor deformation

It is possible that the pendulum or attractor has deformed while under vacuum, as did some of our earlier masses. If the pendulum inlays deformed slightly, it will only act to alter our metrology, not the presence or absence of a violation. If the attractor deformed, it is possible that the null-ness of the attractor could be compromised, in addition to altering metrology. Re-scanning these masses with the SmartScope will yield sufficient information to make this determination.

11.2.1.6 Interferometer heating

While we performed a test to check whether the interferometer's localized light might have heated the foil enough to open a thermal or outgassing mechanism of communication between the pendulum and attractor, a record-keeping lapse prevented us from making any quantitative statements. We know which run marks a return to laser-on operation, but we don't know with certainty which run began the laser-off period.

While we don't expect this to be a major concern, it's not impossible, and it is testable in a future iteration of the experiment.

11.2.1.7 Casimir

We don't believe that the Casimir effect can get through the isolating foil, but we can't prove it won't (see Section 2.4.2.9). The possibility of a Casimir subtlety should be entertained.

11.2.1.8 Timing (this point added after unblinding had begun, but before the result was known)

Presuming that there is at least a small linear background torque, as expected, it is possible to use the torque data to directly synchronize the attractor-computer and pendulum-computer clocks. If there is an error in our synchronization, correlation analysis may allow its detection and resolution.

11.2.2 If a signal is resolved

It might be new physics, it might be one of the above systematic effects, or it might be something we missed. While that sentence is tautological, it leads us forward. If the signal is new physics, it will survive future tests. It is possible to evaluate many of the above systematics readily, in some cases with external measurements, in others by repeating the measurement with upgrades. If the systematic is something inherent to our apparatus, repeated experiments by others in the field, or with completely different designs, like our group's wedge-pendulum apparatus, will help to improve everyone's understanding of Nature.

Any repetition of our experiment should be performed blind. Remaining blind has helped us so much in identifying previously unknown systematics and properly characterizing known ones.

11.2.3 If a signal is not resolved

Everything in the resolved-signal section above applies. If a measurement is important to the physics community, null or no, it should be rechecked and confirmed by others. The only reasons to stop hunting for both greater sensitivity and experimental certainty are if the level of measurement precision substantially exceeds theoretical prediction or if no major experimental improvements are feasible. In the case of short-range gravity, the former is unlikely to happen in our lifetime, and the latter is precluded by our short list of easy experimental upgrades. PlateWash is worth doing at least one more time.

Chapter 12

Results

Results shown in this chapter display the output of a run of the complete end-to-end analysis using 480 bootstrapped points, divided equally among interferometer-added, interferometer-subtracted, and no-interferometer-correction fits. A more thorough analysis, combining almost 5000 points, is described in Chapter 14.

12.1 Bootstrapped "Force Law"

As described in Section 5.10, we make piecewise linear fits to the data; as close to a model-free fit as one can reasonably get. Figure 12.1 is a reconstruction from thousands of differential measurements. The arbitrary fit does not include any fits for systematic effects. As plotted, there are 480 independent reconstructions, so the scatter does not reflect a traditional one-sigma error-bar. To demonstrate our ability to recover injected signals, we include Appendix B.1.

12.2 Model Constraints

12.2.1 Yukawa Potential

To fit for Yukawa interactions, we follow the recipe outlined in 5.8.1.

The upper two fields in the following figures will be most familiar to those versed in inverse-square law tests. Along the horizontal axis is the putative range of the interaction. Along the vertical axis is its strength relative to gravity. Each circular point in the plot is a best-fit point from the bootstrapped fit. For perspective relative to the best results in the field, the exclusion limits of Kapner and Cook [113, 16] are shown.

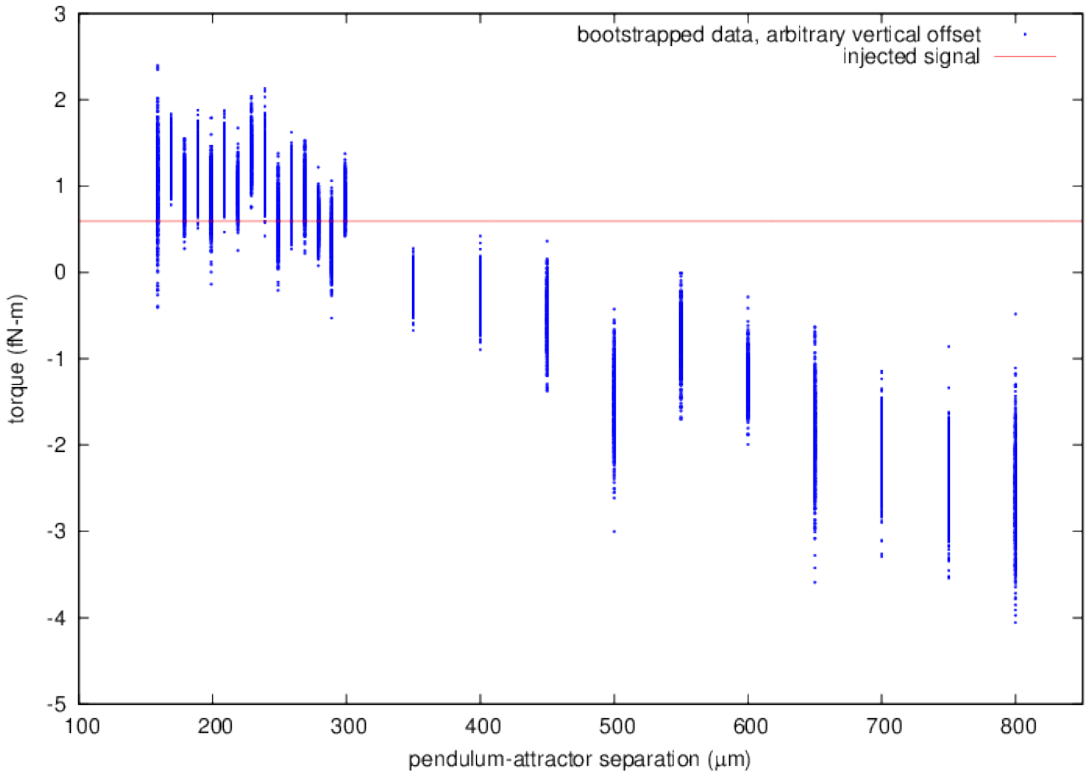

Figure 12.1: Bootstrapped arbitrary fit. This is a piecewise-linear reconstruction of the pendulum as a function of attractor position.

Next is a plot of the best fit points for the linear slope, C in equation 5.1. Different colors are used to show positive and negative slopes on the logarithmic plot.

Finally, we show the reduced χ^2 for each fit. In diagonal crosses, we show all fits. As past fitting algorithms had a tendency to be trapped in local minima immediately after initialization, we drop any points that converge in fewer than 5 iterations of the fit, which applies to essentially none of the simulated-annealing fits. Points surviving that cut are shown in horizontal/vertical crosses. Points surviving all cuts, including minimum and maximum chi-squared (horizontal lines) are shown as circles. χ^2 is of limited quantitative value, as in many torsion balance experiments, per-cut error-bars will tend to underestimate the uncertainties; we expect the relative uncertainties to be meaningful. Furthermore, as we jointly fit for systematic uncertainty, any reduced-χ^2 better than 1 may be due to the over-fitting inherent to the use of conservative systematic uncertainties.

We demonstrate the performance of our fitting algorithm against synthetic torques added to the data in Appendix B.2.

12.2.2 Other models

With these data, it is possible to place useful limits upon other inverse-square law modifications such as a power-law potential, Fat Gravitons, Chameleons, Unparticles, and more. For each distinct scenario, care should be taken to conservatively address systematic uncertainties. For this reason, we limit our results to Yukawa potentials.

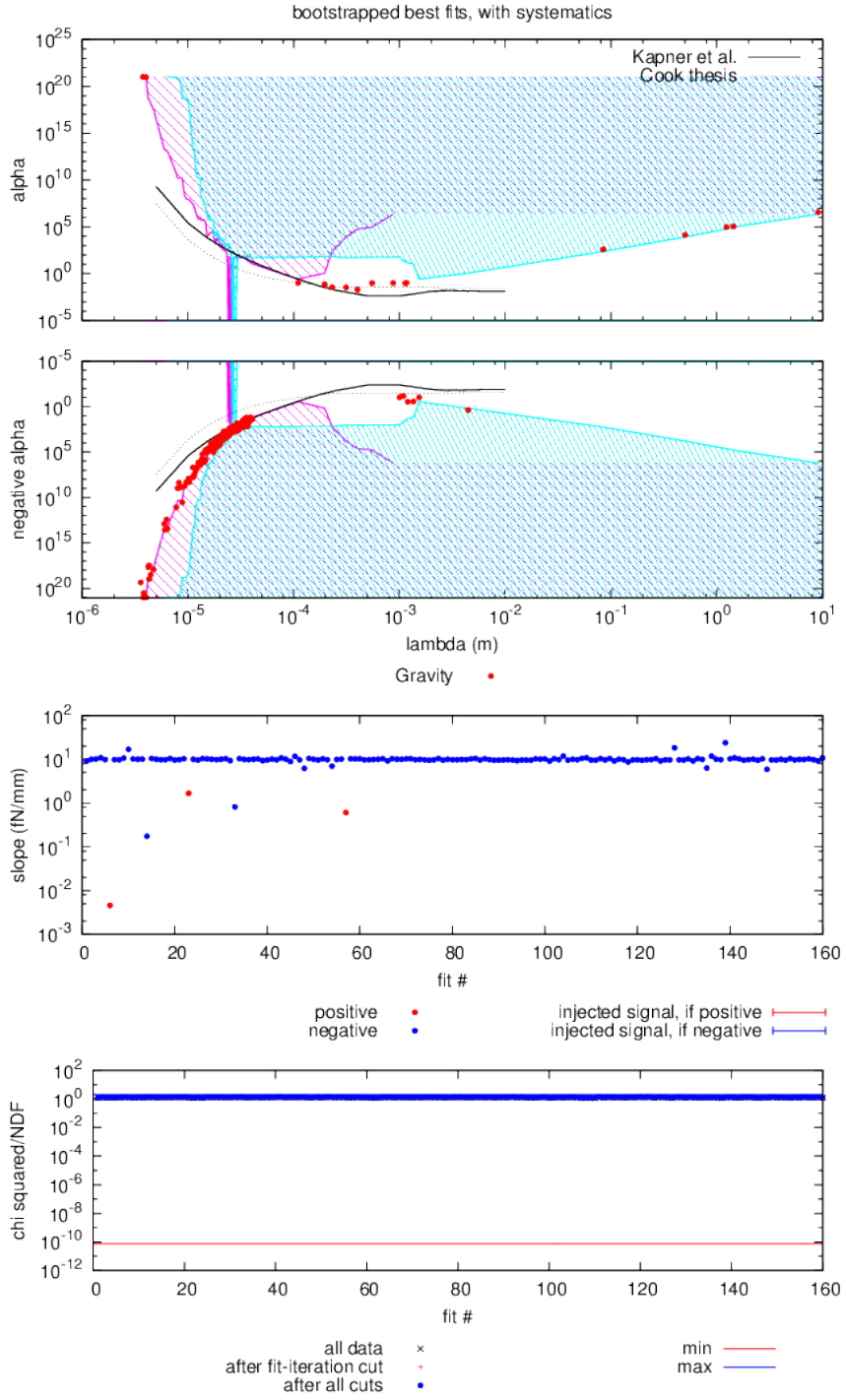

Figure 12.2: Yukawa exclusion limits for gravity alone, with foil systematic subtracted

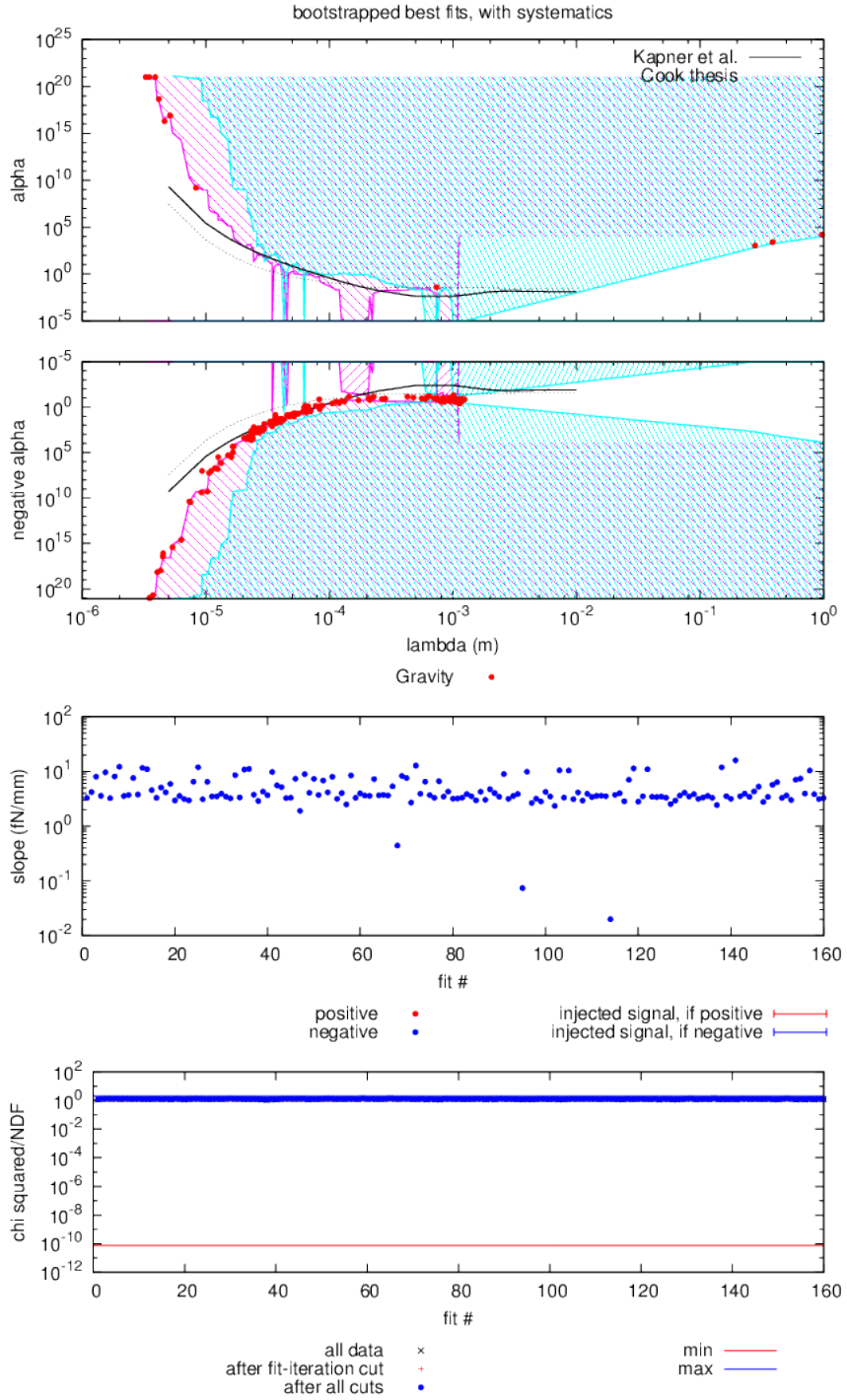

Figure 12.3: Yukawa exclusion limits for gravity alone without foil systematic subtraction

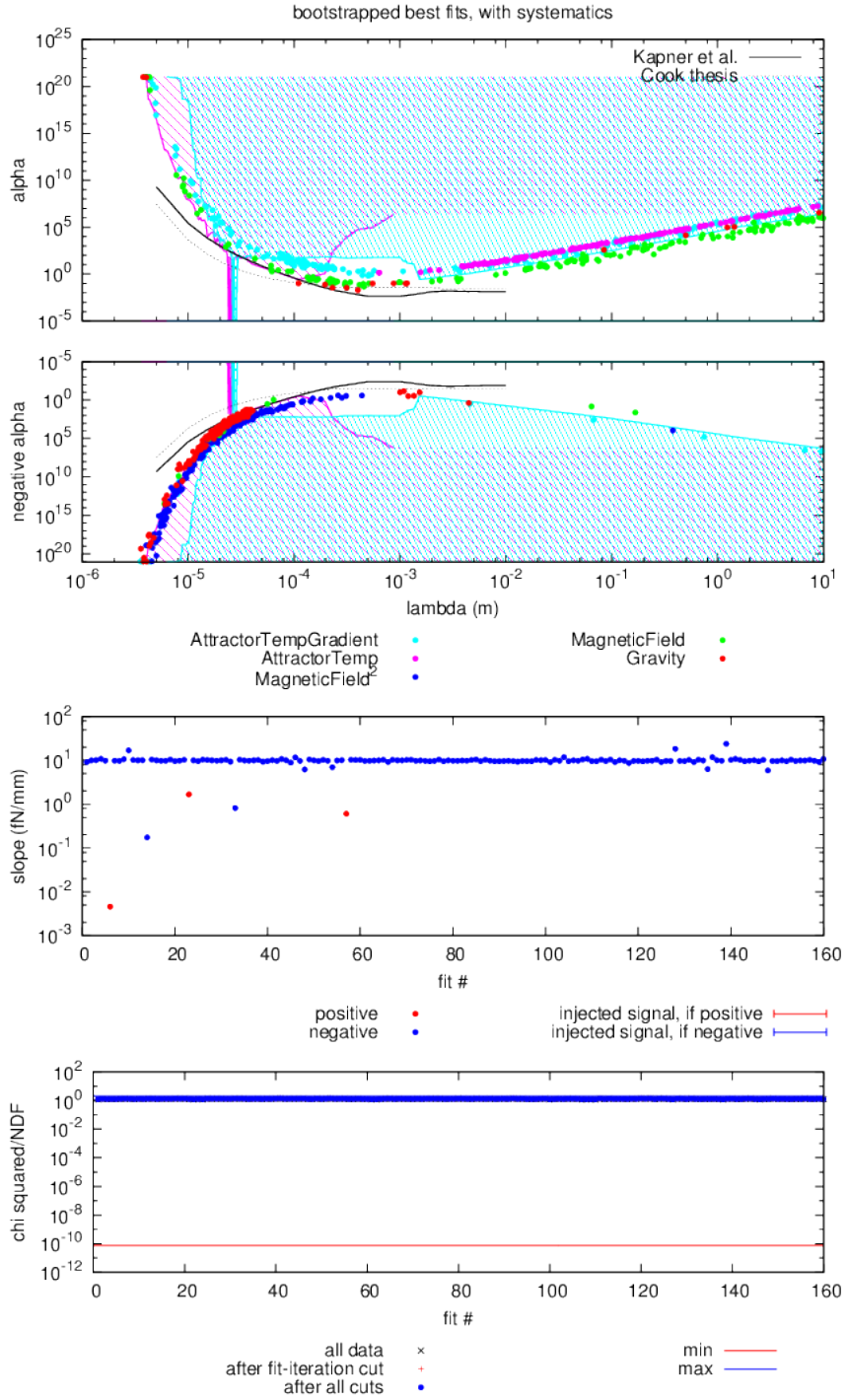

Figure 12.4: Yukawa exclusion plot, showing best-fits for all systematic effects. Foil systematic subtracted.

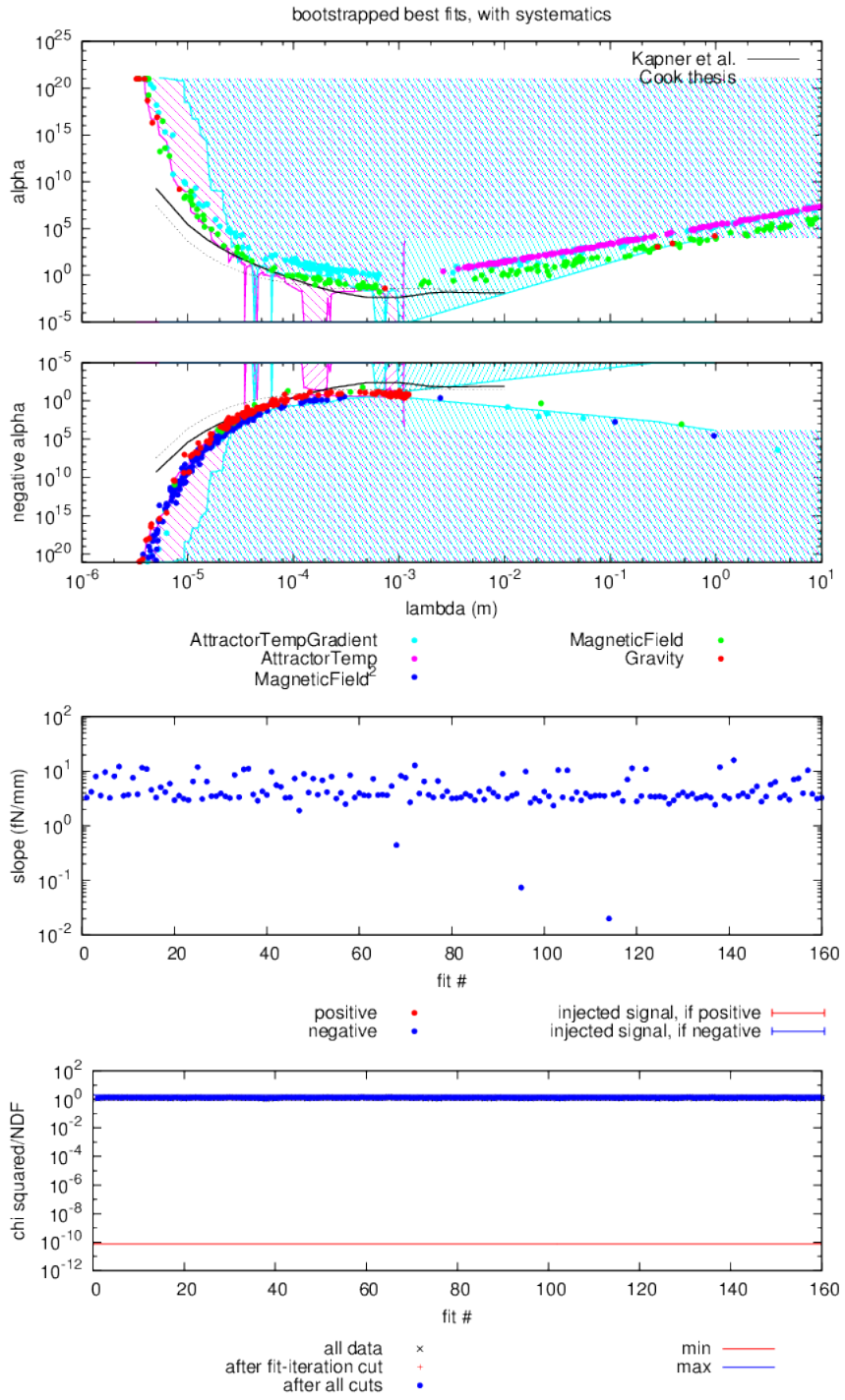

Figure 12.5: Yukawa exclusion plot, showing best-fits for all systematic effects. Foil systematic added.

Chapter 13

Future

13.1 Lessons learned

Scott Schell earned his spot in the acknowledgements when he told a backcountry-ski class, "I think the major difference between an expert and an amateur is whether or not, at the end of the day, they ask and answer the question, 'what have I learned today?' " If we learn from mistakes and avoid repeating them, our expertise can compound itself, exceeding anyone's innate natural talent, skill, or intuition.

The PlateWash experiment began as an attempt to make a lightning strike into new territory. The design is efficient; there is no more powerful geometry for raw signal. The systematic uncertainties for Newtonian gravity are easy, the rest are hard. With an optimistic view toward the latter, it's easy to see how we predicted we'd hit $\alpha = 1$ at 40 μm in $\sim$2008, and reach 20 μm in perhaps 2010. That this result arrives in 2015 with a limit at greater than 50 μm is evidence that we made a mistake, either in our estimate of our capability, in our execution, or in both. We made mistakes in both. I do not mean to say that the construction and execution of this experiment is not an achievement, it is, but we will repeat our mistakes if we do not learn from them.

The field of avalanche safety, and professional risk management more generally, shares its soul with experimental physics: the control of systematic uncertainty. In the case of avalanches, if you underestimate your uncertainty more than a few times, you die. If you overestimate your uncertainty, you never get to experience the joy found among mountains. For physicists, the situation is similar; if we report a violation of the vaunted Newton's Inverse Square Law Of Gravitation, and it's not true, our careers are damaged, if not ended. If we cower in fear of uncertainty, we'll never learn anything new about Nature. Risk-management tools have real value for physicists.

In the modern avalanche world, communication and an understanding of human interactions are prized. The hazards present in avalanche-prone snow can be prominent or subtle; it is the subtle ones that most-often kill the experienced backcountry traveler. In most avalanche accidents among the avalanche-aware, the clues are present, and the error is one of judgment [126, 127]. It is more rare for a group to get in trouble if it has discussed everyone's safety concerns. Everyone has blind spots, sometimes exacerbated by our environments. For us, as physicists, I believe it is the same.

13.1.1 Technical

13.1.1.1 0 tunable parameters, 1 tunable parameter, 2 tunable parameters

Our research group made its name with Equivalence Principle tests [128]. A rotating torsion balance with test masses of equal gravitational mass but differing composition looks for a composition-dependent differential acceleration toward a distant attractor mass. In this experiment, the science signal appears at the rotation frequency. For both earth-fixed and astronomical field masses, the measurement can be distilled to two numbers – the in-phase and out-of-phase components of a torsion balance signal. For these experiments, while the execution and source-mass determination can be challenging, it is relatively straightforward to write down the error budget. For each source of systematic uncertainty, one need only determine the magnitude of the external influence and its coupling to the instrument. From there, the effect on the overall measurement can be estimated.

From these measurements, constraints on interactions with an intrinsic lengthscale, usually a Yukawa length λ, are derived. The uncertainty in these constraints are affected by uncertainty in the field-mass integration, but there are no experimental uncertainties that feed directly into distance-dependent uncertainties. This is a critical difference from our inverse-square law experiments.

In our group's hole and wedge-pendulum experiments [16, 32, 38, 61, 60], the situation is similar, but more complicated. A mass distribution with N-fold symmetry is suspended horizontally on a torsion balance. Below it is placed a similar N-fold-symmetric field mass. The field mass is rotated smoothly, and the $N\omega$ in-phase and out-of-phase torque signals on the torsion balance are measured. In an important difference from the Equivalence Principle tests, the field mass is then displaced relative to the pendulum and the measurement repeated. In this way, these experiments can be viewed as a collection of Equivalence-Principle-style measurements. Each measurement can be used independently to constrain new interactions in the $\alpha - \lambda$ plane. Taken together, however, it is a greater challenge to estimate the ensemble's sensitivity to systematic uncertainties. For each source of systematic error, a model of the systematic effect and its pendulum coupling must be constructed as a function of height.

In the PlateWash experiment, things are even more complicated. For each individual cut, we have the choice of start and end positions. While this extra dimension of flexibility gives us extra discriminating power between specific inverse-square law violations, it also requires that we have an even better model of each systematic effect. With the intentionally myriad choices of start and end positions that we used in this experiment, ascertaining both our sensitivity to a specific inverse square law signal and our uncertainties became a real challenge. The methods outlined in Sections 5.8 and 5.10 now allow us to do this relatively efficiently, but it can be challenging to extract or judge a physics result by eye. The two-parameter approach is powerful for discriminating against instrumental effects, but the loss in clarity is substantial.

13.1.1.2 Size

We don't yet know with confidence what the optimal spatial size of active elements in an inverse square law experiment might be. The smaller the active elements, the more-easily they can be aligned. Larger sizes yield larger lever arms for torsion balances and more-uniform gravitational fields. The very direct geometry of the Huazhong apparatus [18] addresses the former concern, providing both easy alignment and a long lever arm at the possible expense of larger Newtonian gravity backgrounds. The scaling laws for proximity-enhanced noise with distance/area are strongly dependent upon the underlying phenomena (patch fields, gas molecule rearrangement, proximity-enhanced gas damping, etc), which are, at best, partially understood.

We will continue to experiment with size in an effort to improve our signal to noise ratio.

13.1.1.3 Foil-displacement

We spent about a year trying to find ways to avoid the need to measure the foil position before taking on the problem directly with the foil interferometer. In our opinion, for any attractor-displacing short-distance gravity experiment, a measure of the isolating foil's displacement is essential. Without such a measurement, substantial systematic-uncertainty questions will linger. The foil interferometer is one of the major successes of this experiment.

13.1.1.4 Capacitor/contact potential

The decision to simultaneously remove both the bypass capacitors and the contact-potential system was over-eager, though we didn't know it at the time. Had we ascertained that the systematic signal we saw was related only to the capacitors, we could have significantly improved the sensitivity of the overall measurement through decreased noise and decreased pendulum-foil coupling. Indeed, the most important single upgrade of the experiment will be to reconnect those systems.

It remains important for us to determine the mechanism by which the capacitor-modulated resonance couples to the pendulum/pendulum readout. Addressing that coupling, which probably resides in the auto-collimator pre-amplifier, may yield improved noise and decreased systematic sensitivity.

13.1.2 Strategic

The most important errors we've made are ones of strategy. Just as months in the lab can save you hours in the library [129], small strategic errors can have large consequences over time.

13.1.2.1 Error Budget

Perhaps the most important single concept when planning an experiment is the formation of a complete error budget.

Our single biggest mistake is the lack of a complete quantitative error budget. This seems like a childish oversight, but it's inherent to every modern short-range gravity experiment because it is extremely difficult to do. When we proposed the experiment, we submitted an approximate error budget for a single square-wave operation of the experiment, focused on $\alpha = 1$ and $\lambda \sim 20 - 70$ μm. As the exclusion limits span orders of magnitude in both λ and α, it is difficult to build reliable models for the influence of each systematic at all points in the $\alpha - \lambda$ plane.

Error budgets provide the guidance necessary for keeping experiments from going adrift. To operate without them is to operate without a rudder, reliant upon intuition alone for direction.

13.1.2.2 Blitzkrieg vs Campaign

While it was first imagined as a short-range equivalence principle test, and again as a competing design idea to the early Eöt-Wash 10-hole pendulum design, PlateWash, as it exists now, was sold in 2006-7 as a chance to jump ahead of the existing technology.

Our action mirrored this blitzkrieg philosophy. The goal was to push for the single best measurement we could get, as quickly as we could. We put off complete solutions to systematic concerns in favor of estimates and a promise to tie up loose ends when the measurement was complete. When the experiment was first assembled, our first day of data easily surpassed the first Eöt-Wash short-range experiments; we were on our way, led by the huge potential of PlateWash's efficient design.

Our commitment to the lightning-strike approach, where we would get a great measurement on our first try, led us astray; the first strikes didn't work. Complete focus on delivering a respectable result instead of iteratively delivering coarse but reliable results slowed our progress. Without building a foundation of

complete measurements on which we could stand, grow, and improve, we condemned ourselves to inefficiency.

13.1.2.3 Complexity

PlateWash is a very simple experiment. Its simplicity is perhaps its second-strongest feature (behind the intrinsically-null nature of the short-range field from the attractor). In pursuing the random attractor motion, blind analysis, and audit-able/repeatable analysis software, I severely underestimated the implied complexity of each. Correct execution of all of them are necessary to yield a maximally-believable result, but each one has generated a year or more's worth of work. Constrained complexity must be a hallmark of future short-range experiments. PlateWash has led the way into new areas and generated new tools to simplify each of these complex tasks, but we must be mindful of the pitfalls to which it's been subject.

13.1.2.4 Blindness

Prior to PlateWash, our group has only done one measurement effectively blind, when an important experimental parameter was undocumented during the duration of the experiment and subsequent analysis. It was not by choice. The approach taken with this work is designed to directly address the (reasonable) objection that "no experiment is ever truly blind" – hinting at the reality that most blind experiments unblind before making their work public. The repeatable analysis, open git commit log, and the public unblinding of the experiment are all designed to address this concern.

The systematic influences in known short-range gravity experimental configurations are too delicate to work with without a blind. How else will you know when to stop looking for uncertainties? The next giant leap in this field will come with an entirely new and simpler experimental idea; one where it is obviously free from large classes of systematic uncertainty. In the early days of those experiments, perhaps no blind will be requisite.

As this measurement comes to a close, while the blind remains stressful, it is also refreshing. We have done what we can to mitigate known systematic effects. The result, whatever it will be, is as unaffected by our own biases as we could manage. If our measurement is to be repeated, it should be performed blind.

13.1.3 Social

It may have been a mistake to have insisted so strongly that we follow the paths of replicable analysis and blindness. Both approaches are new to our research group. To depart on a headstrong approach to an analysis without gaining consensus from within the team was dangerous. Without buy-in from everyone involved, team members' interest in helping with the details of the project dropped considerably. Everyone

in the project agrees that these challenging approaches have merit, but substantial disagreement regarding how much those merits are worth remains. All of the senior members of the team are truly excellent at non-blind measurements; it may have been a more-efficient use of everyone's time and resources to have proceeded down that path instead, even if that path is intrinsically biased toward zero. From an opposite viewpoint, the efficiencies of non-blind analysis become increasingly clear as a project nears completion, something to which a junior researcher can be blind. In either case, we failed to coalesce behind a single plan, to our considerable detriment.

13.2 Upgrade plans

With a complete parallel-plate measurement under our belts, we're in a position to make important small upgrades. The biggest wins to be had are in modest improvements to the hardware, especially in tuning the contact potential. That single change has the potential to reduce both the largest source of statistical noise and the foil-coupling systematic. That single change, alongside easy improvements to the experiment's alignment, has the potential to produce a superior result within months.

13.2.1 Hardware

13.2.1.1 Contact potential

As noted in Sections 2.4.1.4, 2.4.1.6, and elsewhere, our inability to set the contact voltage was the biggest single blocker for this measurement. The elevated electrical potential difference between the pendulum and the foil is probably responsible for the level of swing coupling we see, limits the distance of closest approach between the pendulum and the foil, and surely enhanced the pendulum/foil coupling that forced us to consider a systematic correction (Section 10.2). In addition, the presence of a possible foil/attractor contact potential may have led to the observable short-range attractor foil coupling that makes a pendulum/foil systematic correction so daunting.

The fix is clear. We have very nice fully-isolated battery-powered contact potential setting devices built and ready to go. We need only reconnect them, now that we understand that our systematic signal was due to the 'capacitor systematic' and not something in the contact-potential system.

13.2.1.2 Pendulum Battery

As we showed with Figure 2.31, it is probable that the pendulum had its own trouble with an irreducible on-board battery, formed from the contact of the dissimilar metals of the pendulum body and the dense

inlays. Alternatively, it is also possible that different regions of the foil surface held slightly different effective electrical potential. Either way, as measured, this voltage may have contributed ~ 100 mV to the contact potential. A monolithic pendulum with a much-smaller area of closest approach, limited only to the active mass, may yield substantially better noise performance. If the dense/close side of the pendulum is sufficiently thick, it is also possible to imagine attaching a dense protruding "inlay" to a lightweight pendulum body (perhaps glass or aluminum), minimizing the pendulum moment of inertia but still allowing complete control over the dense material/foil contact potential.

These approaches are only possible with the simultaneous use of a foil interferometer. The intentional loss of symmetry inherent to these designs requires that even more care be given to the foil's motion. While a tuned contact potential should limit the electrostatic effects of any swing, gas dynamics and other effects presently canceled by approximate symmetries may become more pronounced.

13.2.1.3 Pendulum flatness

Our pendulum was a prototype, pressed into service when our superior pendulum failed. As seen in Figure 2.10, a new pendulum could be made flatter with known fabrication techniques. We can surely get improved metrology, decreased separation, and improved active area with a flatter pendulum. A refined pendulum is likely to be implemented in parallel with the pendulum-battery reducing design mentioned above.

13.2.1.4 Autocollimator

The autocollimator used in this iteration of the experiment was a coarse workhorse. The wide dynamic range, software-defined rotation, and two-dimensional operation were all quite useful to making this measurement possible. The noise floor of the autocollimator, however, wasn't too far away from the intrinsic noise of the torsion balance. Improved readout would allow us to run higher-gain feedback loops, with larger differential terms. The stiffer loop performance would allow us to operate at shorter separations and at higher attractor frequencies.

Furthermore, the encoder systematic (Section 8.2) most-likely emerged through electrical crosstalk between the autocollimator leads and the encoder. At present the leads that run between the photodiode and the autocollimator preamplifiers are 5-10 cm in length. By relocating the preamplifier board and improving the shielding further, we may be able to decrease the size of any encoder systematic. While it would require considerable testing to be trustworthy, the benefits of full-time encoder operation would be considerable. The attractor could operate in feedback, decreasing the positioning scatter by at least an order of magnitude, and there would be no need to use an air pressure gauge as a position monitor.

13.2.1.5 UV LED

Our understanding of exactly which physical mechanism increases the noise in the torsion pendulum at short pendulum/foil separation is minimal. One possibility is that the pendulum is being buffeted by outgassing from the foil, perhaps enhanced by squeeze-film effects. Another is that adsorbed materials on the foil surface may be slowly rearranging themselves, causing fluctuations of the effective electrostatic interactions between pendulum and foil. If either of these two are the culprit, as suggested by [92], it's possible that retracting the pendulum away from the foil and illuminating the active surfaces with ultraviolet light might improve the noise. Through photoelectric-like effects, UV light is known to significantly accelerate outgassing rates and expel adsorbed surface layers. Trying out a UV light-emitting diode, as we've already used in our LISA [130] experiment, would be straightforward, and might yield an improvement in noise.

13.2.1.6 Interferometer improvements

As described in Section 2.3.9.2, the interferometer may be limited by polarization-related noise. If we were to switch to entirely polarization-maintaining fiber, it's possible that the pickup and drift we see could be reduced. We have begun to acquire the necessary fiber components to do so.

In addition, it would be a great help in simplifying the analysis and controlling for the foil systematic if we could hold the interferometer in its sensitive region of operation for the duration of the experiment. That could be accomplished by physically moving the interferometer fiber tip with slow-control feedback or by changing the laser temperature enough to shift the laser wavelength. The new infrared laser is extremely resistant to temperature change, which may force us to develop a mechanism of moving the fiber tip in a smooth and controllable fashion. The geared and levered-down servo that presently positions the fiber tip is too coarse for the job.

It has been suggested that we should replace the 90/10 coupler with a fiber circulator to increase the amount of light that reaches the detector by more than a factor of ten. Early experiments on tilt-meter interferometer test apparatus by Krishna Venkateswara suggest that a circulator will be a drop-in replacement for the existing coupler and will yield increased sensitivity.

13.2.2 Software/Strategy

13.2.2.1 Position choices

Much of the difficulty in the data analysis stemmed from the use of a large ensemble of attractor positions instead of two or three. If we had instead chosen to operate with one or two position differences, the analyses of both systematic uncertainty and potential new physics would have been drastically simplified.

The increased immunity to position-related systematic effects may not have been worth the massive cost in complexity. While blindness is still a challenge, it's equally possible to blind such a few-position measurement.

13.2.2.2 Interleaved systematic studies

The strategy of taking science data as quickly and comprehensively as possible and then following up with systematic studies was very pragmatic. Future measurements would benefit from interleaving systematic studies with science measurements (see [131] for a beautiful and extreme example). Each cut of science data could be immediately followed by repeating the cut with a varied systematic parameter (magnetic field, interferometer light power, attractor bias voltage, etc.). This isn't possible with all sources of systematic uncertainty (imposed temperature gradients take thousands of seconds to equilibrate), but it would relax the requirement that the instrument remain untouched and stable for months at a time. In addition, in-line systematic studies offer the opportunity to monitor the expected sensitivity of an entire measurement campaign in real time.

13.2.2.3 Blindness

Now that we have completed an entire measurement cycle with a blind measurement, we must revisit as a group whether or not a blind has been helpful. I contend that it has been essential to producing an unbiased result. Whether we choose to go blind for a second iteration or not is perhaps less important than whether we agree on whether we should. Whichever method we choose, we need to be all in it together.

13.2.2.4 Python?

Octave has been a good and reliable tool throughout this work. While the Octave project continues to improve and work toward drop-in compatibility with MATLAB, much of the open-source effort has moved into the SciPy packages for Python. Writing future analysis code in Python will make it more widely readable and permit access to mathematical libraries that are audited by many more users.

A particular sticking point is our reliance upon the excellent Gnuplot plotting package. Data often flows out of an Octave program, gets plotted in or analyzed a little in Gnuplot, and then is no longer accessible to the tools in Octave. At least a man-month has been spent finding ways to smoothly move data from one to the other and back. A search for a more integrated analysis and plotting environment will be worthwhile.

Chapter 14

Conclusions

This chapter was constructed by combining the results of many runs of the unblinded analysis in order to accommodate the thesis-submission deadline. Building the analysis from the git commits noted in the chapter will allow reconstruction of these results. The best-fit points used to construct the confidence-interval plots below are included in the project's git repository to aid in reproducibility and alternate analysis. The preceding document was built using the "Ten Fewer Microns" analysis denoted in Section 14.2.

14.1 Unblinding

We unblinded the experiment in public on May 5, 2015. The talk can be seen in its entirety on the internet [132]. The unblinding process began using git commit hash 5d34825dd0aa6a4e1da560448ecce3dfcdc5ffff, as documented on Twitter [133] and elsewhere. The initial appearance of the exclusion (Figure 14.2) and arbitrary-fit (Figure 14.1) plots was extremely puzzling.

Hasty inspection made it clear that something had gone "weird", as it appeared that the torque-variance cut (Section 5.6) had failed to cut away substantial quantities of noisy data. With the blind restriction lifted, we could look directly at the central values of the torque data, plotted in Figure 14.4. It appeared that a section of data differed substantially from the rest, but didn't appear to be flagged as a systematic test.

The following day, we determined that the well-resolved and apparently anomalous data were from the tests associated with the "capacitor systematic" (Section 8.3). This made plain the fact that systematic test data for which we do not make a global fit were included into the global fit data, a clear clerical error. In all, nine runs were included in the global fit that should not have been. These runs comprised the "capacitor systematic" tests and the electrical-crosstalk test.

The capacitor-systematic data included a lengthy run with a well-resolved and, compared with the science

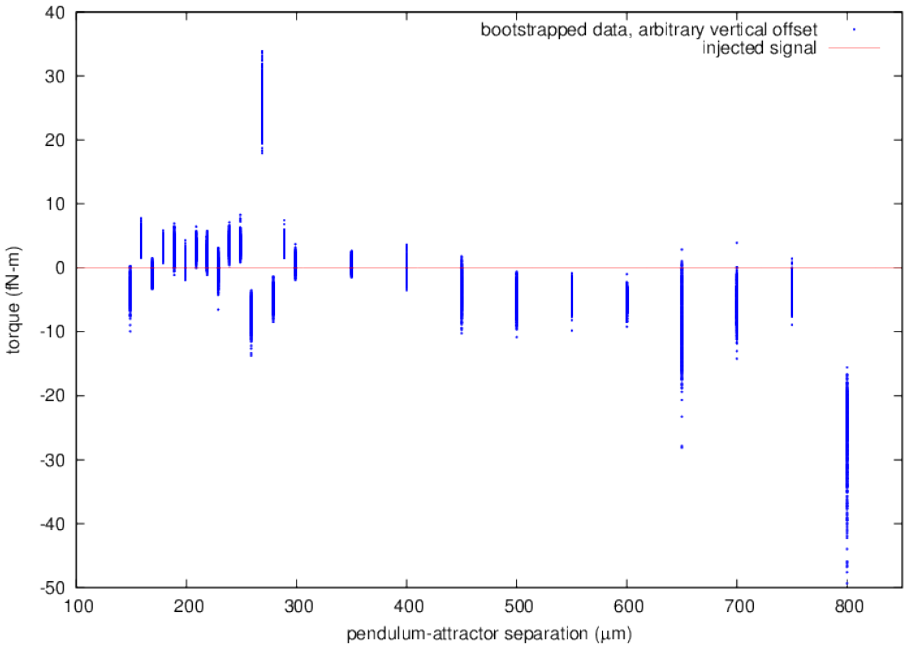

Figure 14.1: Unblinded arbitrary-fit plot, precisely as-unblinded. Something is clearly amiss.

data, gigantic signal. See Figure 9.1 for the well-resolved >100 fN-m signal. This necessarily altered the behavior of our adaptive variance cut (Section 5.6), allowing extra transients into the final dataset. The combined inclusion of a well-resolved signal and added non-Gaussian outliers made the result meaningless.

With that knowledge, on May 7, we began fitting again with those irrelevant runs removed. The changes are summarized with the commit hash 6f9be5dea4c24126fb9bbc7b8f3cbf8155c76651. In order to generate a smoother fit with higher bootstrapped statistics in limited time, we parallelized the fitting process further with commit 537b7685094cde1d00dd3e464758fb32732c5e38. The aggregated outputs from several repetitions of the fit were used to generate the confidence intervals shown here. The fit results are included in the document's code repository. The resulting fits and the associated confidence intervals are shown in Figures 14.5 and 14.6.

These results should be regarded as the unblinded scientific outcome of the experiment. Jens Gundlach points out that the fact that such a blatant clerical error might have been allowed to be unblinded publicly is evidence the blind analysis was truly blind. If we had peeked into the raw data or fit outcomes, we surely would have repaired the problem.

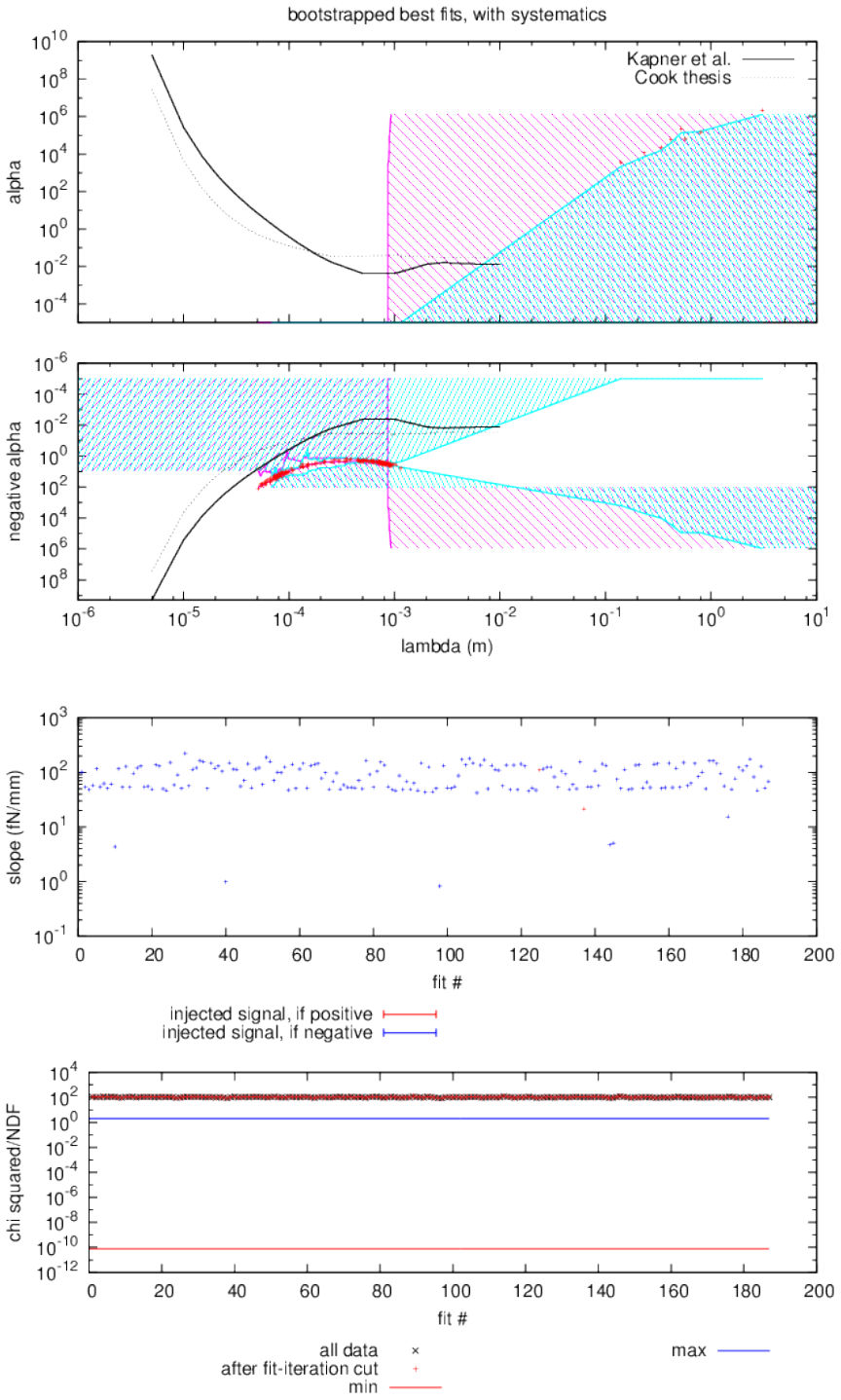

Figure 14.2: Alpha-lambda plot, precisely as unblinded. Our apologies for the plotting artifacts; the distribution of the best-fit points in the upper two panels is of greatest importance. Interferometer correction subtracted. A signal is clearly resolved. Note poor χ^2.

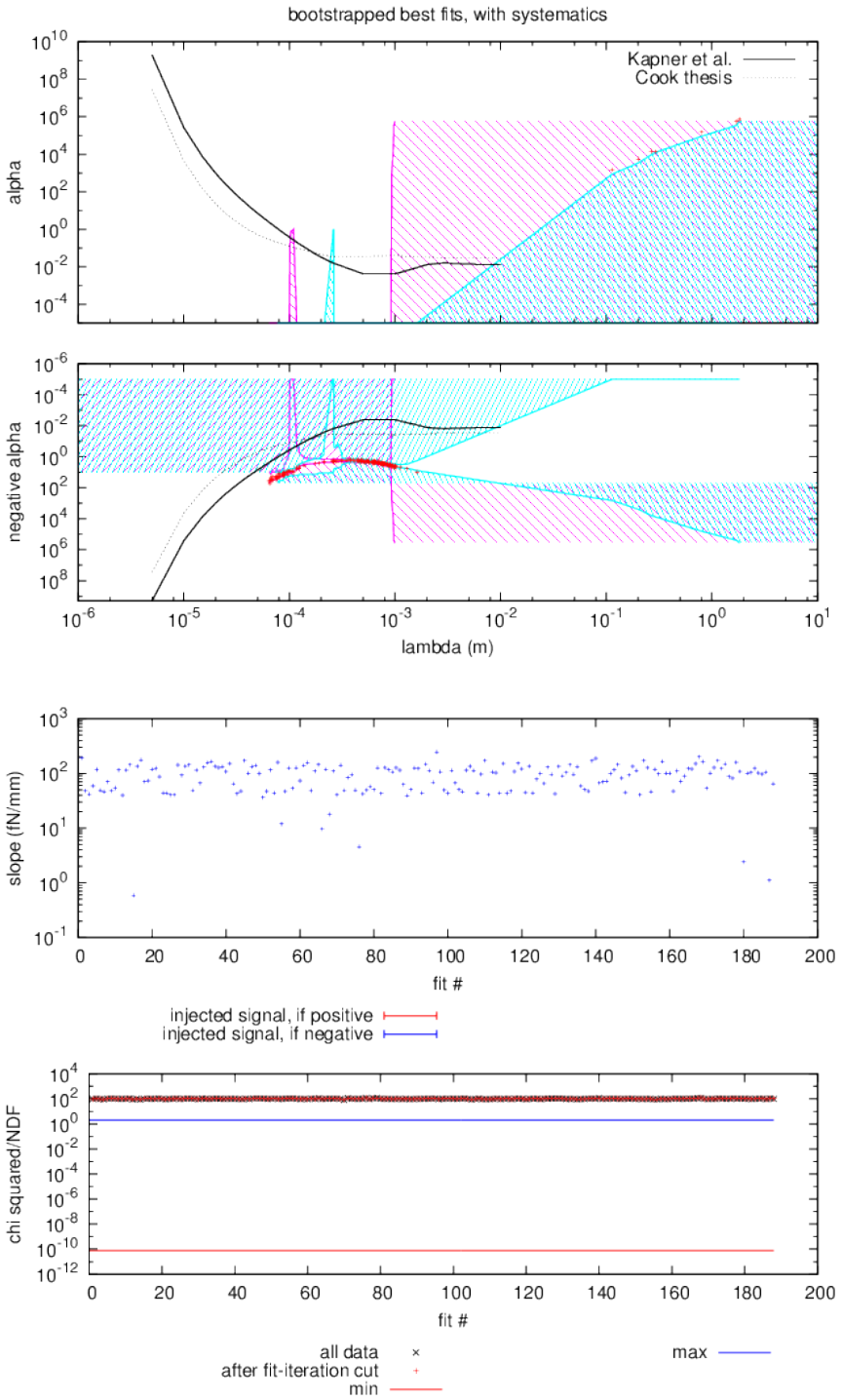

Figure 14.3: Alpha-lambda plot, precisely as unblinded. Our apologies for the plotting artifacts; the distribution of the best-fit points in the upper two panels is of greatest importance. Interferometer correction added. A signal is clearly resolved. Note poor χ^2.

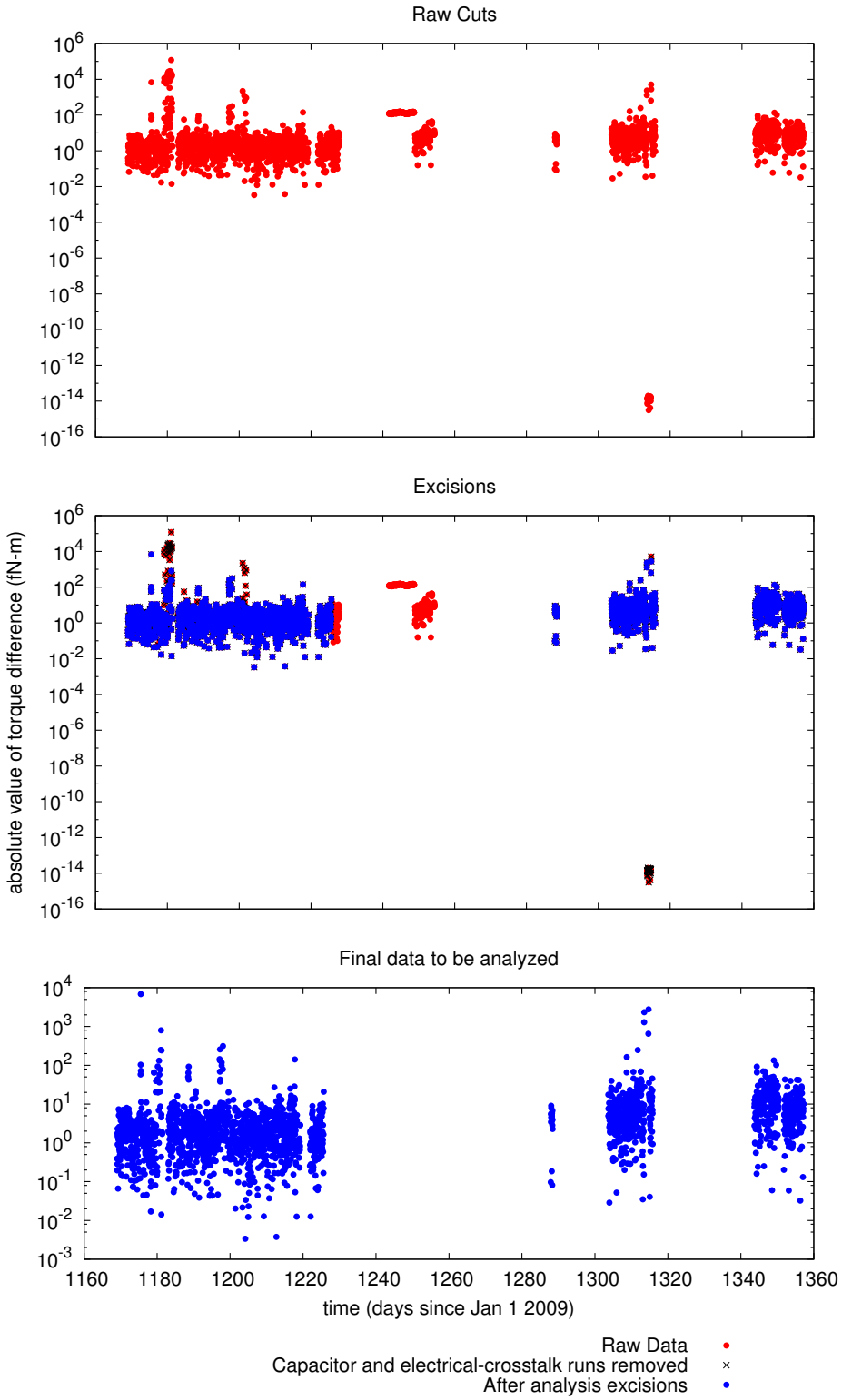

Figure 14.4: Per-cut torques, uncut, after each excision, and in the now-final form. The red-circle "Raw Data" points that appear without overlying points in the middle frame were the incorrectly-included capacitor-test and electrical-crosstalk test runs.

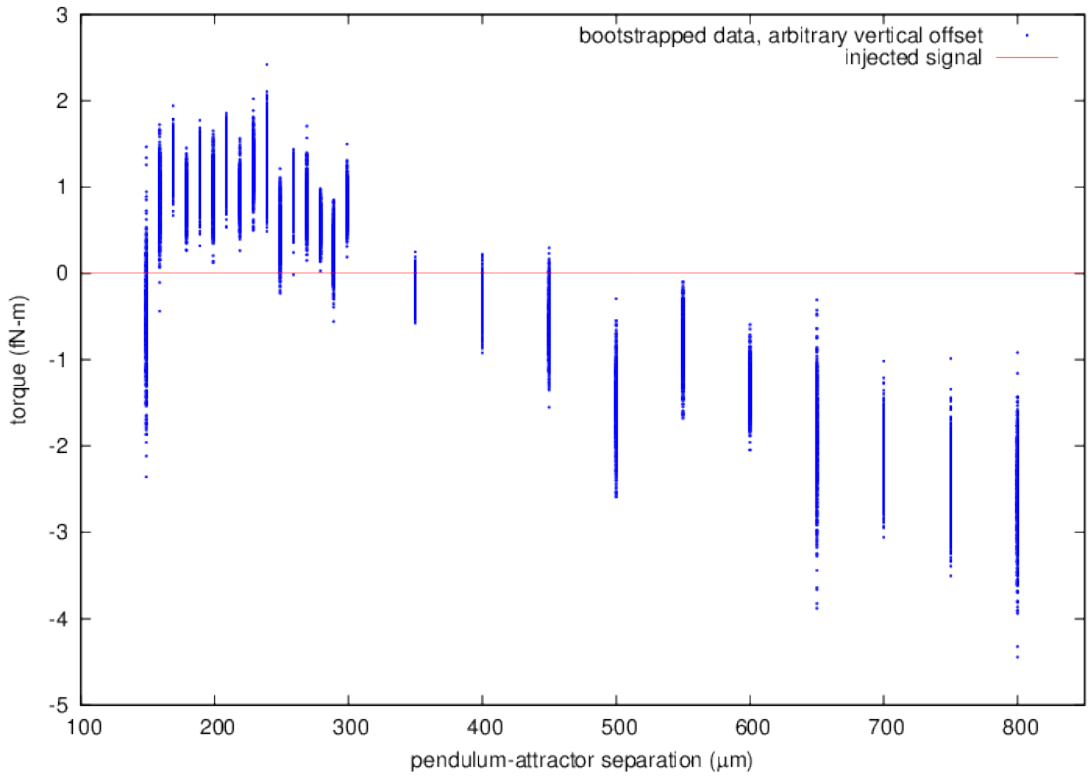

Figure 14.5: Arbitrary-fit torque, as unblinded, with capacitor-test and crosstalk-test runs removed.

14.2 Ten fewer microns

It is clear from Figure 14.5 that there's a somewhat-resolved signal at the shortest distances. Comparison with the foil interferometer signal, Figure 14.7, shows a possible culprit. As noted in the pre-unblinding caption from Figure 10.2, residual foil-touching is an easy candidate for such a signal. We immediately ran an alternate analysis using a subset of the data ($\sim$ 2325 cuts instead of $\sim$ 2391, or 97.2% of the unblinded dataset), removing the closest 10 microns of the data. As seen in Figure 10.2, this removes any preferred direction of short-range interferometer signal. We have not explored any other modification of the excisions. This change is described in commit bec5a46fd98392eb35ff94573710246f95122900. As in the section above, we parallelized the code further, as in commit 9e8c5cd11ecc322d1ce47e8fd01587fc2af1b28d. The fits are shown in Figures 14.8 and 14.9. **We regard this "ten fewer microns" dataset as our most reliable result.**

To provide further understanding of the result, we additionally include a zoomed-in plot (Figure 14.10) and a plot with three-sigma confidence intervals (Figure 14.11). The three-sigma plot should represent 95% confidence intervals in a strict interpretation of the look-elsewhere effect.

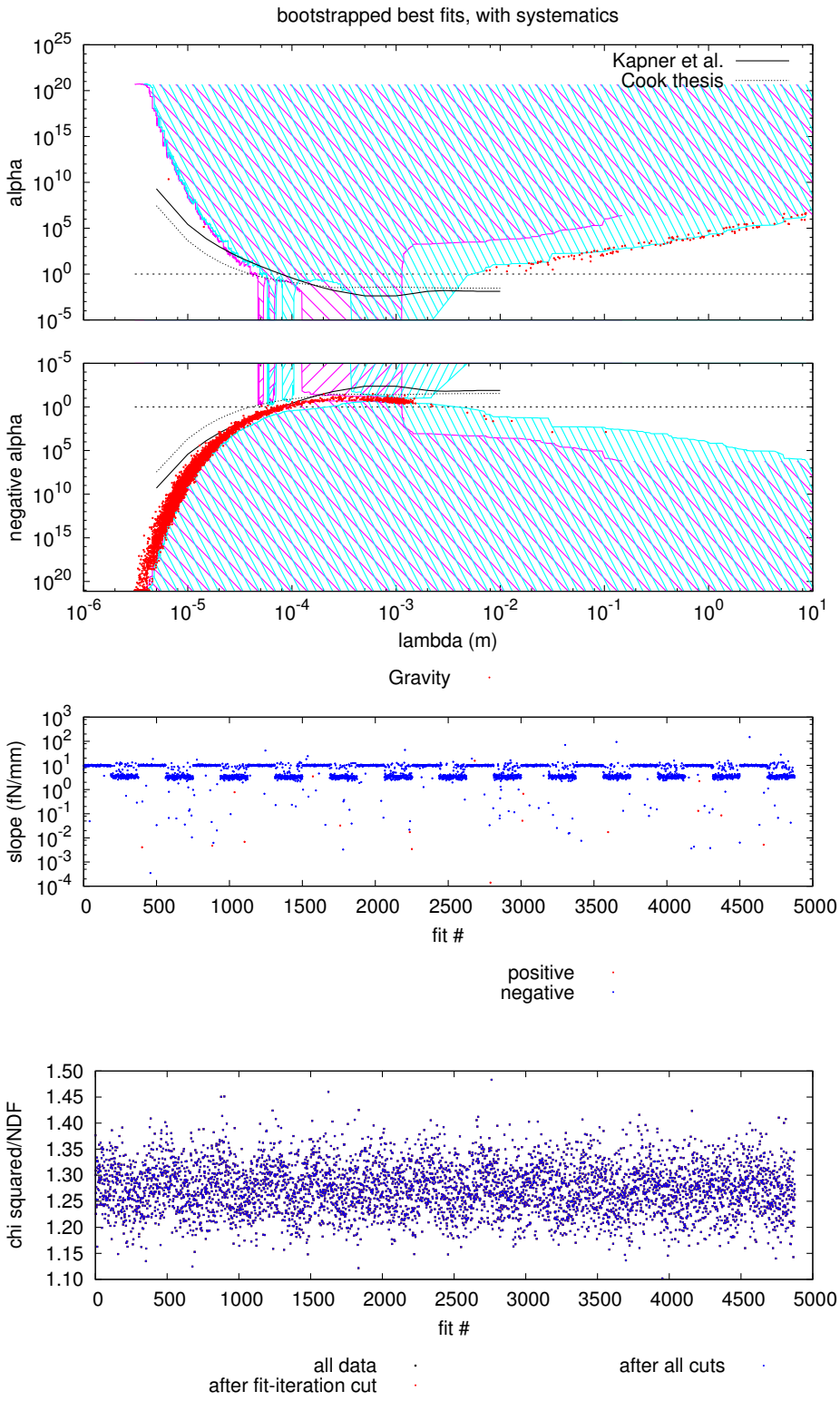

Figure 14.6: Yukawa Fit, with irrelevant runs removed. Interferometer correction is both added and subtracted, creating the alternating pattern in the fitted slope. A signal is weakly resolved in this analysis at sub-gravitational strength with negative α and $\lambda \sim 1$ mm.

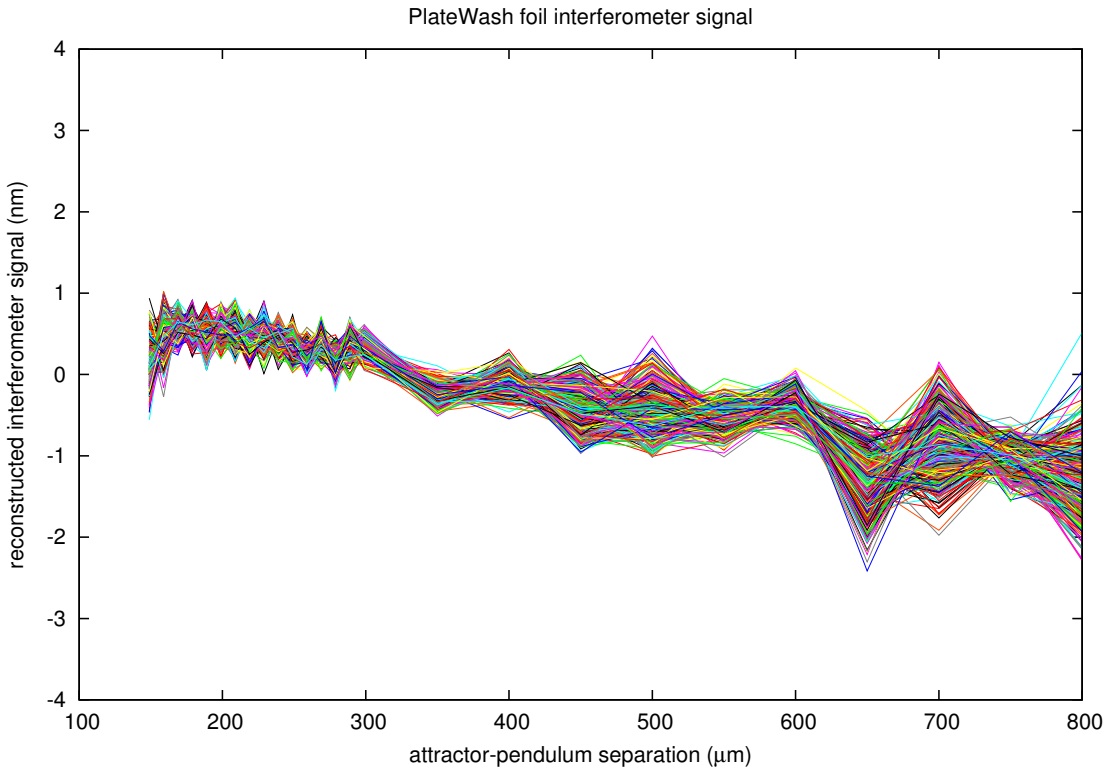

Figure 14.7: As-unblinded foil-interferometer signal reconstruction

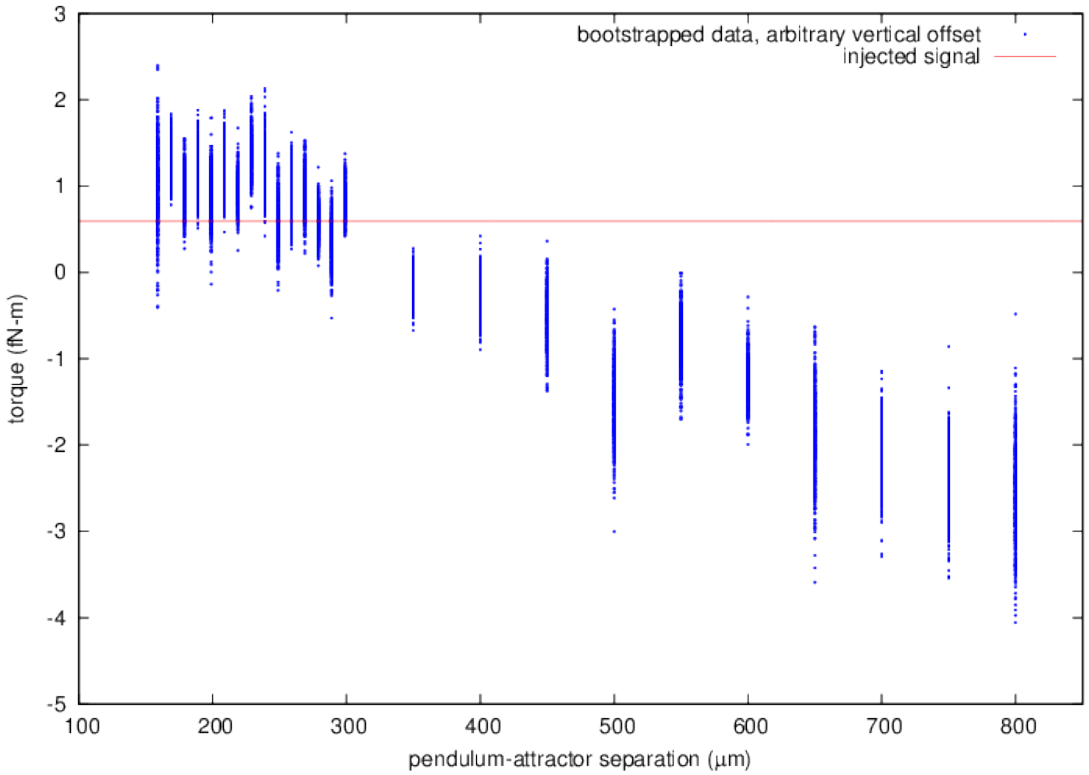

Figure 14.8: Arbitrary-fit torque reconstruction, ten fewer microns.

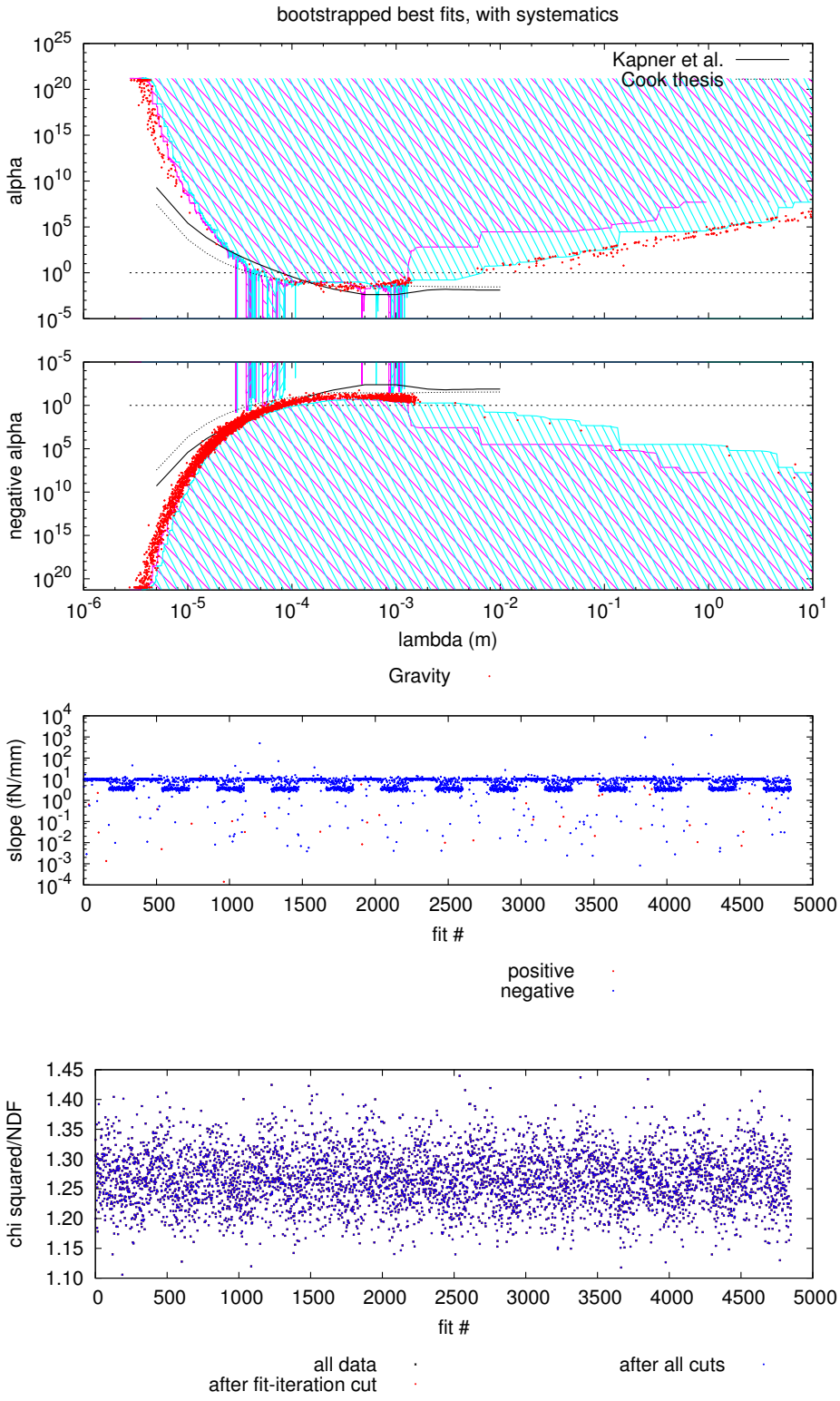

Figure 14.9: Yukawa fit, ten fewer microns. The interferometer signal has been both added and subtracted, creating the alternating pattern in the fitted slope. A signal is weakly resolved both slightly below 100 μm and near 1 mm.

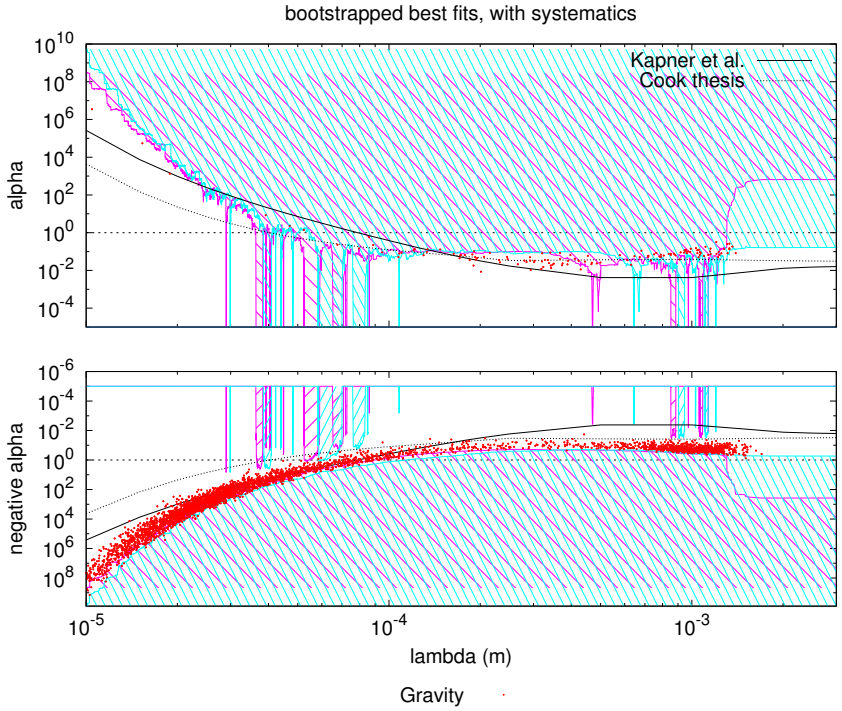

Figure 14.10: Enlarged version of Figure 14.9. A signal is weakly resolved both slightly below 100 μm and near 1 mm.

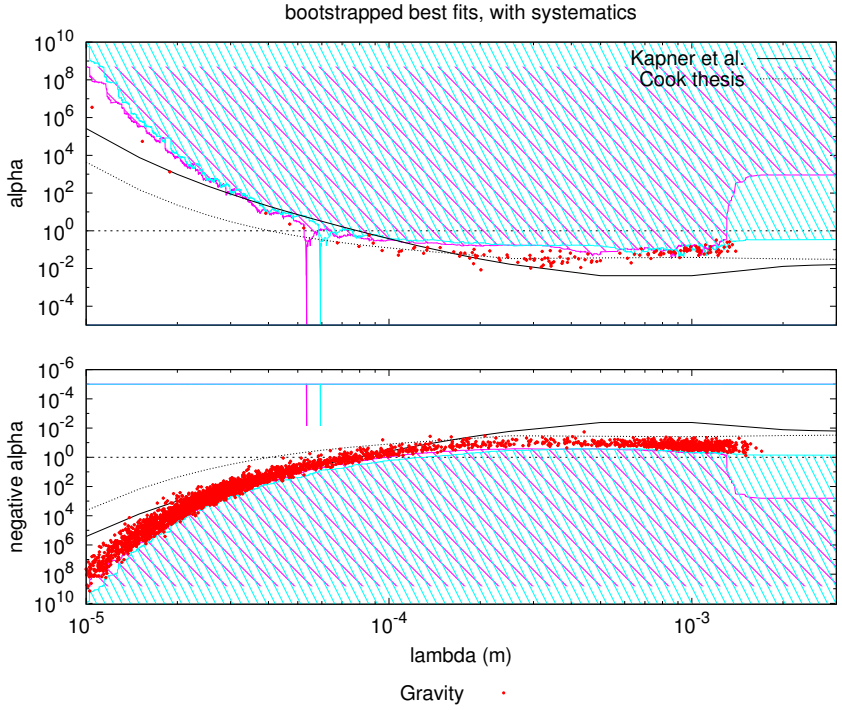

Figure 14.11: Enlarged version of Figure 14.9, generated with 3-σ confidence intervals. As there are < 100 independent regions sampled by the confidence-interval determination, this represents a strict accommodation of the look-elsewhere effect at 95% confidence. No significant signal is observed.

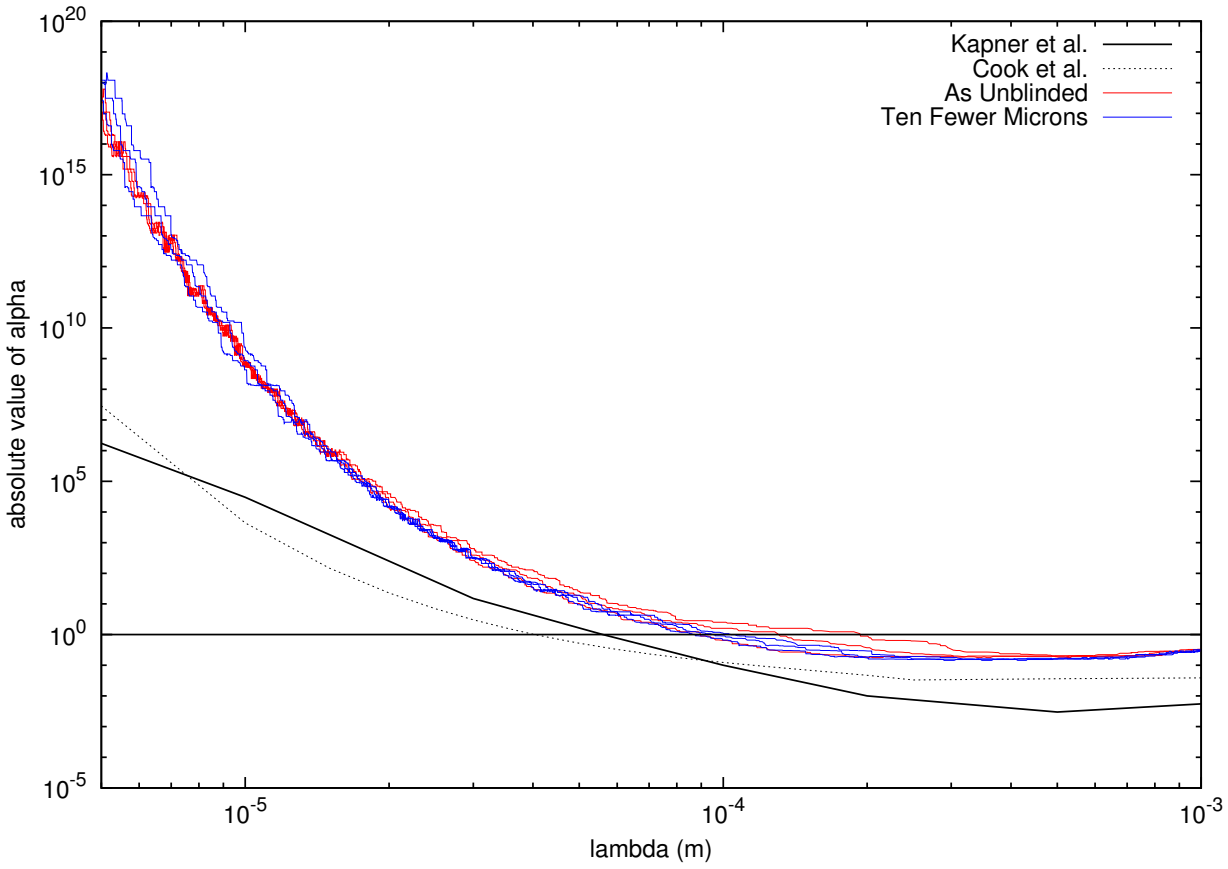

Figure 14.12: 95% limits on the Yukawa-parametrized $|\alpha|$

14.3 Absolute $|\alpha|$ limits

It is a technical hurdle to take the inclusion limits from the preceding sections and reliably generate exclusion limits for $|\alpha|$. Fortunately, the bootstrapped data allow us an easy path forward. Just as with the inclusion limits, we take the approach of "Turner Smoothing" (Section 5.9), and move a fixed-number bin (an integer-multiple of 20, to make accurate 95% limits) left-to-right through the best-fit points. For each binning, we sort the points into increasing order of $|\alpha|$, discard the top 5% of points, and take the next-largest point (19th of 20, 38th of 40, 57th of 60) as representative of the 95th percentile. The result is plotted in Figure 14.12.

While this approach yields a pretty and easily-interpreted plot, we caution that it is more meaningful to consider the two-sided plot of the confidence intervals. It is not possible to combine two experiments nor compare weakly-resolved signals using a $|\alpha|$ plot, while two-sided limits can be combined and compared.

14.4 Conclusion

We demonstrated a thorough approach to parallel-plate inverse-square law tests. The parallel-plate geometry allows an efficient null test for new physics, at the cost of many attractor-position-correlated systematic uncertainties. Our experiment incorporated a number of systematic-uncertainty mitigating technologies: pneumatic drive, random attractor positioning, and the critically-important interferometric foil-position monitor. Our analysis was performed largely blind; the approach to systematic uncertainty was determined without reference to any effect upon the final result. The analysis, from raw data to finished product, is reproducible. The analysis was unblinded in public.

After addressing the obvious clerical error in our run configuration and making a more-conservative interferometer-guided cut, no evidence for a significant violation of Newton's inverse-square law is observed. Using the traditional Yukawa parametrization (Equation 1.1), our results are are consistent with the inverse-square law ($|\alpha| < 1$) for λ between 104 μm and 1320 μm with 95% confidence. This limit is expected to improve to 95 μm with the improved fit resolution permitted by even more computation of best-fit points.

The code that generates this entire body of work is freely available:
https://www.github.com/4kbt/PlateWash/ and by request from the author.

Appendix A

Yukawa Variance and the Detwiler Square

A.1 The fitting plan:

The fitting plan:

1. Define a fitting function,

$$\begin{aligned} F(x, B_{1\ldots n}|\alpha_{1\ldots n}, \lambda_{1\ldots n}) &= \pi G h \left(\frac{w^2}{4} - j^2\right)(\rho_h - \rho_l)\,\rho_a \sum_i \alpha_i B_i \lambda_i^2 \\ &\quad \left(1 - e^{-(t_p - t_I)/\lambda_i}\right)\left(1 - e^{-t_a/\lambda_i}\right)\left(1 - e^{-t_I/\lambda_i}\right) e^{-x/\lambda_i} + Cx \end{aligned}$$

.

2. Let the thickness corrections be $T(\lambda) = \rho_a \left(1 - e^{-(t_p - t_I)/\lambda_i}\right)\left(1 - e^{-t_a/\lambda_i}\right)\left(1 - e^{-t_I/\lambda_i}\right)$

3. So,

$$\begin{aligned} F(x, B_{1\ldots n}|\alpha_{1\ldots n}, \lambda_{1\ldots n}) &= \pi G h \left(\frac{w^2}{4} - j^2\right)(\rho_h - \rho_l) \sum_i \alpha_i B_i \lambda_i^2 T(\lambda_i) e^{-x/\lambda_i} + Cx \\ &= Q \sum_i \alpha_i B_i \lambda_i^2 T(\lambda_i) e^{-x/\lambda_i} + Cx \end{aligned}$$

4. Define $G(x, y\ldots) = F(x, \ldots) - F(y, \ldots)$

5. Measure $\Delta\tau(x_A, x_B, B_1, \ldots, B_n, \sigma_{x_A}, \sigma_{x_B}, \sigma_{B_1}, \ldots, \sigma_{B_n})$.

6. Define $P(a, \bar{a}, \sigma_a) = \frac{1}{\sigma_a\sqrt{2\pi}} e^{-(a-\bar{a})^2/2\sigma_a^2}$
7. Compute $\bar{G} = \int_{-\infty}^{\infty} G(x, y, B_{1\ldots n}, \alpha_{1\ldots n}, \lambda_{1\ldots n}) P(x, \bar{x}, \sigma_x) dx P(y, \bar{y}, \sigma_y) dy \prod_i P(B_i, \bar{B}_i, \sigma_i) dB_i$
8. Compute $\bar{G} = \bar{F}_{x_A} - \bar{F}_{x_B}$
9. Compute $\sigma^2_{F_{x_A}} = \int_{-\infty}^{\infty} F(x, B_{1\ldots n}, \alpha_{1\ldots n}, \lambda_{1\ldots n})^2 P(x, \bar{x}, \sigma_x) dx P(y, \bar{y}, \sigma_y) dy \prod_i P(B_i, \bar{B}_i, \sigma_i) dB_i - \bar{F}^2$ and $\sigma^2_{F_{x_B}}$
10. Compute $\sigma^2_G = \sigma^2_{F_{x_A}} + \sigma^2_{F_{x_B}}$
11. Compute χ^2

$$\chi^2 = \sum_{measurements} \frac{(\Delta\tau - \bar{G})^2}{\sigma_\tau^2 + \sigma_G^2}$$

A.2 Integrals

Linear $\int_{-\infty}^{\infty} aP(a, \bar{a}, \sigma_a) da$
$= \frac{1}{\sigma_a\sqrt{2\pi}} \bar{a}\sqrt{2\sigma^2\pi} = \bar{a}$

Quadratic $\int_{-\infty}^{\infty} a^2 P(a, \bar{a}, \sigma_a) da$
$= \bar{a}^2 + \sigma_a^2$

Yukawa - Detwiler observation The χ^2 approach outlined here was enabled by Jason Detwiler's observation that the following calculation should be possible. If it were not true, a closed-form expression for χ^2 might have been too unwieldy to use.

$\int_{-\infty}^{\infty} e^{-a/\lambda} P(a, \bar{a}, \sigma_a) da$
$= \frac{1}{\sigma_a\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-a/\lambda} e^{-(a-\bar{a})^2/2\sigma_a^2} da$
$= \frac{1}{\sigma_a\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(a^2 - 2(\bar{a} - \frac{\sigma^2}{\lambda})a - \bar{a}^2)/2\sigma_a^2} da$
Completing the square,
$(\bar{a} - \frac{\sigma_a^2}{\lambda})^2 = \bar{a}^2 - 2\frac{\sigma_a^2}{\lambda}\bar{a} + \frac{\sigma_a^4}{\lambda^2}$
$= \frac{1}{\sigma_a\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-((a-(\bar{a}-\frac{\sigma_a^2}{\lambda}))^2 + 2\frac{\sigma_a^2}{\lambda}\bar{a} - \frac{\sigma_a^4}{\lambda^2})/2\sigma_a^2} da$
$= e^{\sigma^2/2\lambda^2} e^{-\bar{a}/\lambda}$

Linear Yukawa $\int_{-\infty}^{\infty} ae^{-a/\lambda} P(a, \bar{a}, \sigma_a) da$
$= e^{-\frac{\bar{a}}{\lambda} + \frac{\sigma_a^2}{2\lambda^2}} (\bar{a} - \frac{\sigma_a^2}{\lambda})$

Additive Yukawa $\int_{-\infty}^{\infty} e^{-x(1/\lambda_1+1/\lambda_2)} P(a,\bar{a},\sigma_a)$

This is just a restatement of Jason's trick, so

$= e^{\sigma_a^2(1/\lambda_1+1/\lambda_2)^2/2} e^{-\bar{a}(1/\lambda_1+1/\lambda_2)}$

A.3 $\bar{F}$

$\bar{F} = \int_{-\infty}^{\infty} F(x, B_{1\ldots n}, \alpha_{1\ldots n}, \lambda_{1\ldots n}) P(x,\bar{x},\sigma_x) dx \prod_i P(B_i, \bar{B}_i, \sigma_i) dB_i$

$\bar{F} = Q \sum_i \alpha_i \bar{B}_i \lambda_i^2 T(\lambda_i) e^{\sigma_x^2/2\lambda_i^2} e^{-\bar{x}/\lambda_i} + C\bar{x}$

A.4 σ_G^2

$\sigma_G^2 = \sigma_{F_{x_A}}^2 + \sigma_{F_{x_B}}^2$

$\sigma_{F_x}^2 = \int_{-\infty}^{\infty} F(x, B_{1\ldots n}, \alpha_{1\ldots n}, \lambda_{1\ldots n})^2 P(x,\bar{x},\sigma_x) dx \prod_i P(B_i, \bar{B}_i, \sigma_i) dB_i - \bar{F}_x^2$

Let $L_{ij} = 1/\lambda_i + 1/\lambda_j$.

$$\begin{aligned}
\sigma_{F_x}^2 &= Q^2 \sum_i \sum_{j\neq i} \alpha_i \alpha_j \bar{B}_i \bar{B}_j \lambda_i^2 \lambda_j^2 T(\lambda_i) T(\lambda_j) e^{\sigma_x^2 L_{ij}^2/2 - \bar{x} L_{ij}} \\
&\quad + Q^2 \sum_i \alpha_i^2 (\bar{B}_i^2 + \sigma_{B_i}^2) \lambda_i^4 T(\lambda_i)^2 e^{2\sigma_x^2/\lambda_i^2 - 2\bar{x}/\lambda_i} \\
&\quad + 2CQ \sum_i \alpha_i \bar{B}_i \lambda_i^2 T(\lambda_i) e^{\sigma_x^2/2\lambda_i^2 - \bar{x}/\lambda_i} (\bar{x} - \frac{\sigma_x^2}{\lambda_i}) \\
&\quad + C^2(\bar{x}^2 + \sigma_x^2) - \bar{F}_x^2
\end{aligned}$$

This works great, but speed and stability of numerical evaluation urges us farther.

$$\begin{aligned}
\sigma_{F_x}^2 &= Q^2 \sum_i \sum_{j\neq i} \alpha_i \alpha_j \bar{B}_i \bar{B}_j \lambda_i^2 \lambda_j^2 T(\lambda_i) T(\lambda_j) e^{\frac{\sigma_x^2}{2} L_{ij}^2 - \bar{x} L_{ij}} \left(1 - e^{-\frac{\sigma_x^2}{\lambda_j \lambda_i}}\right) \\
&\quad + Q^2 \sum_i \alpha_i^2 \left(\sigma_{B_i}^2 + \bar{B}_i^2 \left(1 - e^{-\sigma_x^2/\lambda_i^2}\right)\right) \lambda_i^4 T(\lambda_i)^2 e^{2\sigma_x^2/\lambda_i^2 - 2\bar{x}/\lambda_i} \\
&\quad - 2CQ\sigma_x^2 \sum_i \alpha_i \bar{B}_i \lambda_i T(\lambda_i) e^{\sigma_x^2/2\lambda_i^2 - \bar{x}/\lambda_i} \\
&\quad + C^2 \sigma_x^2
\end{aligned}$$

With this knowledge, we can compute χ^2!

If σ_x is zero, then

$$\sigma_{F_x}^2 = Q^2 \sum_i \alpha_i^2 \lambda_i^4 T(\lambda_i)^2 \sigma_{B_i}^2 e^{-2\bar{x}/\lambda_i}$$

Appendix B

Fit Injections

B.1 Arbitrary Fits

The bootstrapped "arbitrary force law" plots, computed as described in Section 5.10 are plotted in this section. These are piecewise linear fits to the data; as close to a model-free fit as one can reasonably get. Fits with signals added to the real (or simulated, if the thesis is built blind) data are shown alongside the injected signals.

B.2 Yukawa Fits

In this appendix we show, for completeness, the response of the global fitting algorithm to signals injected into the science data. Vertical/horizontal cross points are the location of injected signals, to show that we can recover Yukawa signals. These plots are of the same type as those described in Section 12.2.1, but for many different systematic signal injections and with the foil-systematic signal added, subtracted, and left uncorrected.

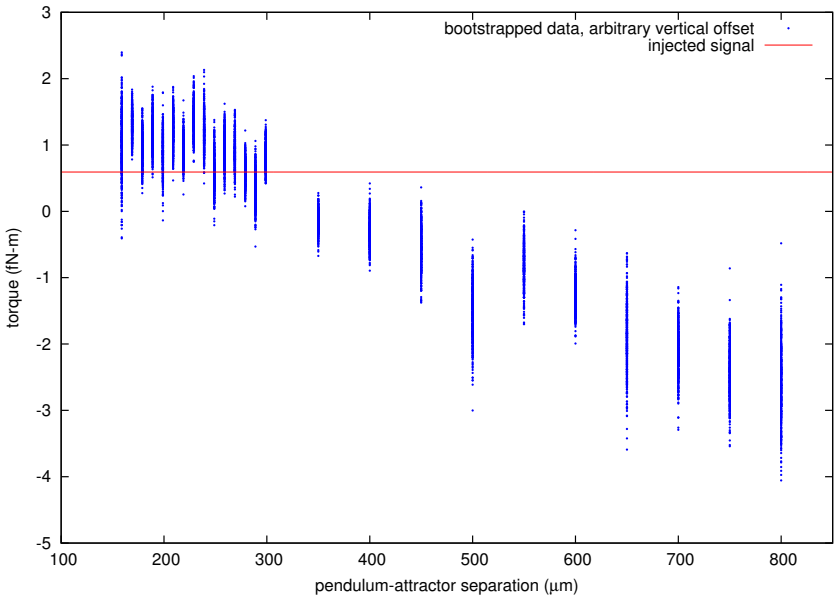

Figure B.1: Bootstrapped arbitrary fit with injected linear slope 0 N-m/m, Yukawa interaction with $\alpha = 0$, $\lambda = 200$ μm. Curves moved vertically to approximately agree over the interval [275,300] μm.

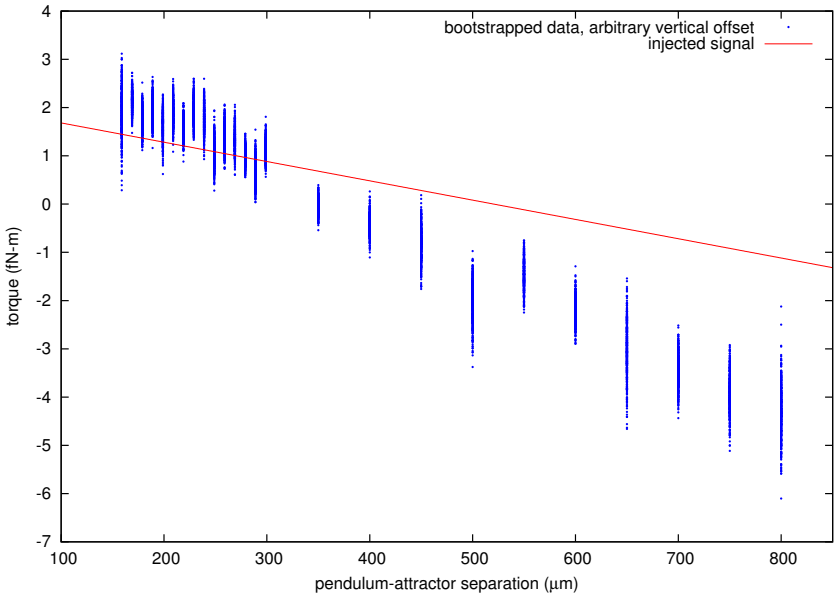

Figure B.2: Bootstrapped arbitrary fit with injected linear slope -4×10^{-12} N-m/m, Yukawa interaction with $\alpha = 0$, $\lambda = 200$ μm. Curves moved vertically to approximately agree over the interval [275,300] μm.

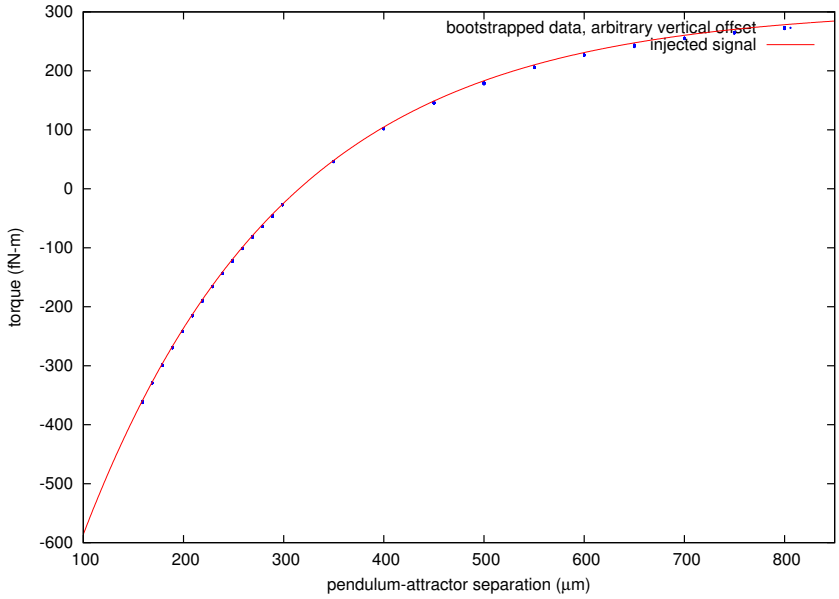

Figure B.3: Bootstrapped arbitrary fit with injected linear slope 6×10^{-12} N-m/m, Yukawa interaction with $\alpha = -100$, $\lambda = 200$ μm. Curves moved vertically to approximately agree over the interval [275,300] μm.

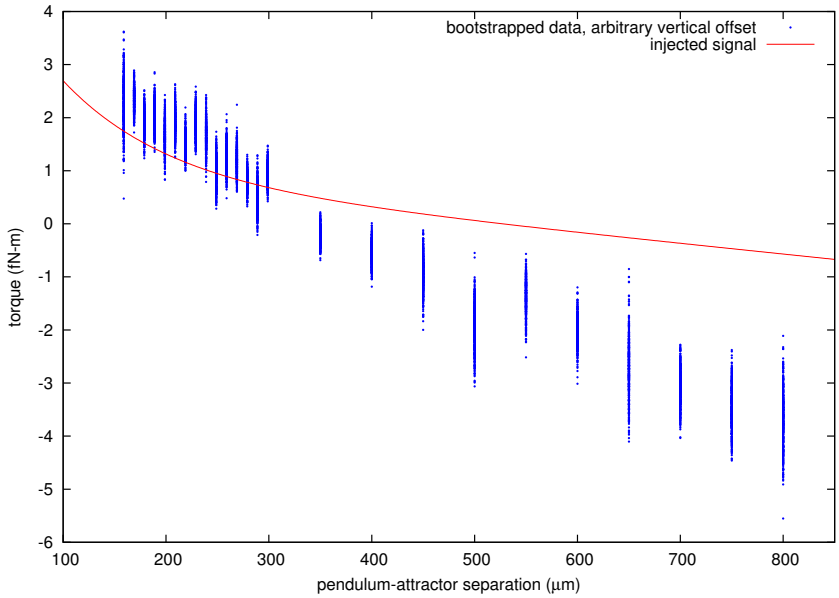

Figure B.4: Bootstrapped arbitrary fit with injected linear slope -2×10^{-12} N-m/m, Yukawa interaction with $\alpha = 1$, $\lambda = 100$ μm. Curves moved vertically to approximately agree over the interval [275,300] μm.

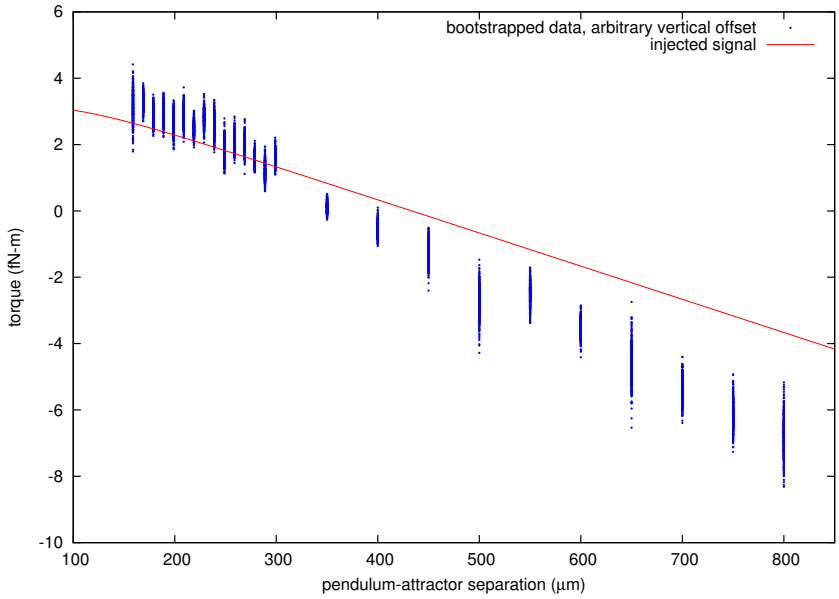

Figure B.5: Bootstrapped arbitrary fit with injected linear slope -1×10^{-11} N-m/m, Yukawa interaction with $\alpha = -1$, $\lambda = 56$ μm. Curves moved vertically to approximately agree over the interval [275,300] μm.

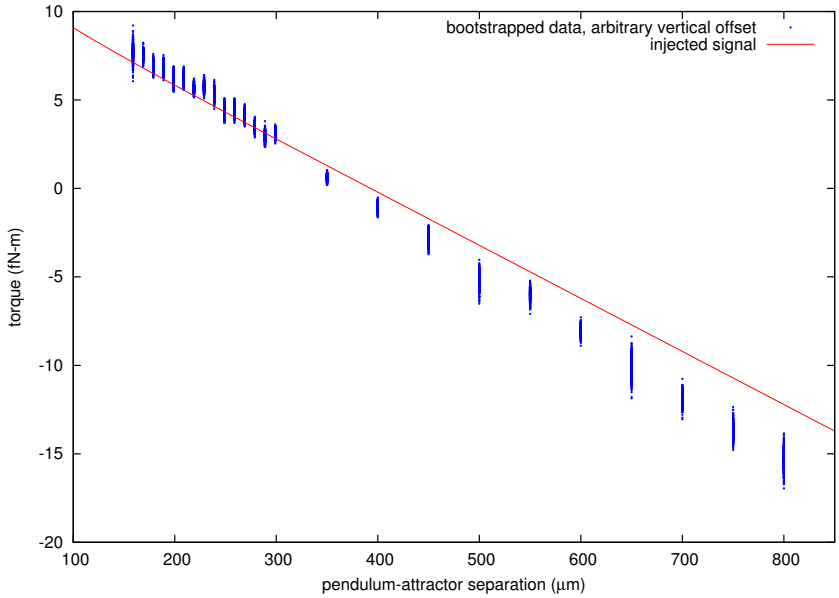

Figure B.6: Bootstrapped arbitrary fit with injected linear slope -3×10^{-11} N-m/m, Yukawa interaction with $\alpha = 1$, $\lambda = 56$ μm. Curves moved vertically to approximately agree over the interval [275,300] μm.

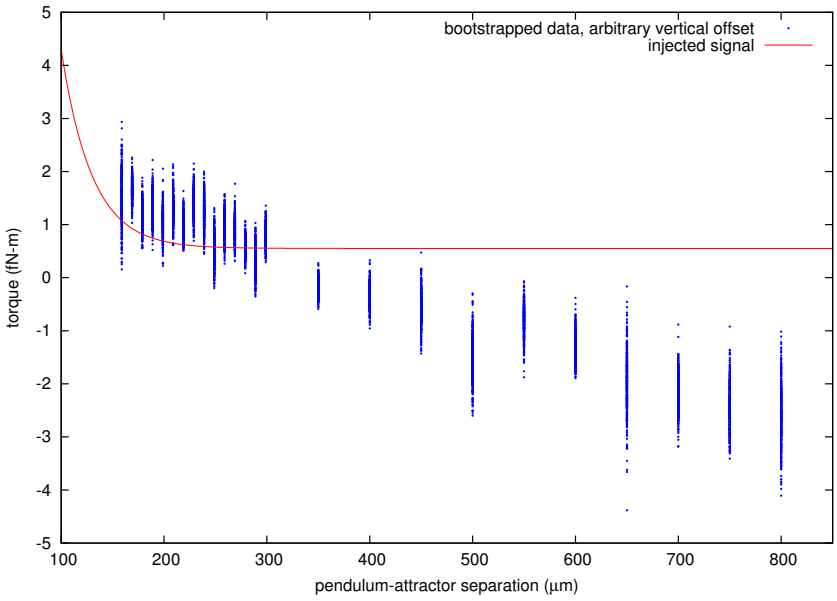

Figure B.7: Bootstrapped arbitrary fit with injected linear slope 0 N-m/m, Yukawa interaction with $\alpha = 200$, $\lambda = 30$ μm. Curves moved vertically to approximately agree over the interval [275,300] μm.

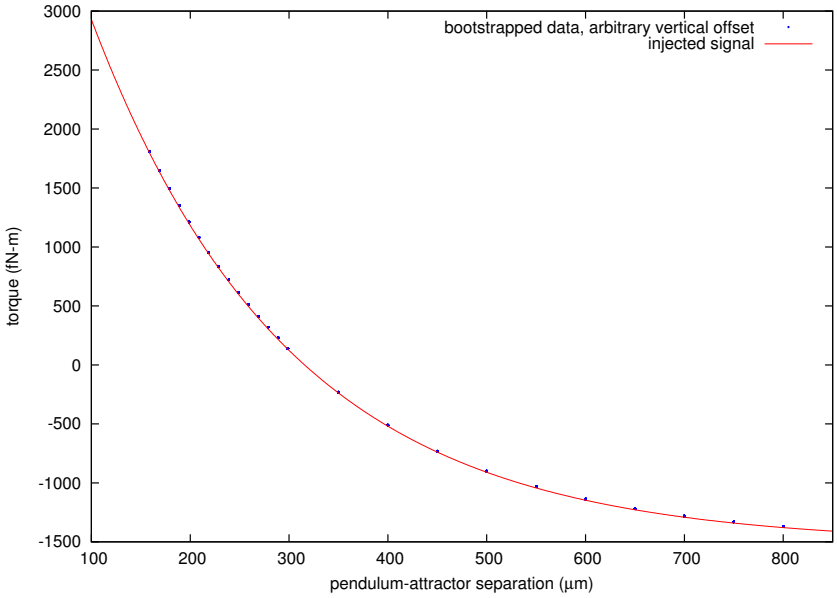

Figure B.8: Bootstrapped arbitrary fit with injected linear slope -1×10^{-11} N-m/m, Yukawa interaction with $\alpha = 500$, $\lambda = 200$ μm. Curves moved vertically to approximately agree over the interval [275,300] μm.

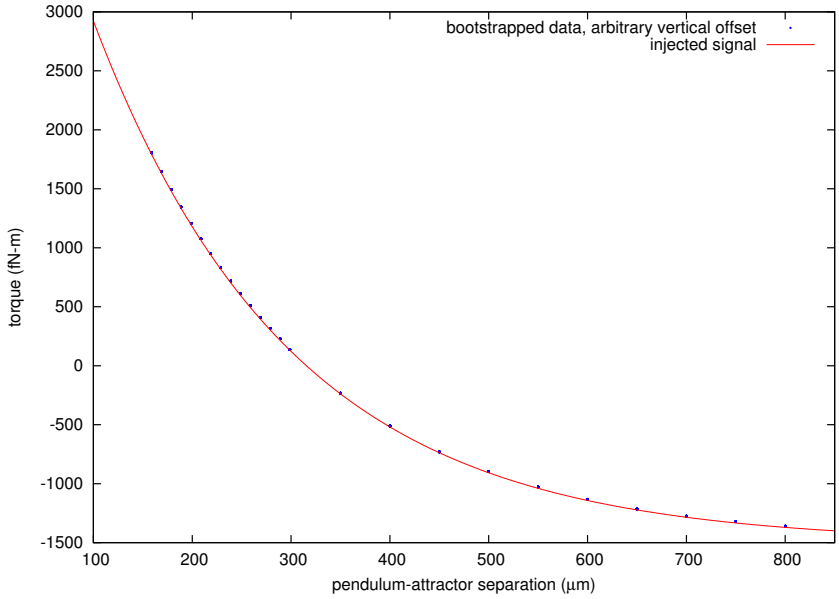

Figure B.9: Bootstrapped arbitrary fit with injected linear slope 1×10^{-11} N-m/m, Yukawa interaction with $\alpha = 500$, $\lambda = 200$ μm. Curves moved vertically to approximately agree over the interval [275,300] μm.

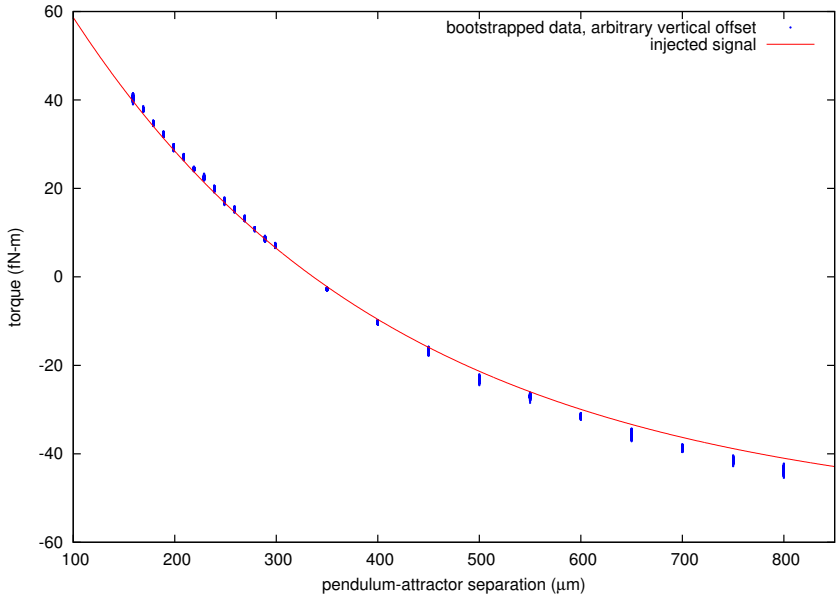

Figure B.10: Bootstrapped arbitrary fit with injected linear slope -4×10^{-12} N-m/m, Yukawa interaction with $\alpha = 6$, $\lambda = 310$ μm. Curves moved vertically to approximately agree over the interval [275,300] μm.

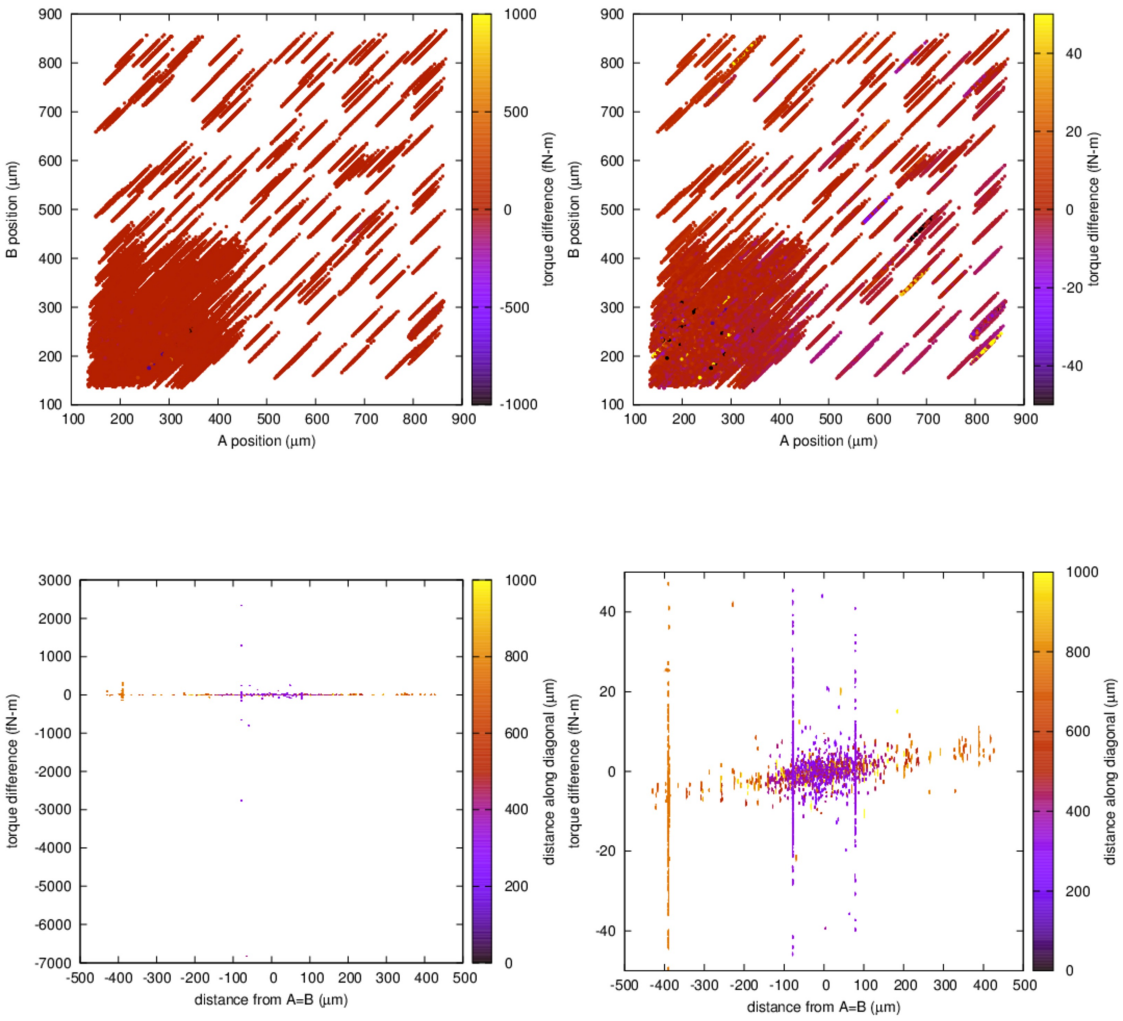

Figure B.11: Triangle plot. A signal was injected with linear slope 0 N-m/m, Yukawa interaction with $\alpha = 0$, $\lambda = 200$ μm. The foil systematic was subtracted from the original data

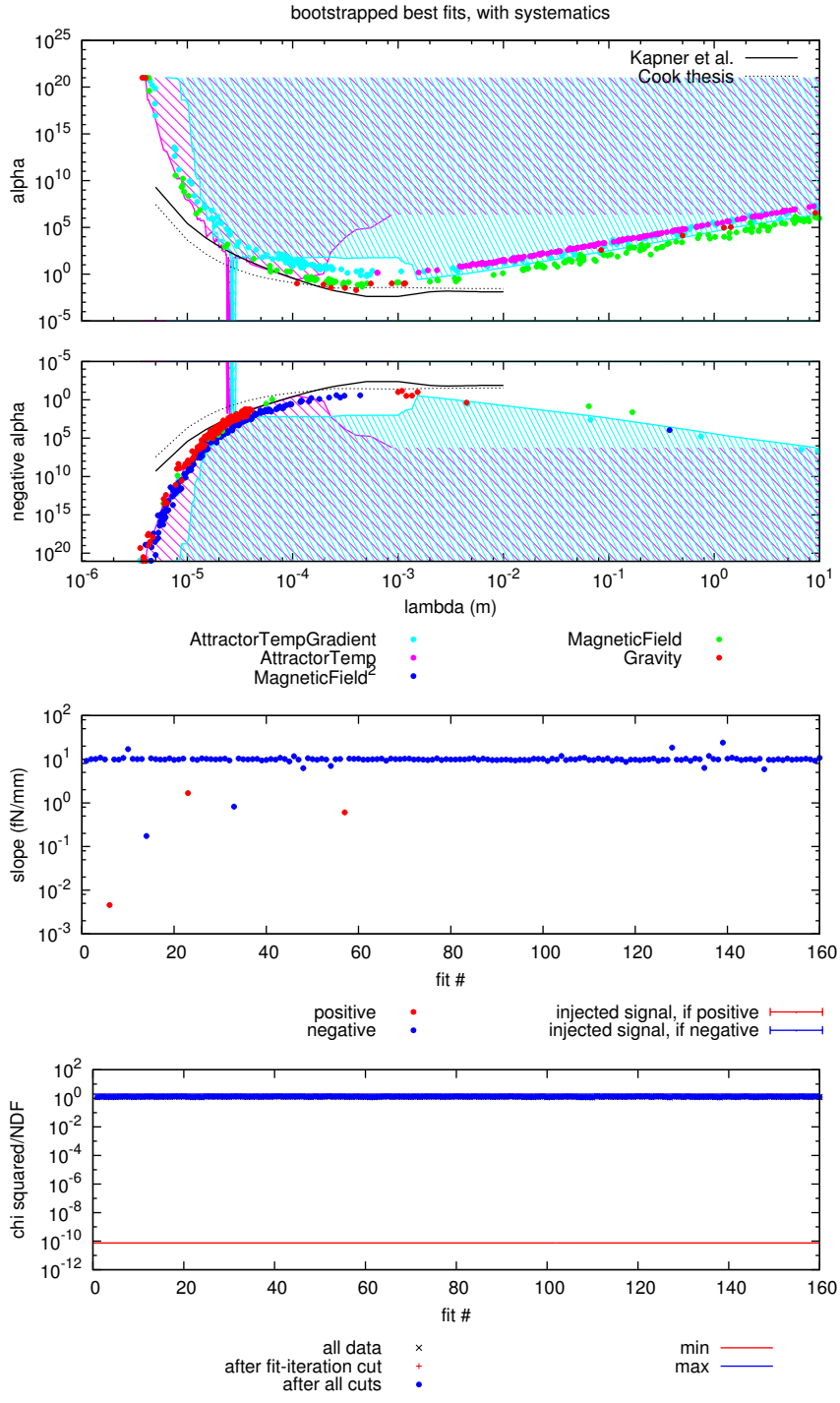

Figure B.12: Yukawa Plot, all systematics. A signal was injected with linear slope 0 N-m/m, Yukawa interaction with $\alpha = 0$, $\lambda = 200$ μm. The foil systematic was subtracted from the original data

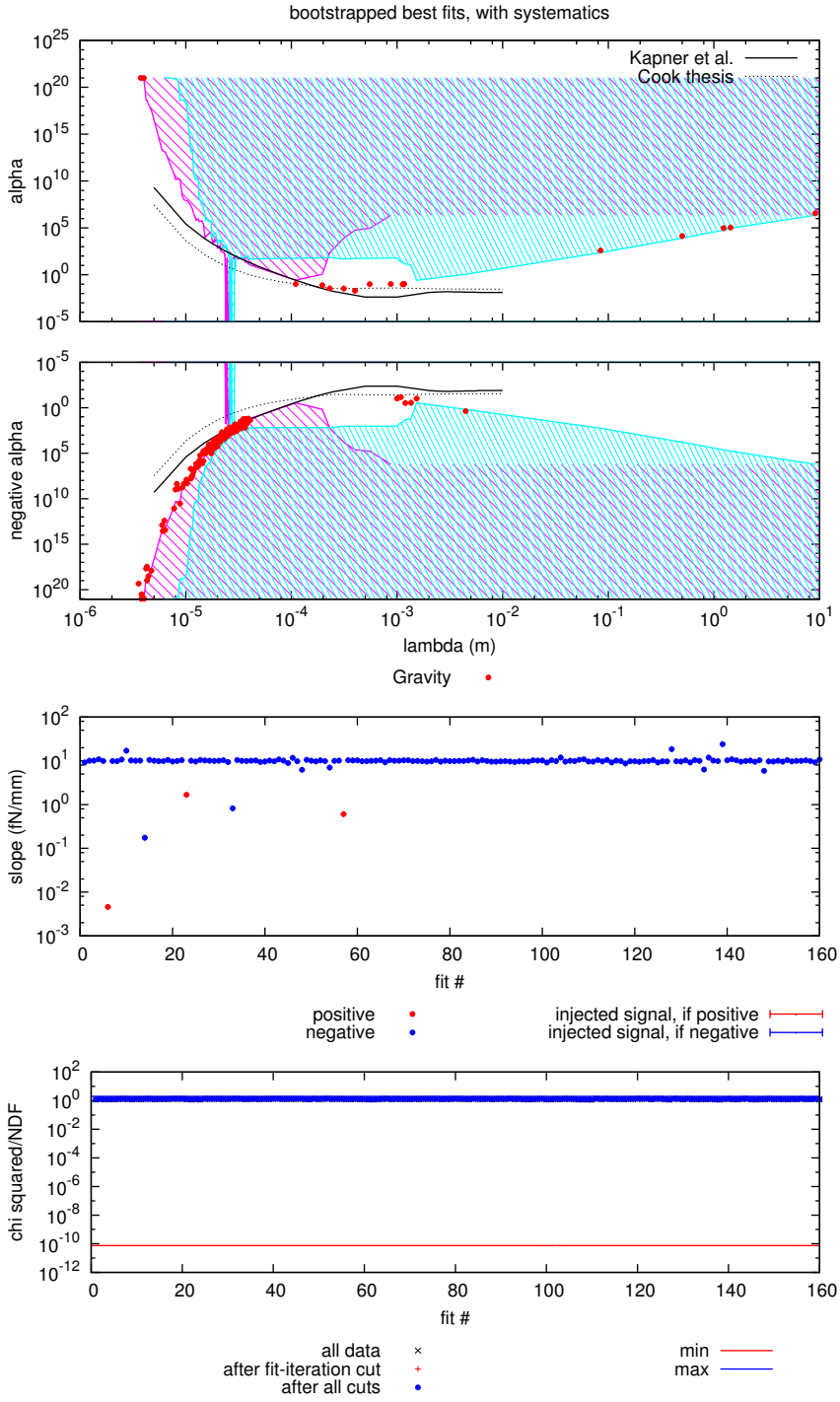

Figure B.13: Yukawa Plot, gravity only. A signal was injected with linear slope 0 N-m/m, Yukawa interaction with $\alpha = 0$, $\lambda = 200$ μm. The foil systematic was subtracted from the original data

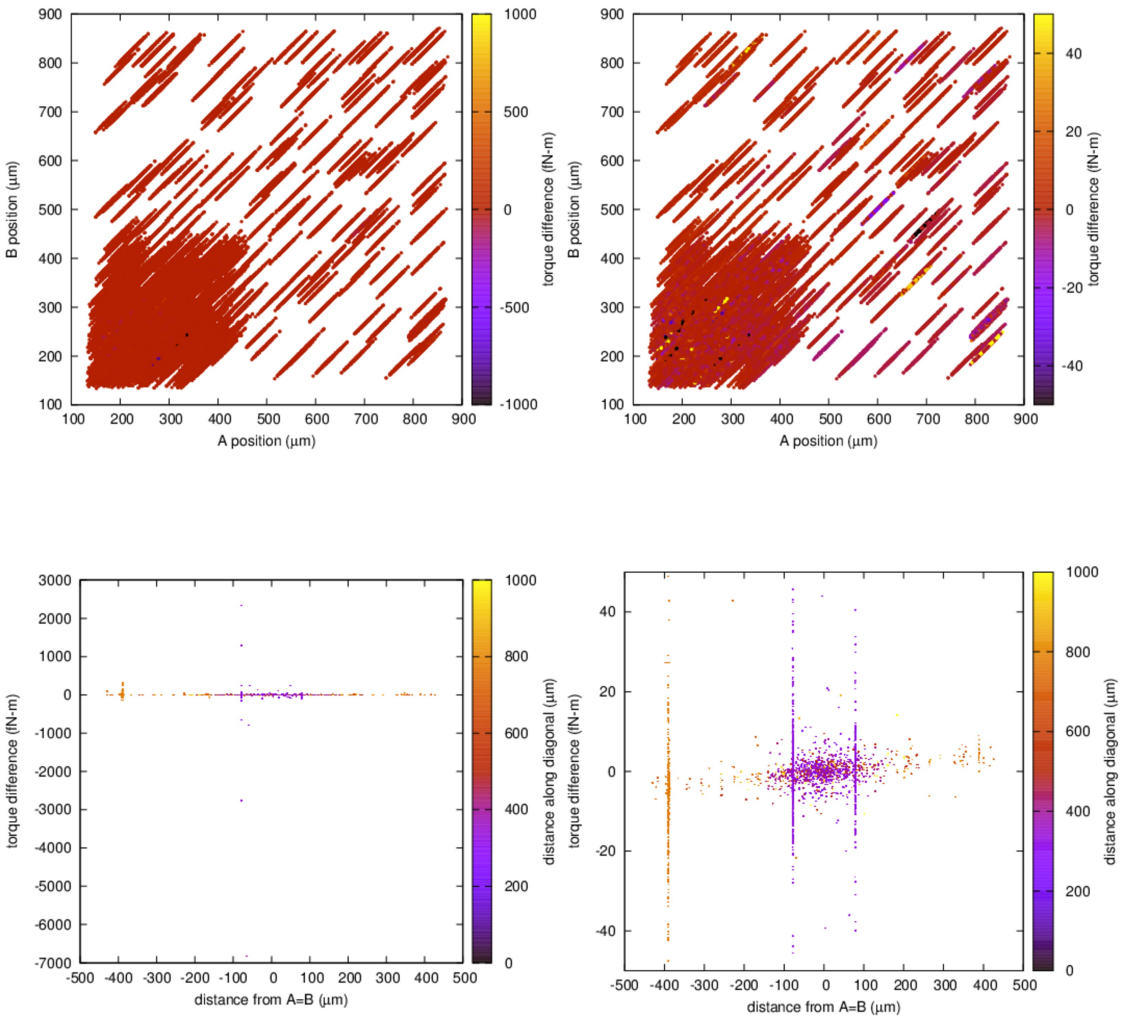

Figure B.14: Triangle plot. A signal was injected with linear slope 0 N-m/m, Yukawa interaction with $\alpha = 0$, $\lambda = 200$ μm. The foil systematic was not corrected.

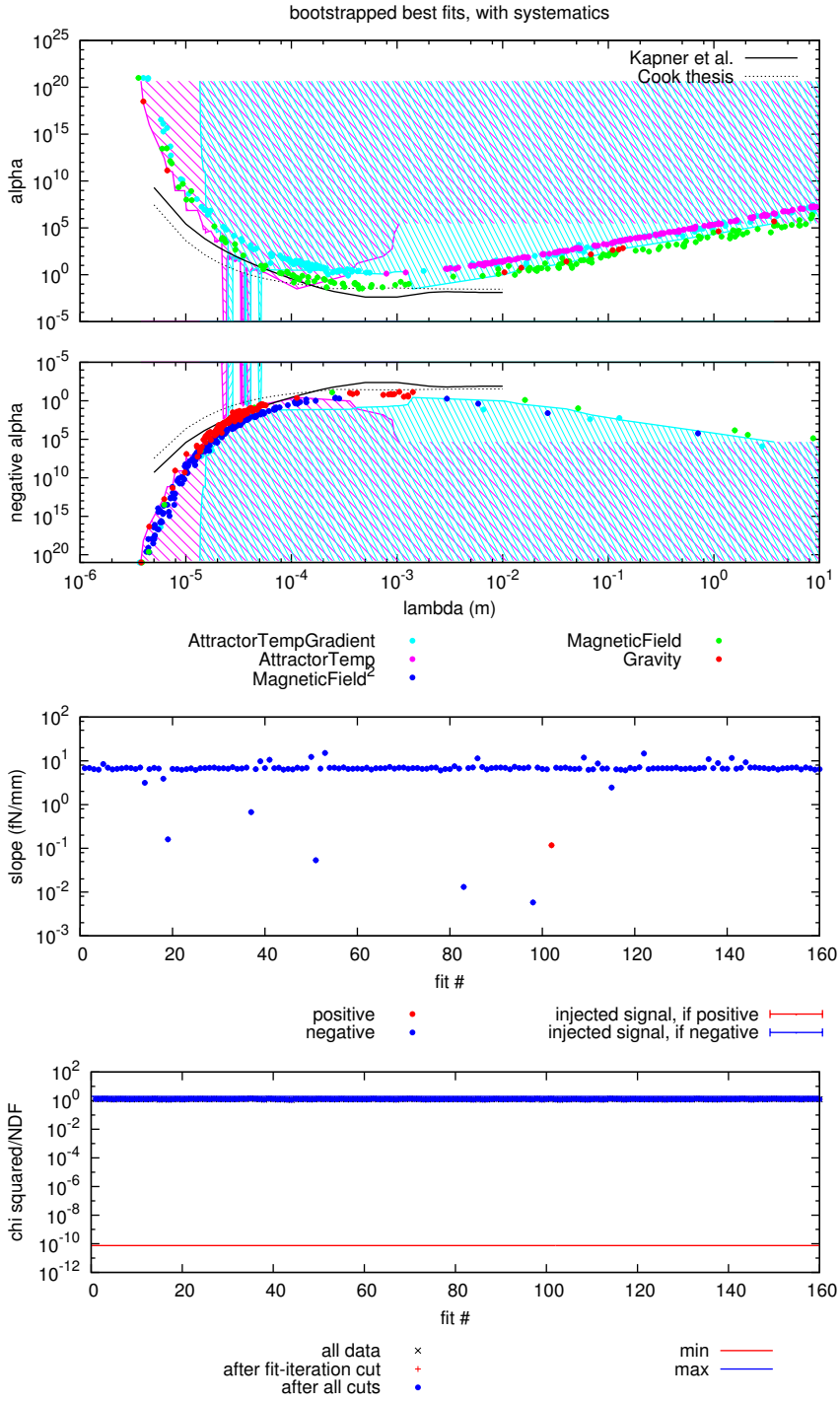

Figure B.15: Yukawa Plot, all systematics. A signal was injected with linear slope 0 N-m/m, Yukawa interaction with $\alpha = 0$, $\lambda = 200$ μm. The foil systematic was not corrected.

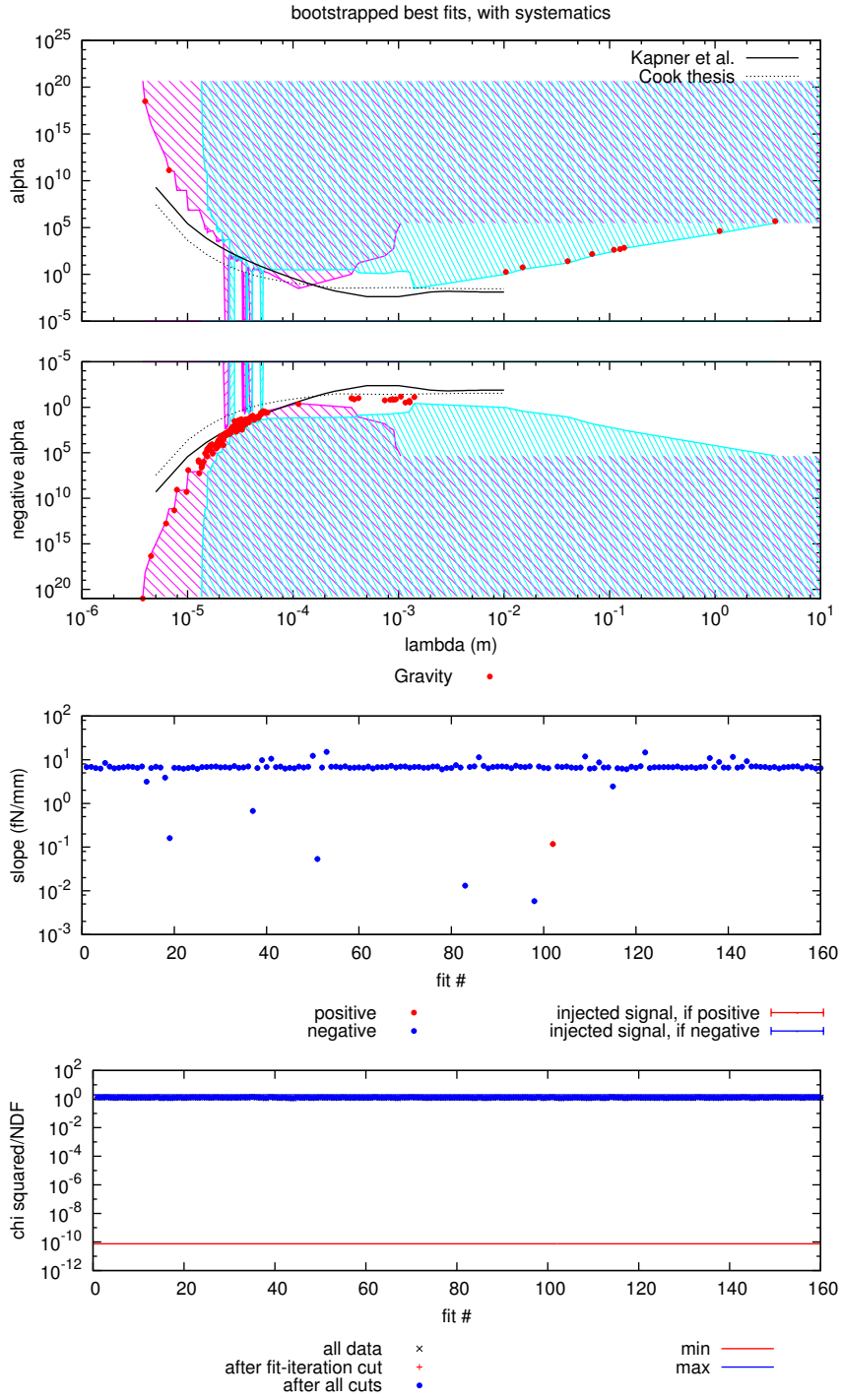

Figure B.16: Yukawa Plot, gravity only. A signal was injected with linear slope 0 N-m/m, Yukawa interaction with $\alpha = 0$, $\lambda = 200$ μm. The foil systematic was not corrected.

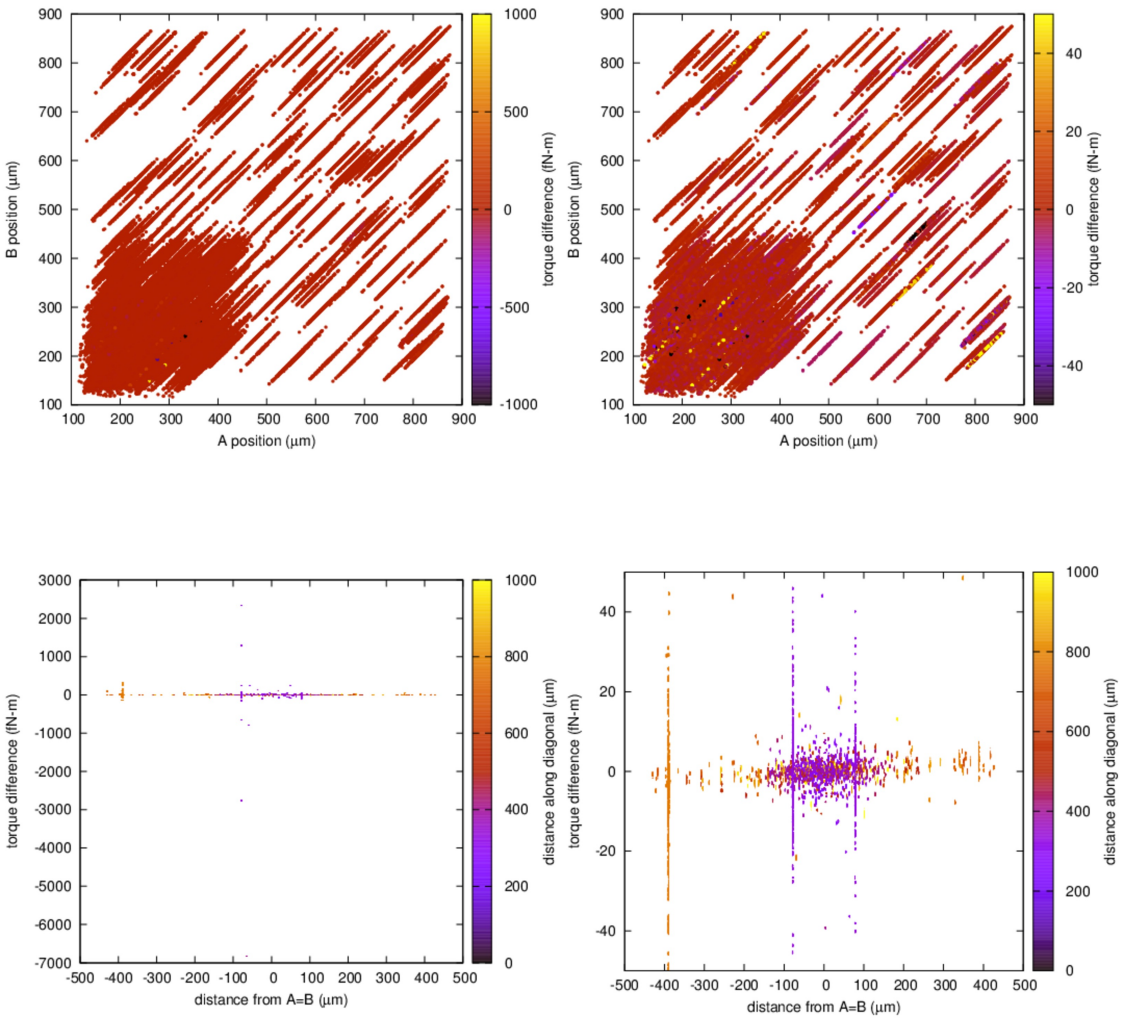

Figure B.17: Triangle plot. A signal was injected with linear slope 0 N-m/m, Yukawa interaction with $\alpha = 0$, $\lambda = 200$ μm. The foil systematic was added to the original data.

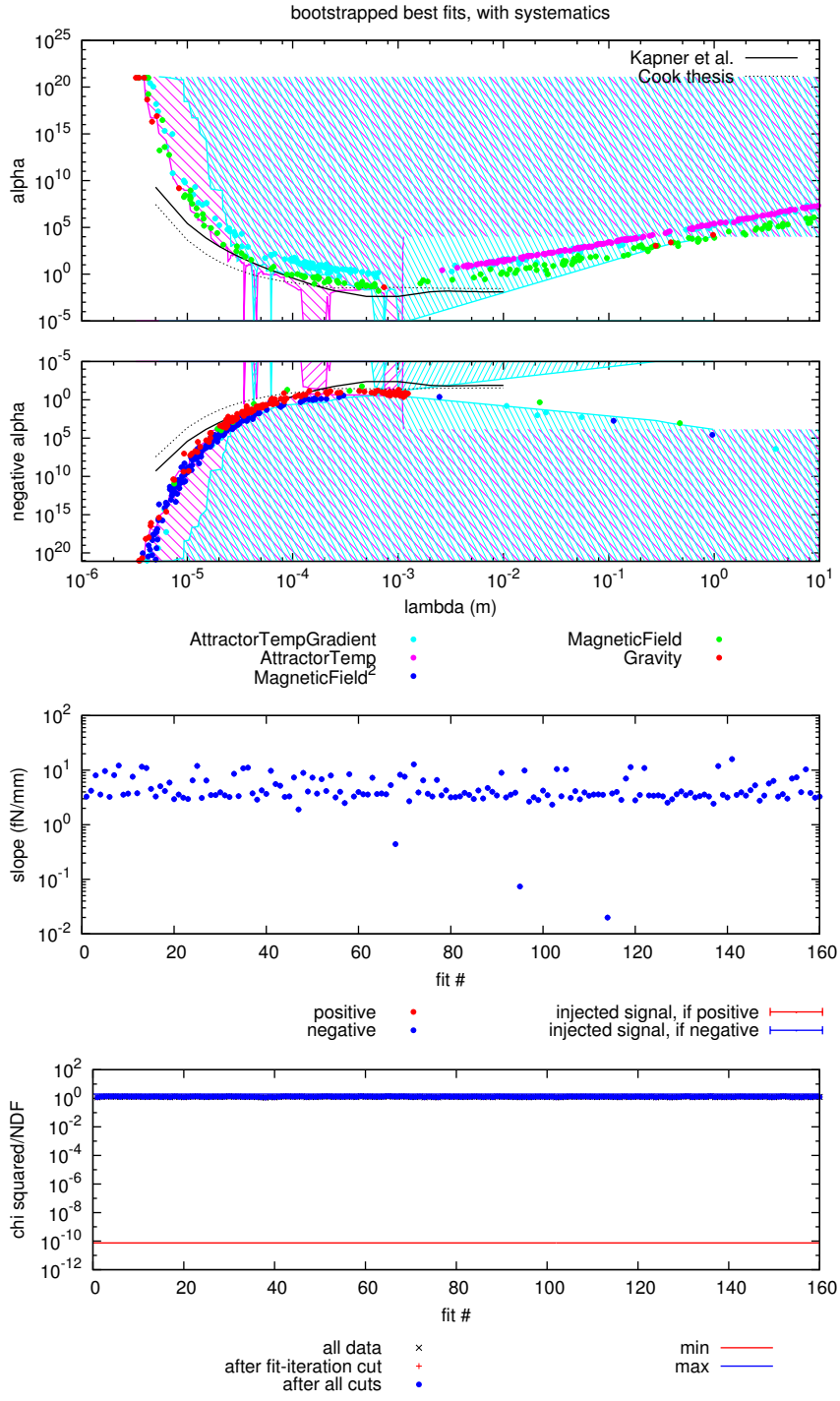

Figure B.18: Yukawa Plot, all systematics. A signal was injected with linear slope 0 N-m/m, Yukawa interaction with $\alpha = 0$, $\lambda = 200$ μm. The foil systematic was added to the original data.

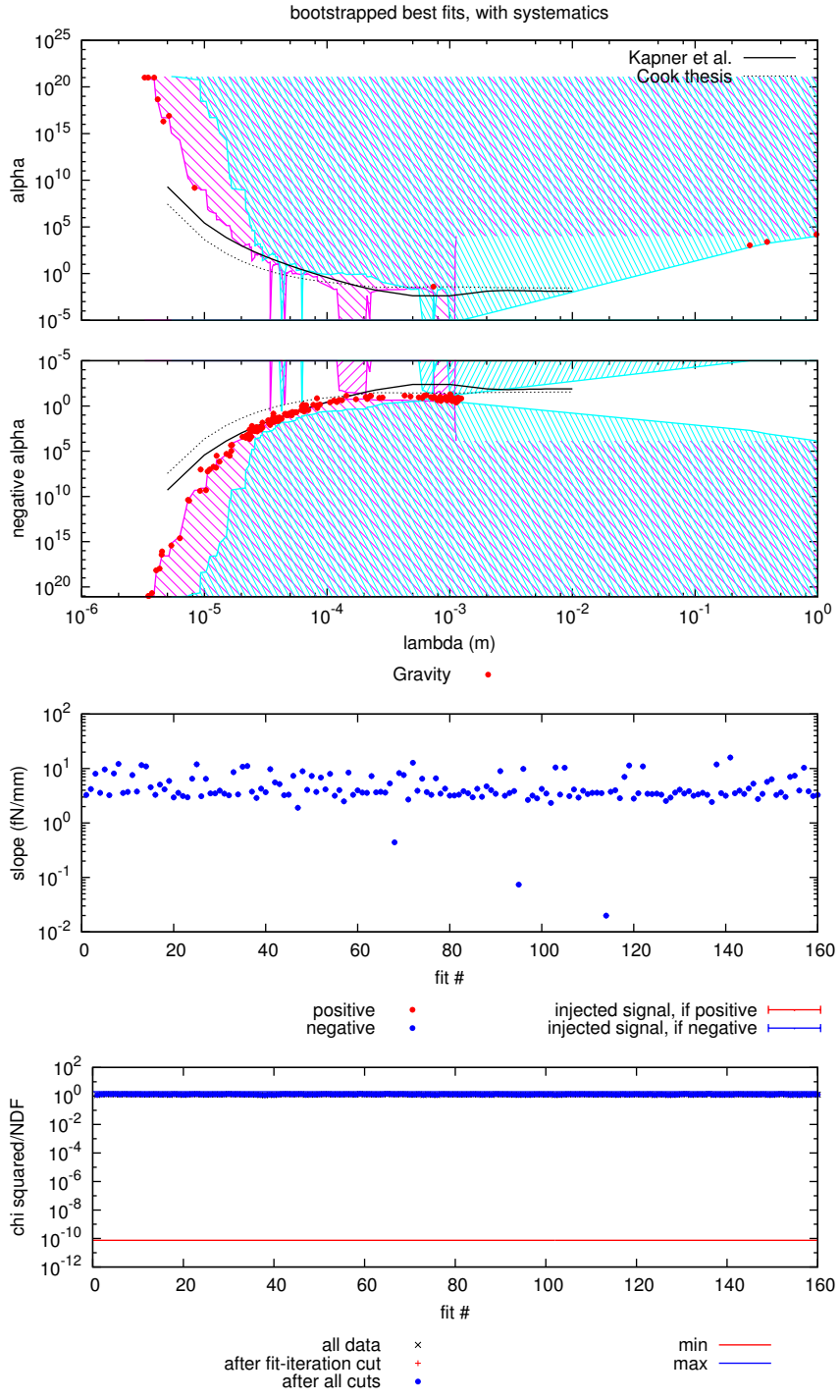

Figure B.19: Yukawa Plot, gravity only. A signal was injected with linear slope 0 N-m/m, Yukawa interaction with $\alpha = 0$, $\lambda = 200$ μm. The foil systematic was added to the original data.

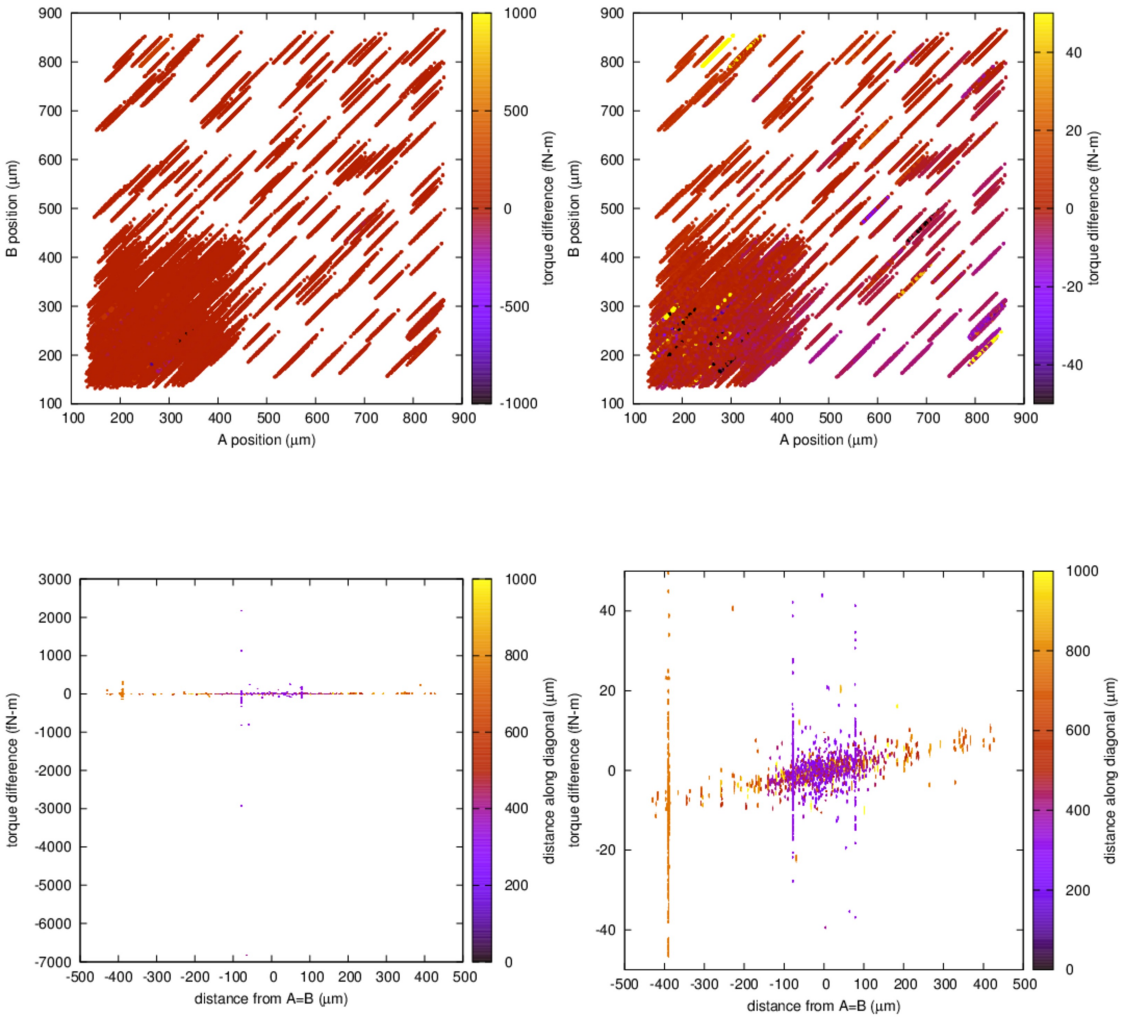

Figure B.20: Triangle plot. A signal was injected with linear slope -4×10^{-12} N-m/m, Yukawa interaction with $\alpha = 0$, $\lambda = 200$ μm. The foil systematic was subtracted from the original data

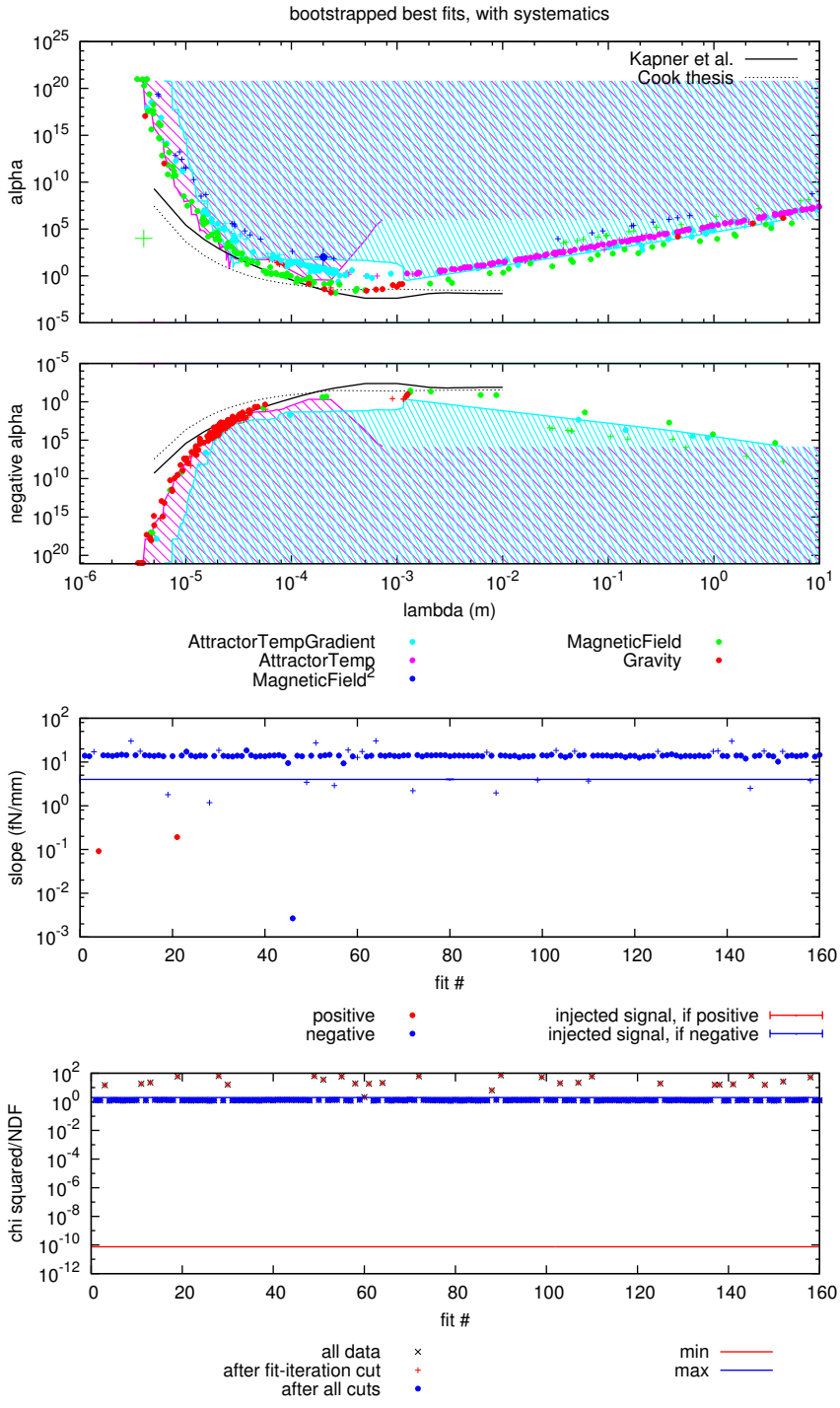

Figure B.21: Yukawa Plot, all systematics. A signal was injected with linear slope -4×10^{-12} N-m/m, Yukawa interaction with $\alpha = 0$, $\lambda = 200$ μm. The foil systematic was subtracted from the original data

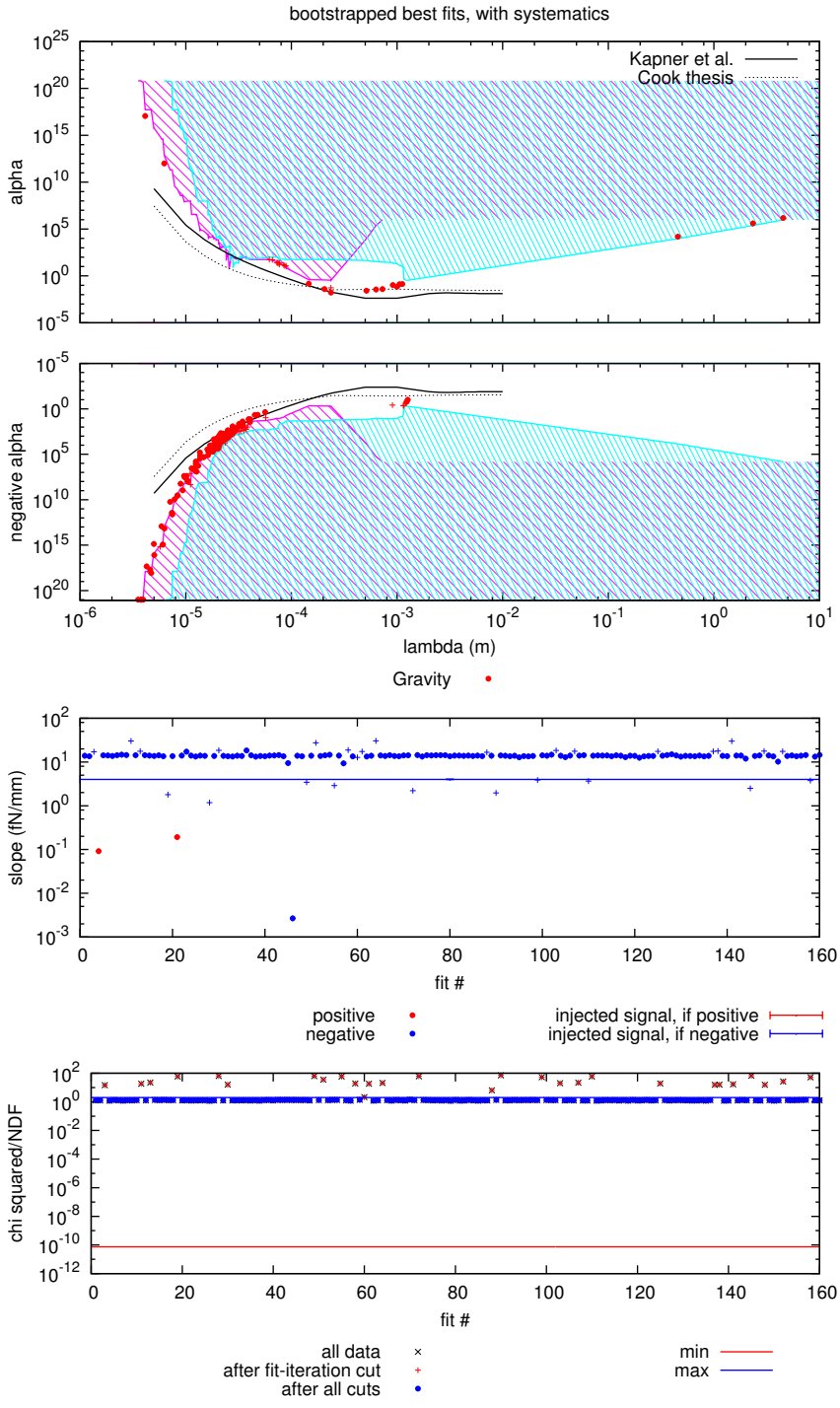

Figure B.22: Yukawa Plot, gravity only. A signal was injected with linear slope -4×10^{-12} N-m/m, Yukawa interaction with $\alpha = 0$, $\lambda = 200$ μm. The foil systematic was subtracted from the original data

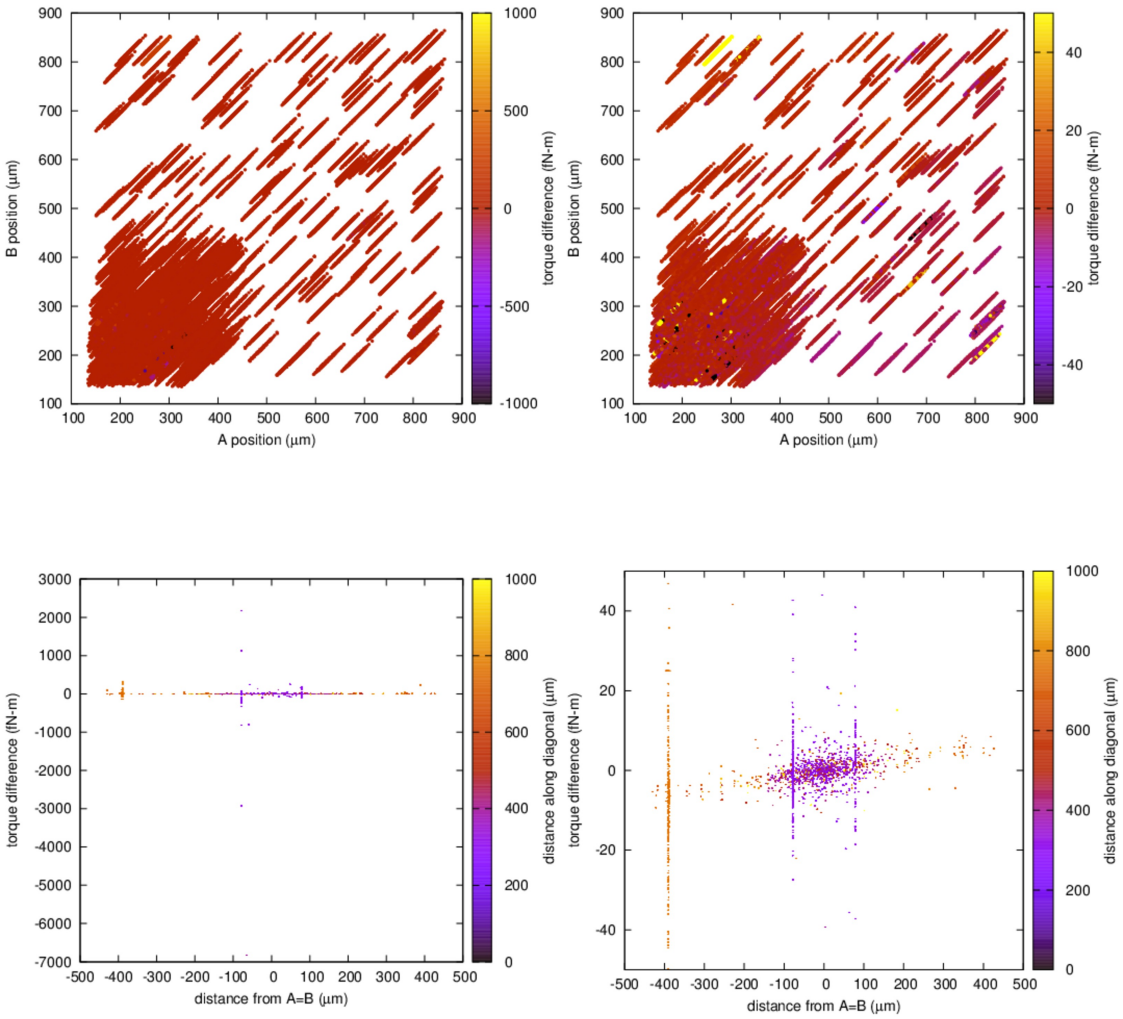

Figure B.23: Triangle plot. A signal was injected with linear slope -4×10^{-12} N-m/m, Yukawa interaction with $\alpha = 0$, $\lambda = 200$ μm. The foil systematic was not corrected.

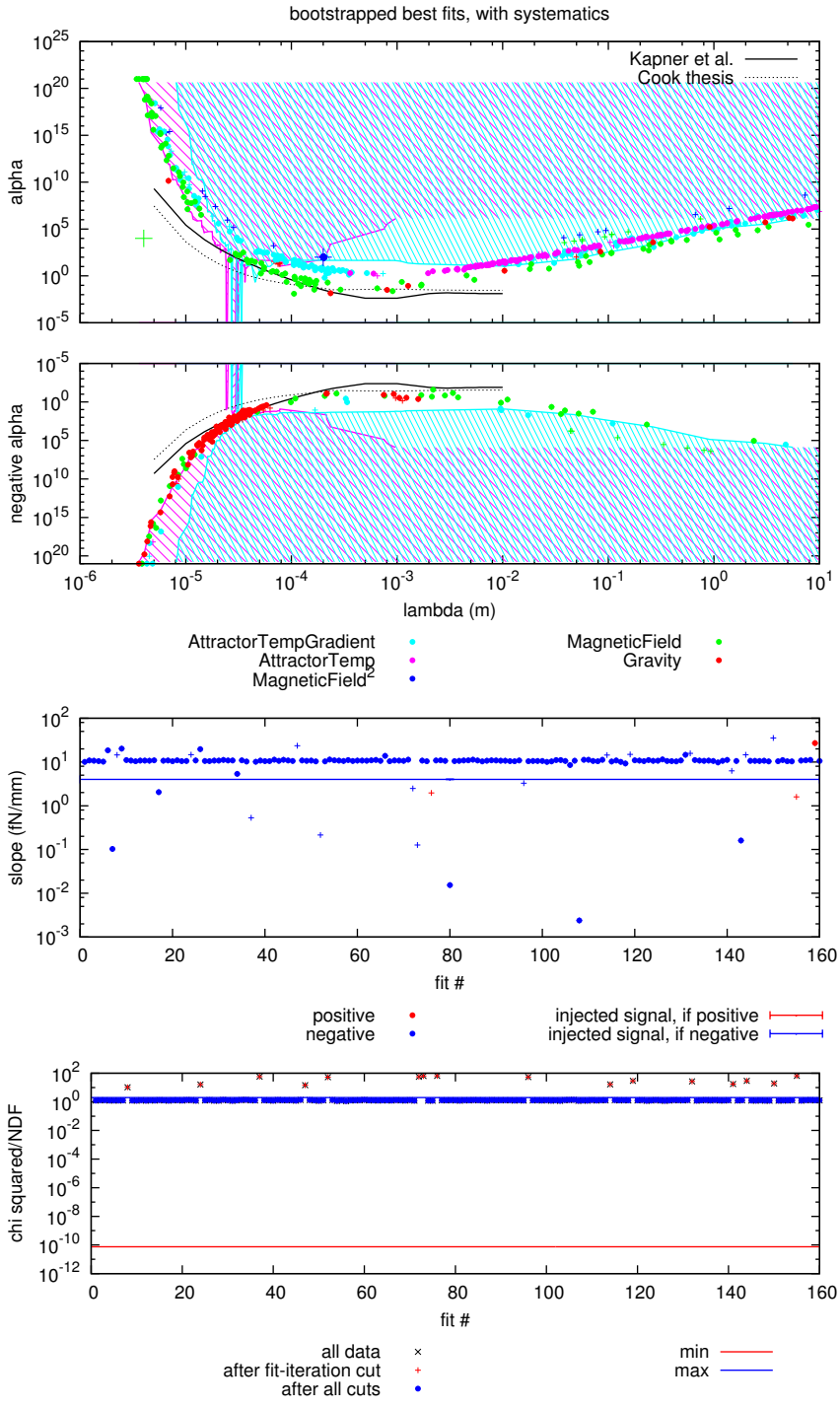

Figure B.24: Yukawa Plot, all systematics. A signal was injected with linear slope -4×10^{-12} N-m/m, Yukawa interaction with $\alpha = 0$, $\lambda = 200$ μm. The foil systematic was not corrected.

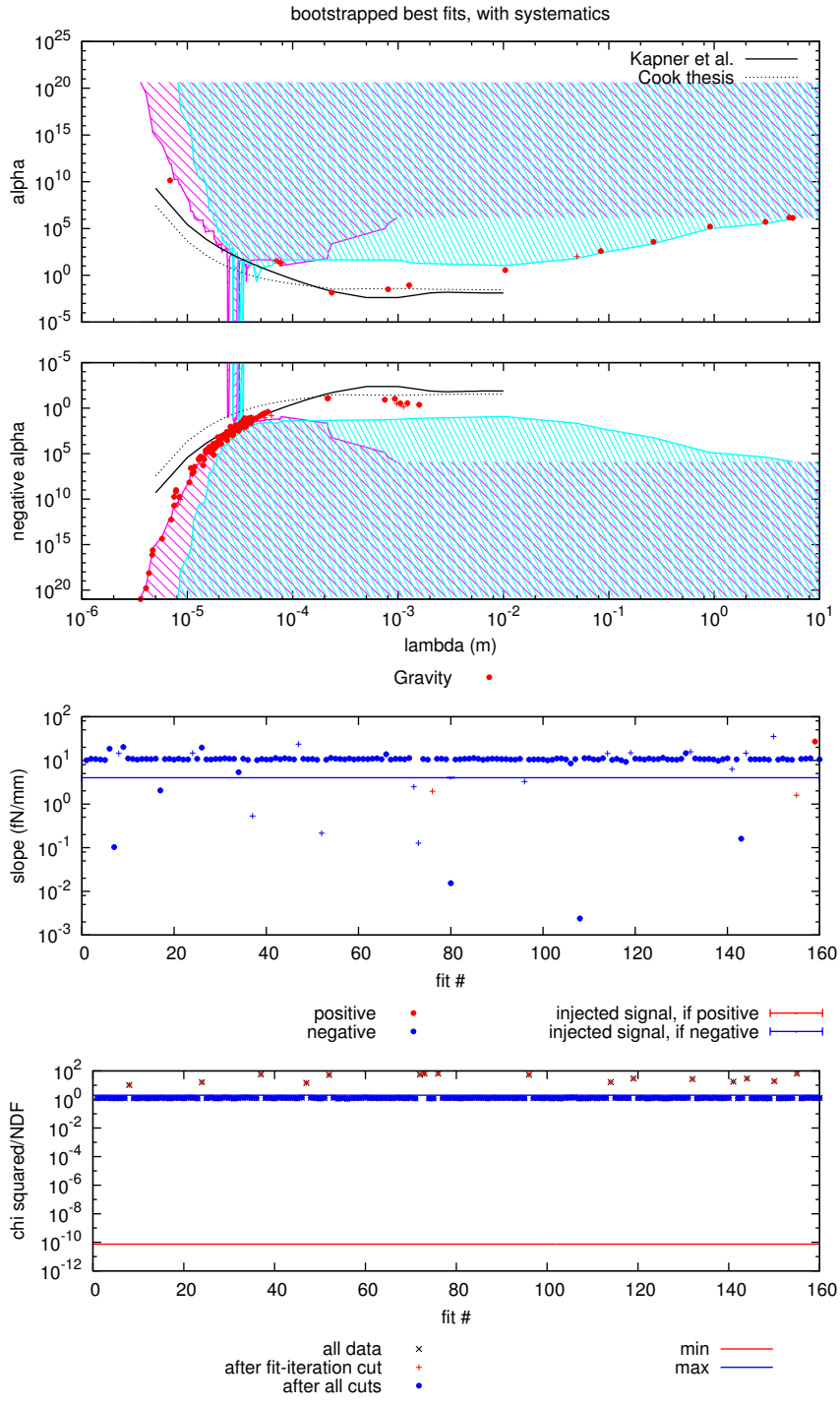

Figure B.25: Yukawa Plot, gravity only. A signal was injected with linear slope -4×10^{-12} N-m/m, Yukawa interaction with $\alpha = 0$, $\lambda = 200$ μm. The foil systematic was not corrected.

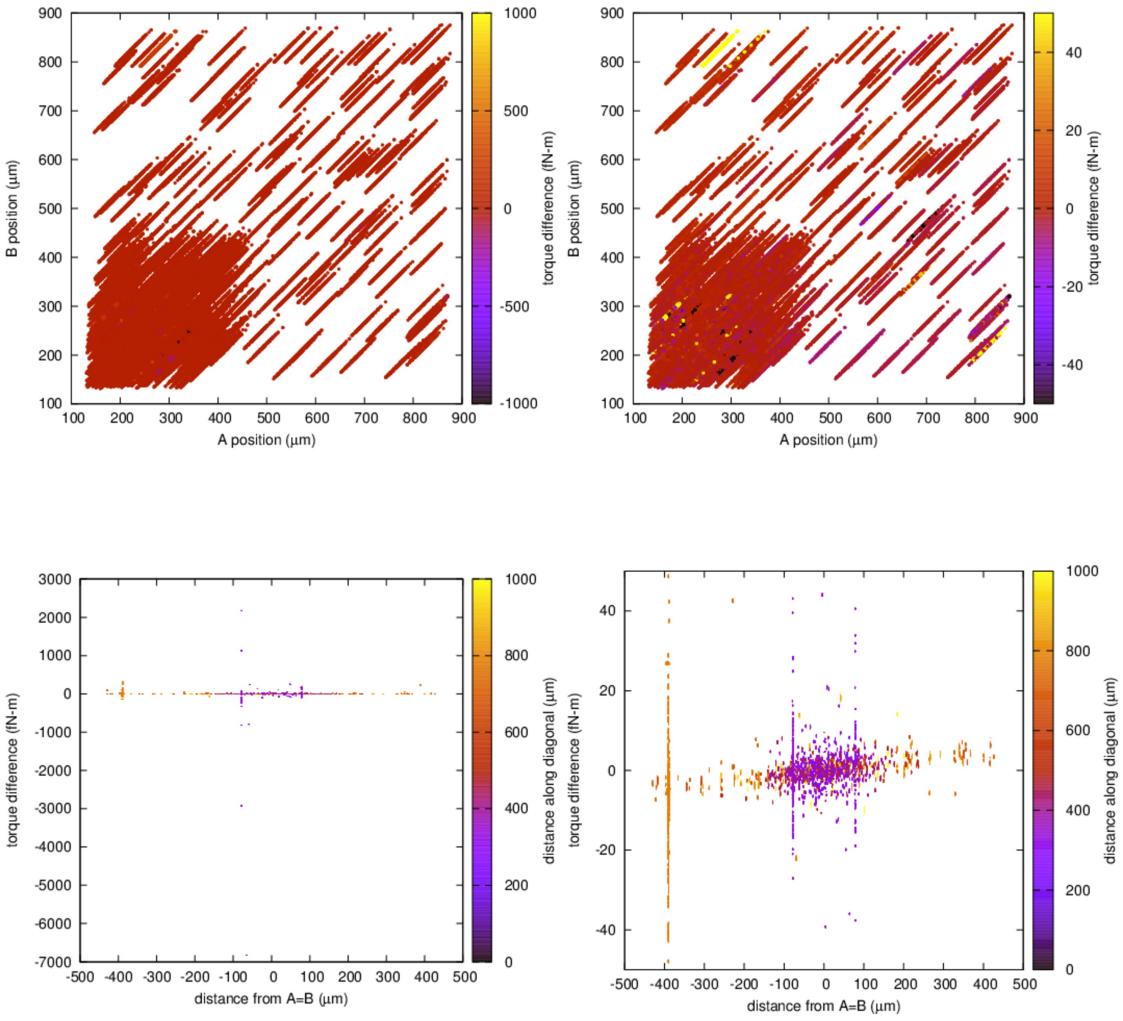

Figure B.26: Triangle plot. A signal was injected with linear slope -4×10^{-12} N-m/m, Yukawa interaction with $\alpha = 0$, $\lambda = 200$ μm. The foil systematic was added to the original data.

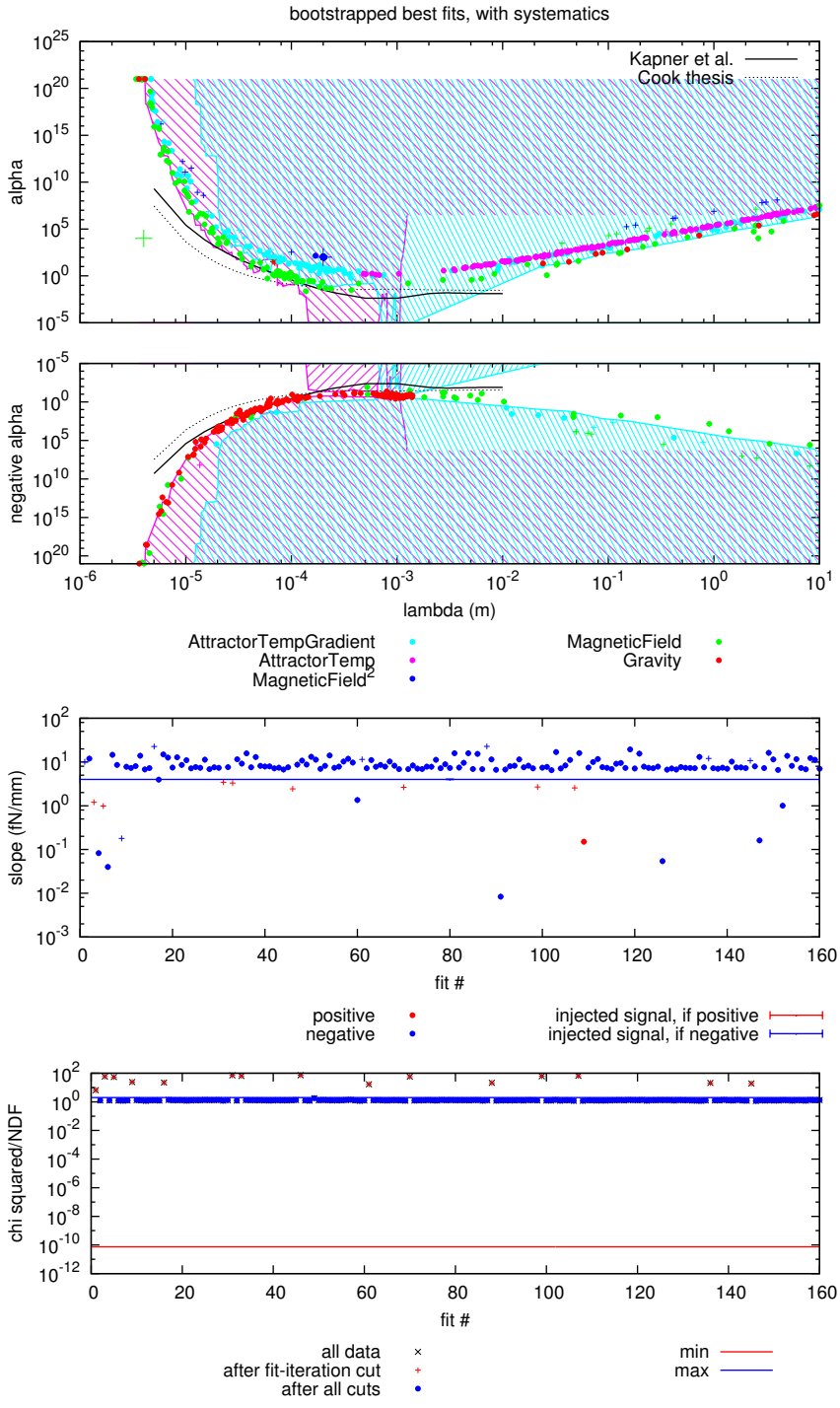

Figure B.27: Yukawa Plot, all systematics. A signal was injected with linear slope -4×10^{-12} N-m/m, Yukawa interaction with $\alpha = 0$, $\lambda = 200$ μm. The foil systematic was added to the original data.

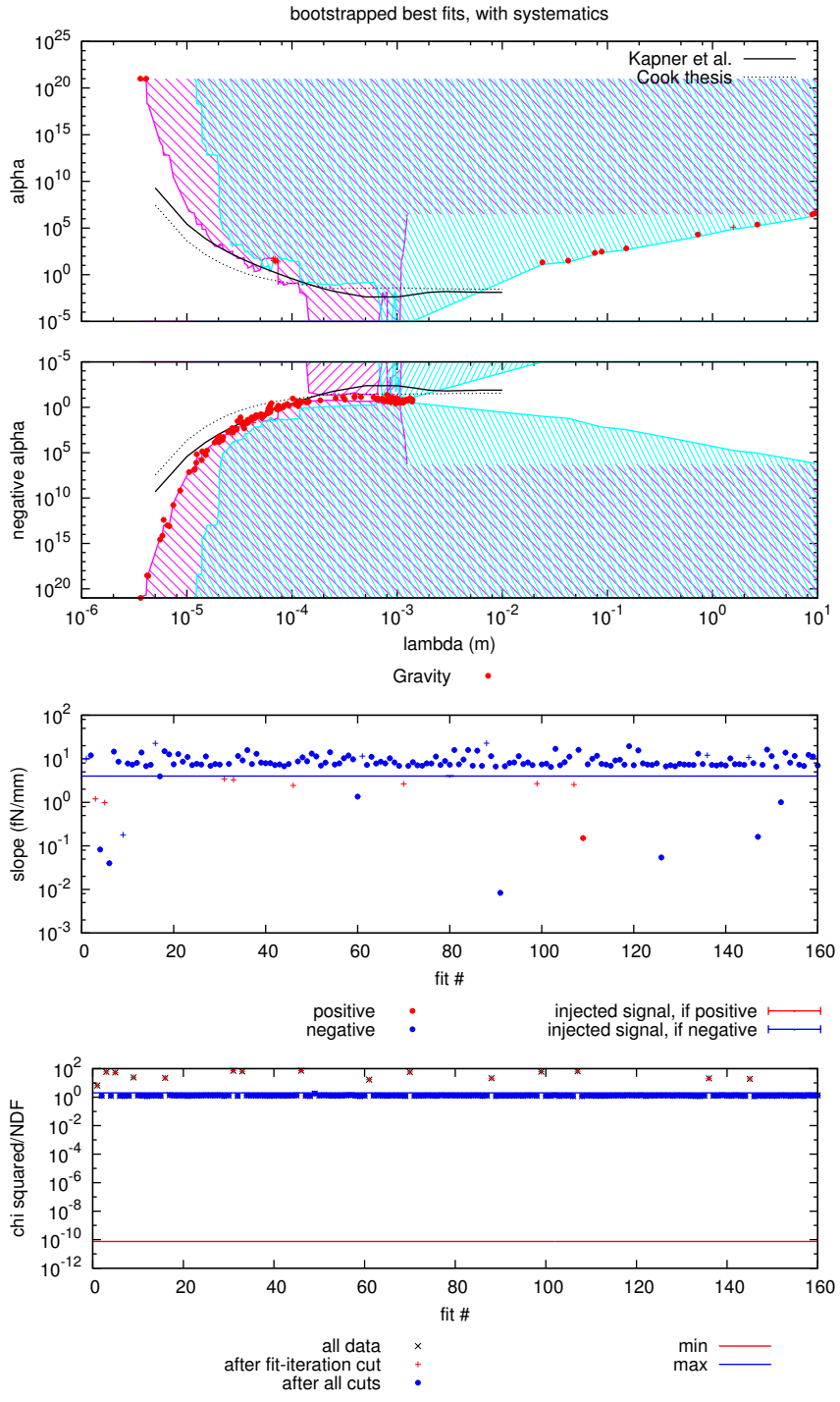

Figure B.28: Yukawa Plot, gravity only. A signal was injected with linear slope -4×10^{-12} N-m/m, Yukawa interaction with $\alpha = 0$, $\lambda = 200$ μm. The foil systematic was added to the original data.

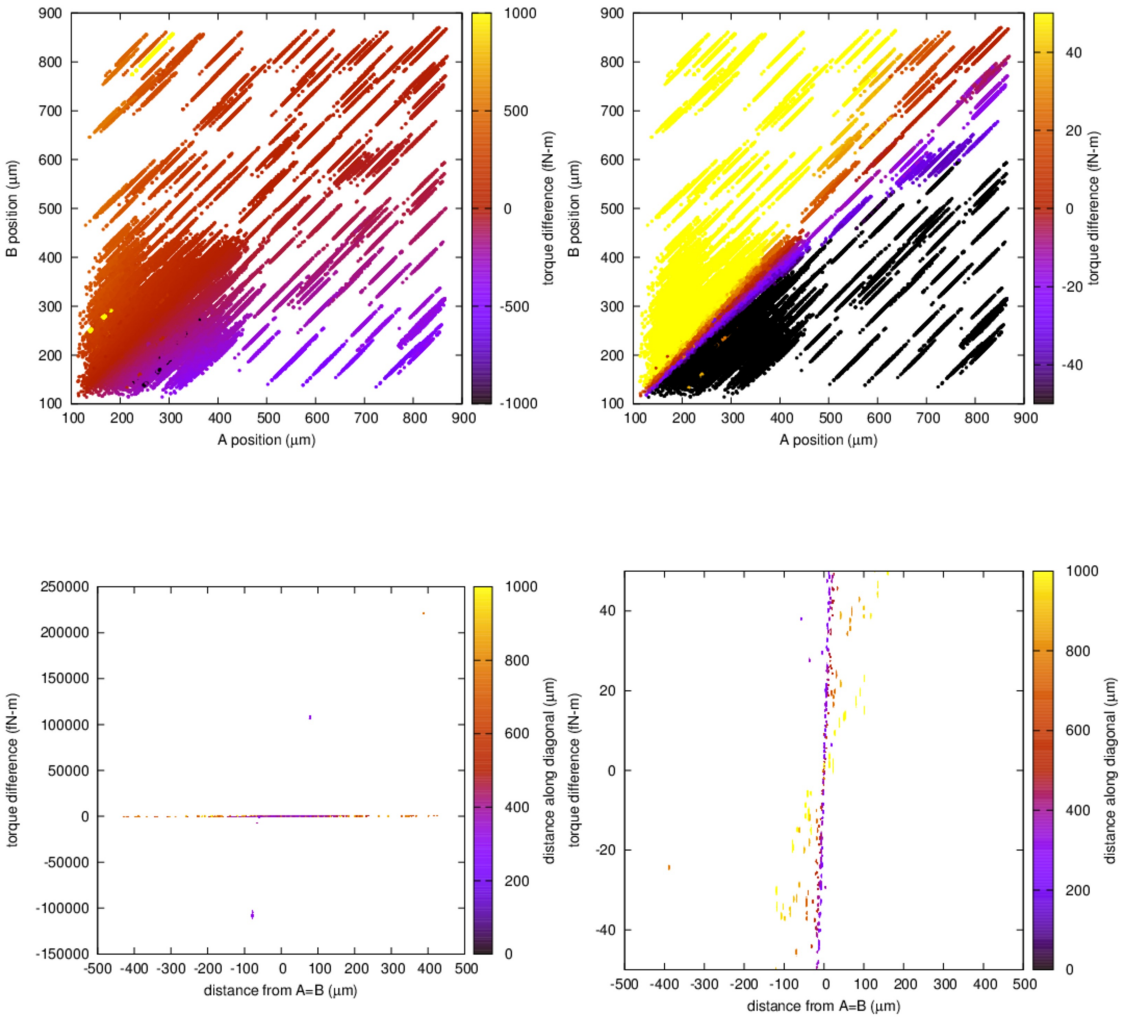

Figure B.29: Triangle plot. A signal was injected with linear slope 1×10^{-11} N-m/m, Yukawa interaction with $\alpha = 100$, $\lambda = 200$ μm. The foil systematic was subtracted from the original data

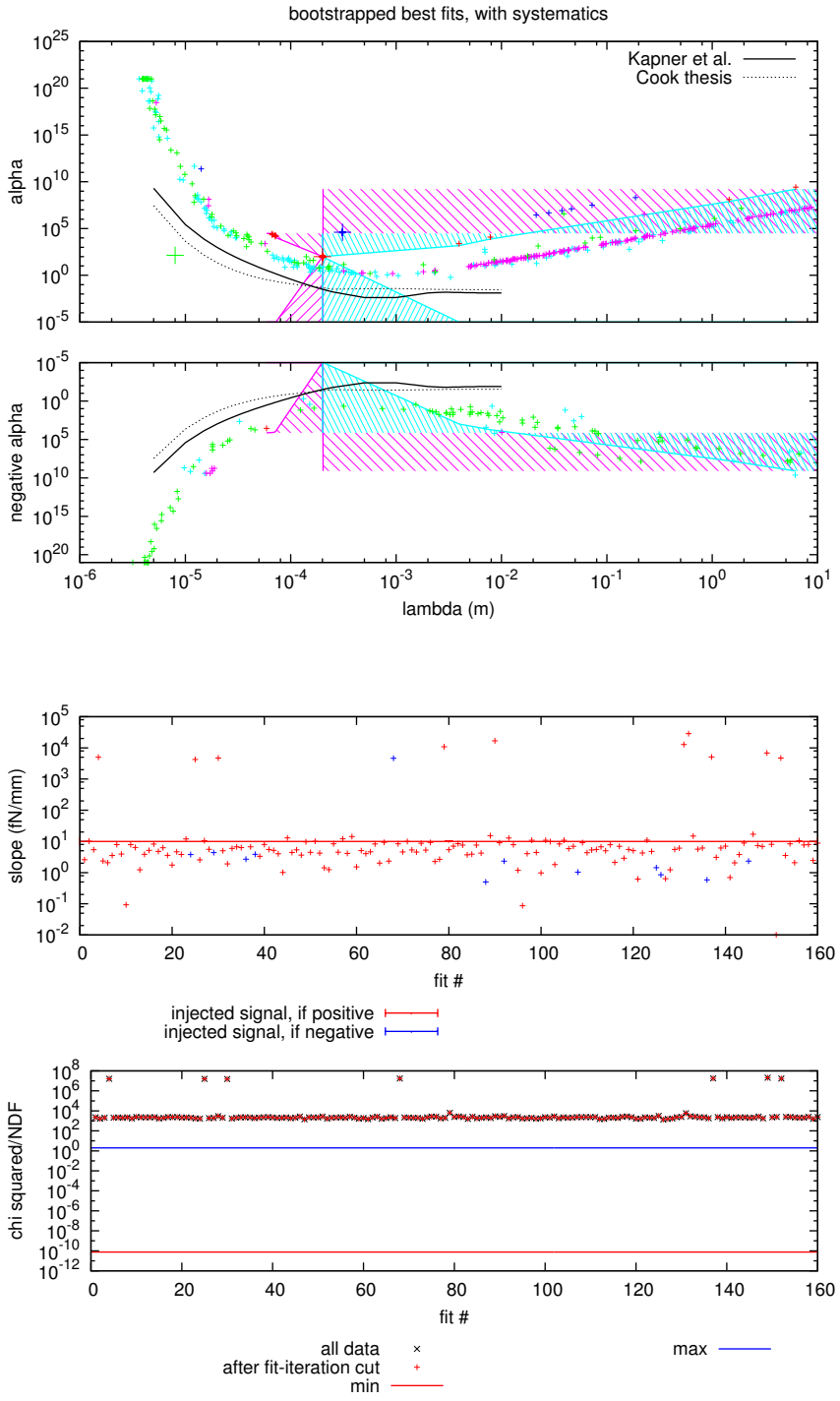

Figure B.30: Yukawa Plot, all systematics. A signal was injected with linear slope 1×10^{-11} N-m/m, Yukawa interaction with $\alpha = 100$, $\lambda = 200$ μm. The foil systematic was subtracted from the original data

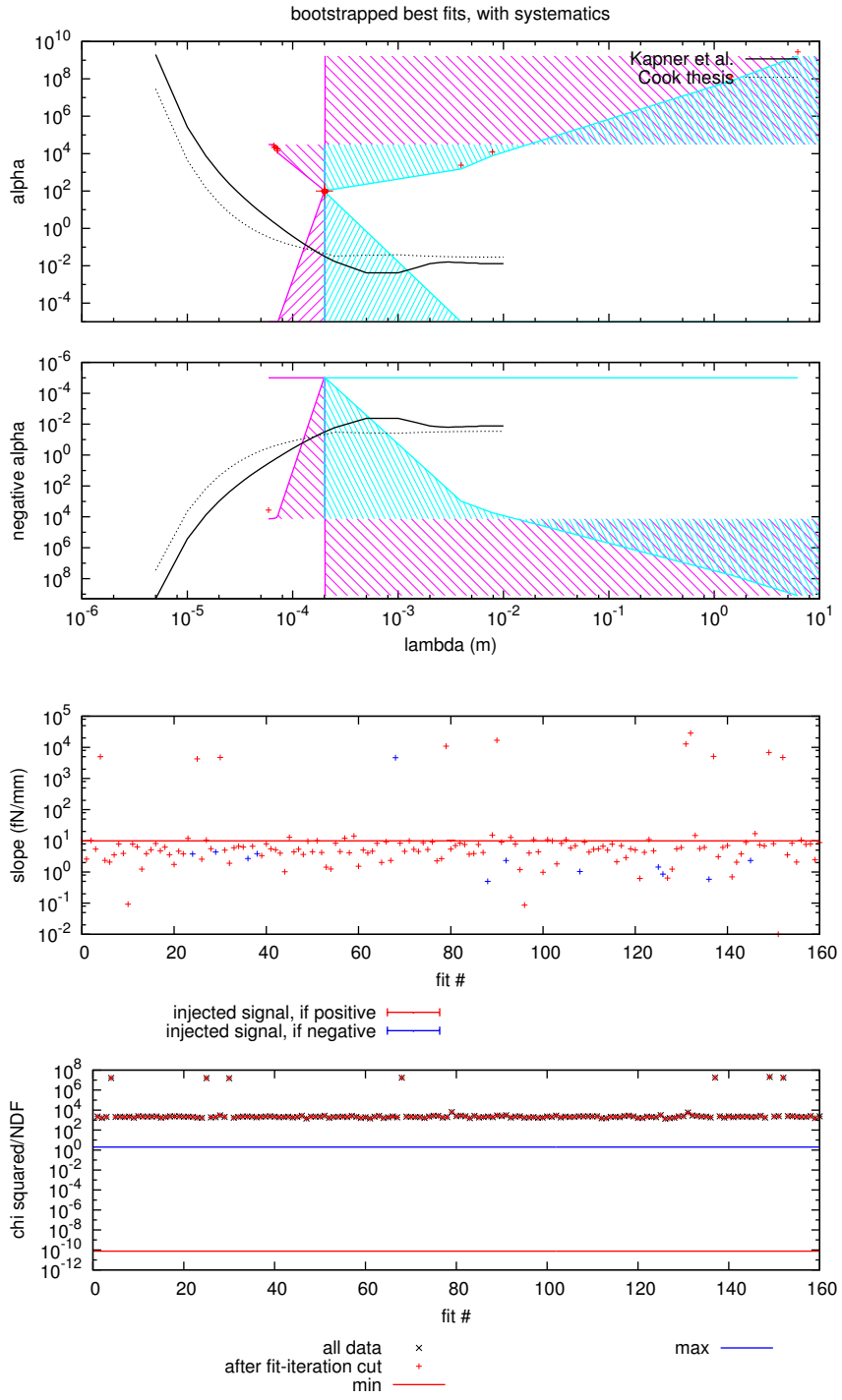

Figure B.31: Yukawa Plot, gravity only. A signal was injected with linear slope 1×10^{-11} N-m/m, Yukawa interaction with $\alpha = 100$, $\lambda = 200$ μm. The foil systematic was subtracted from the original data

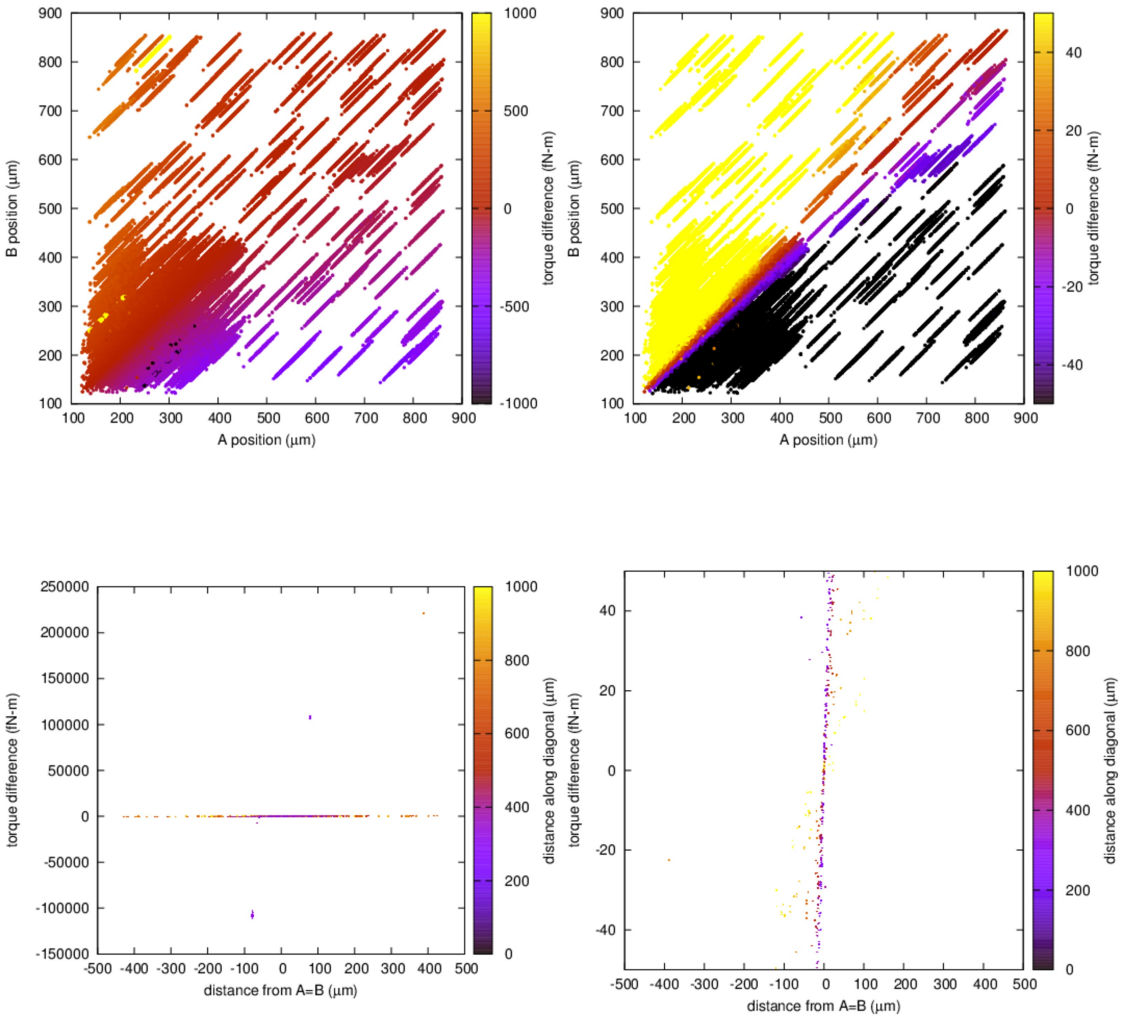

Figure B.32: Triangle plot. A signal was injected with linear slope 1×10^{-11} N-m/m, Yukawa interaction with $\alpha = 100$, $\lambda = 200$ μm. The foil systematic was not corrected.

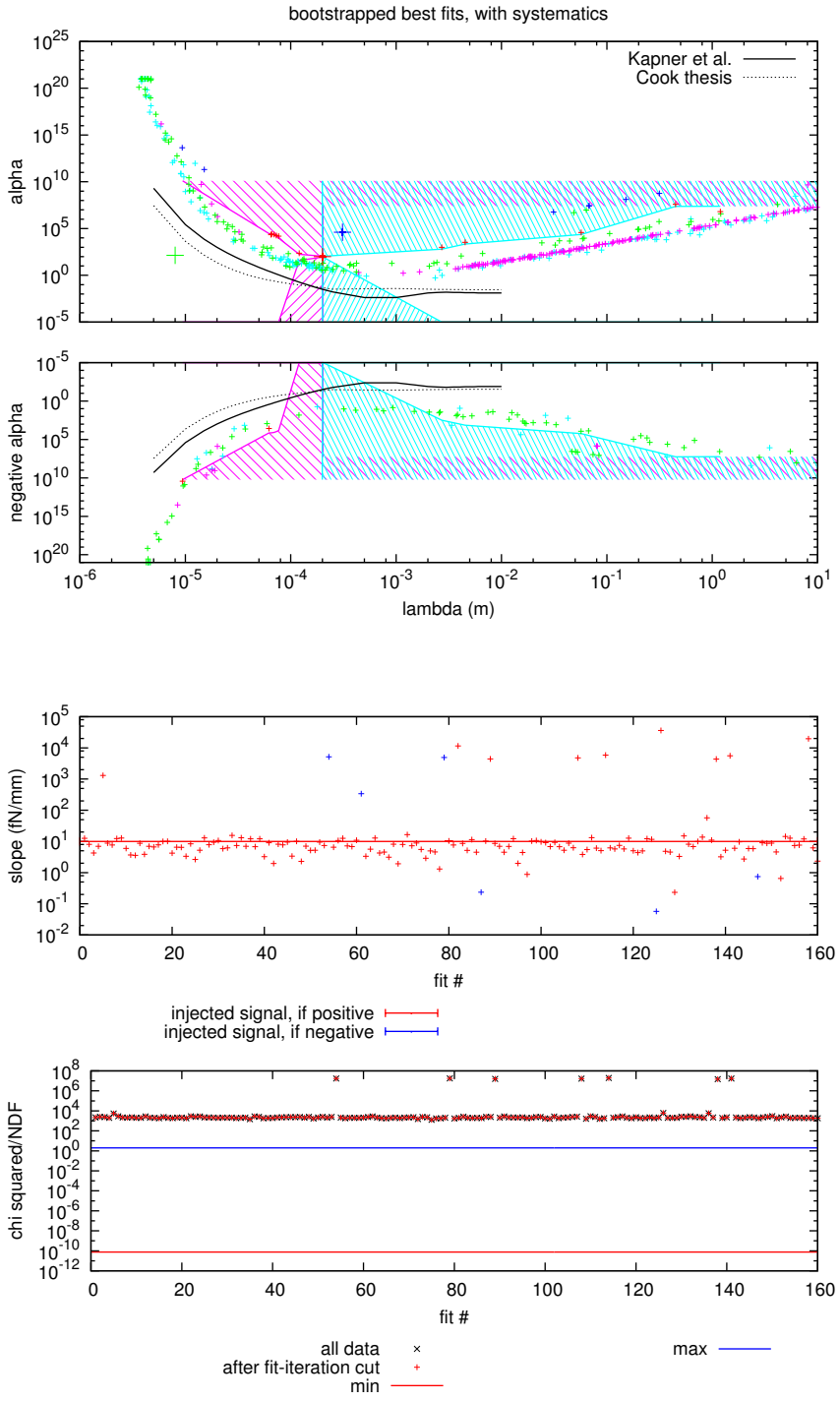

Figure B.33: Yukawa Plot, all systematics. A signal was injected with linear slope 1×10^{-11} N-m/m, Yukawa interaction with $\alpha = 100$, $\lambda = 200$ μm. The foil systematic was not corrected.

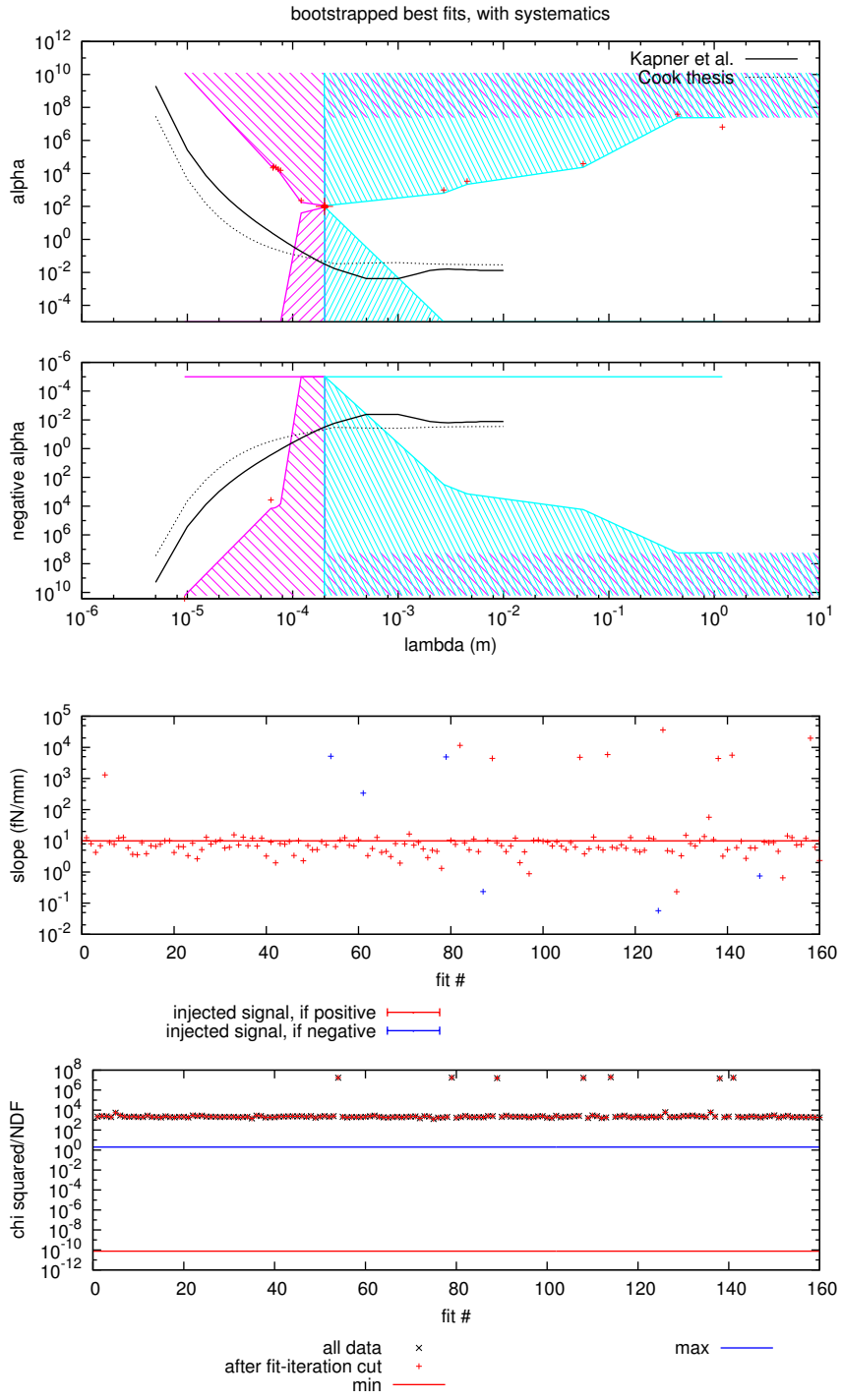

Figure B.34: Yukawa Plot, gravity only. A signal was injected with linear slope 1×10^{-11} N-m/m, Yukawa interaction with $\alpha = 100$, $\lambda = 200$ μm. The foil systematic was not corrected.

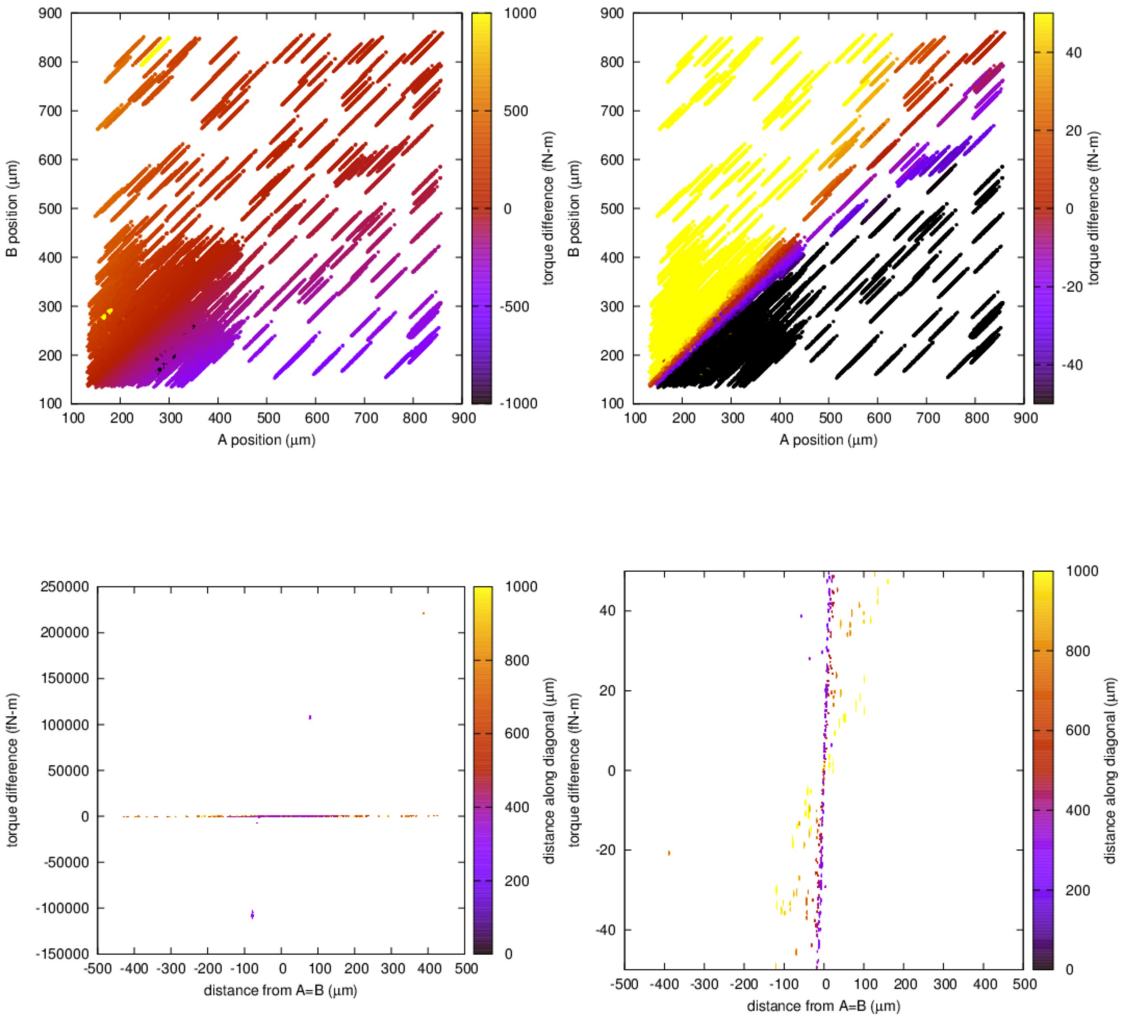

Figure B.35: Triangle plot. A signal was injected with linear slope 1×10^{-11} N-m/m, Yukawa interaction with $\alpha = 100$, $\lambda = 200$ μm. The foil systematic was added to the original data.

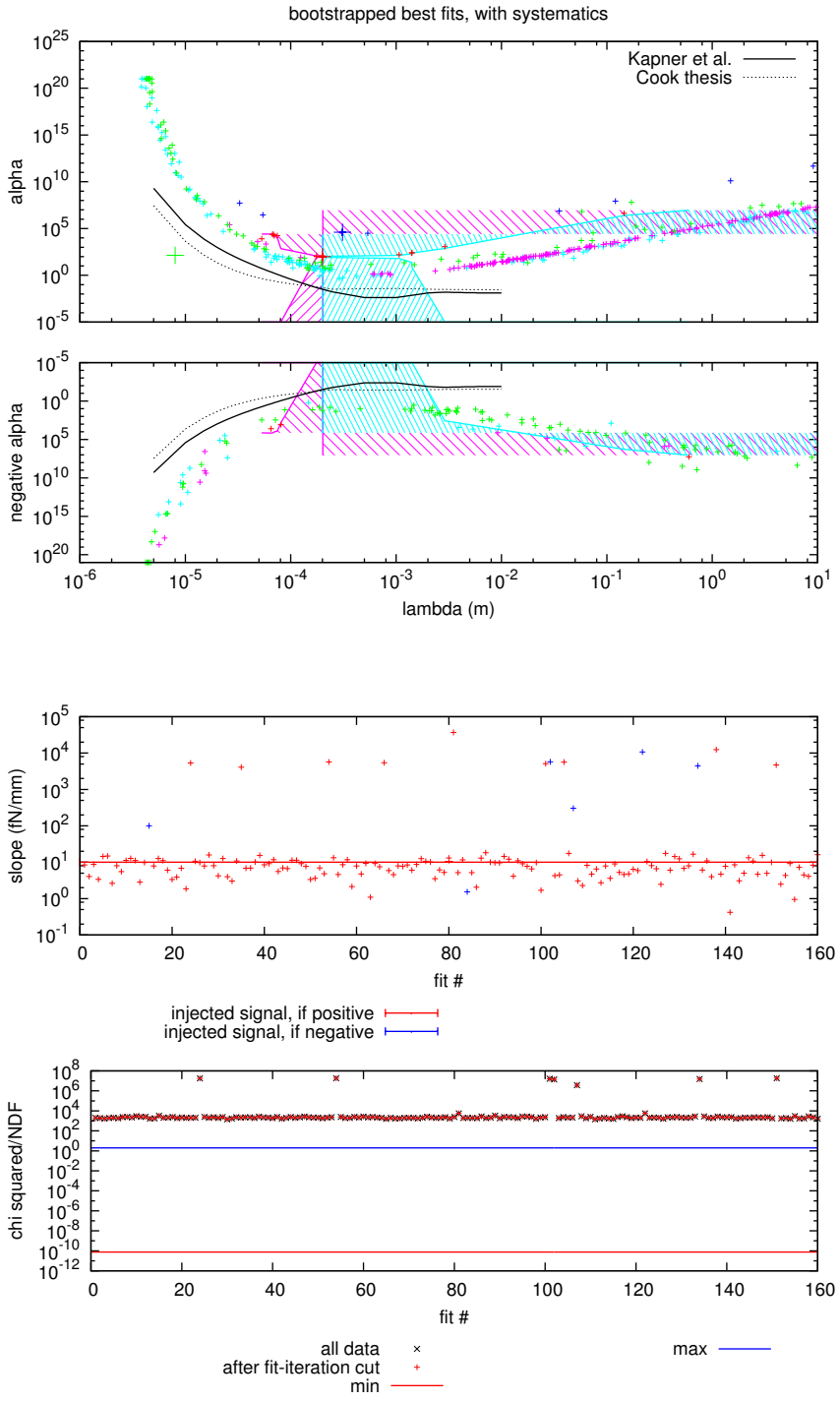

Figure B.36: Yukawa Plot, all systematics. A signal was injected with linear slope 1×10^{-11} N-m/m, Yukawa interaction with $\alpha = 100$, $\lambda = 200$ μm. The foil systematic was added to the original data.

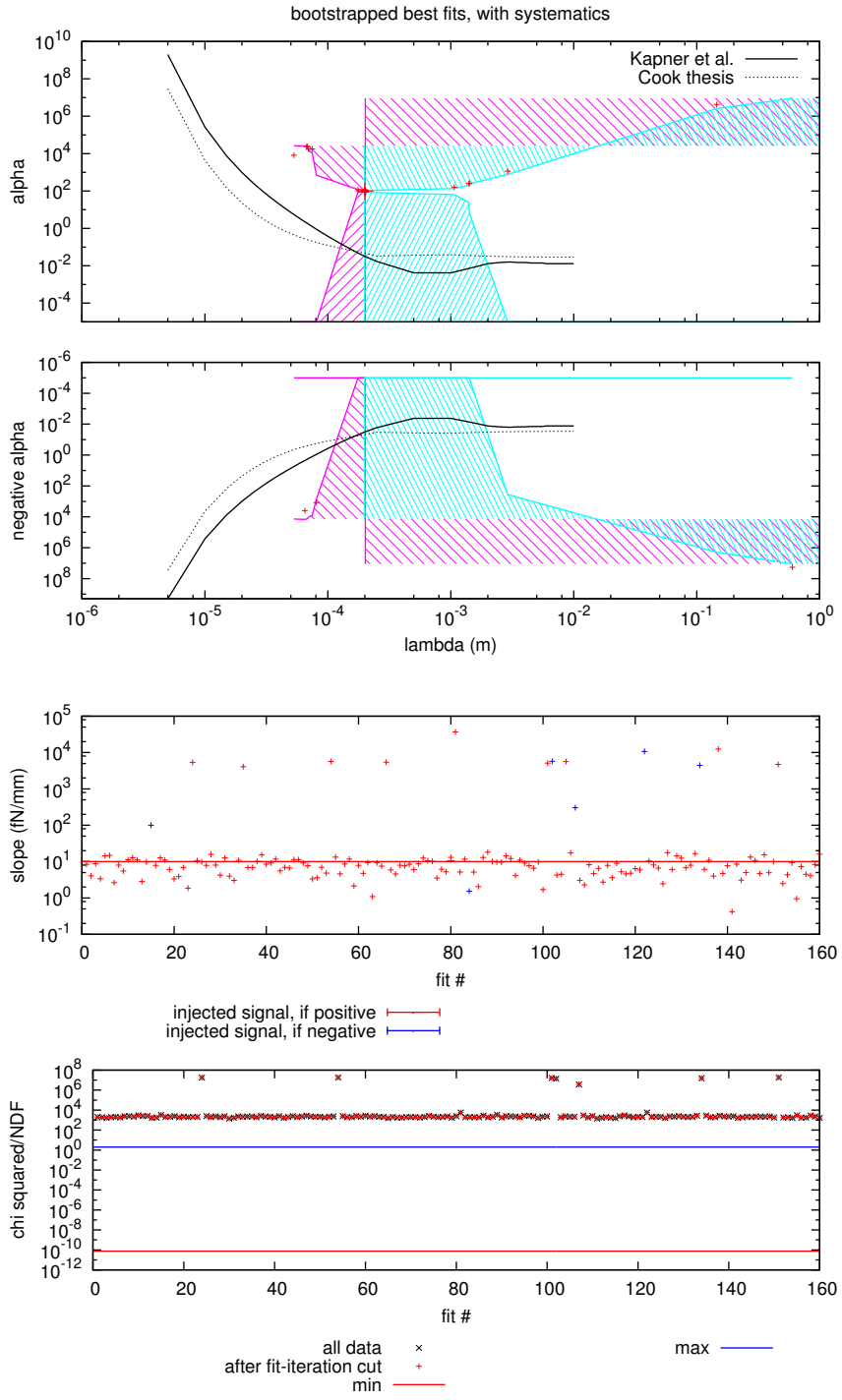

Figure B.37: Yukawa Plot, gravity only. A signal was injected with linear slope 1×10^{-11} N-m/m, Yukawa interaction with $\alpha = 100$, $\lambda = 200$ μm. The foil systematic was added to the original data.

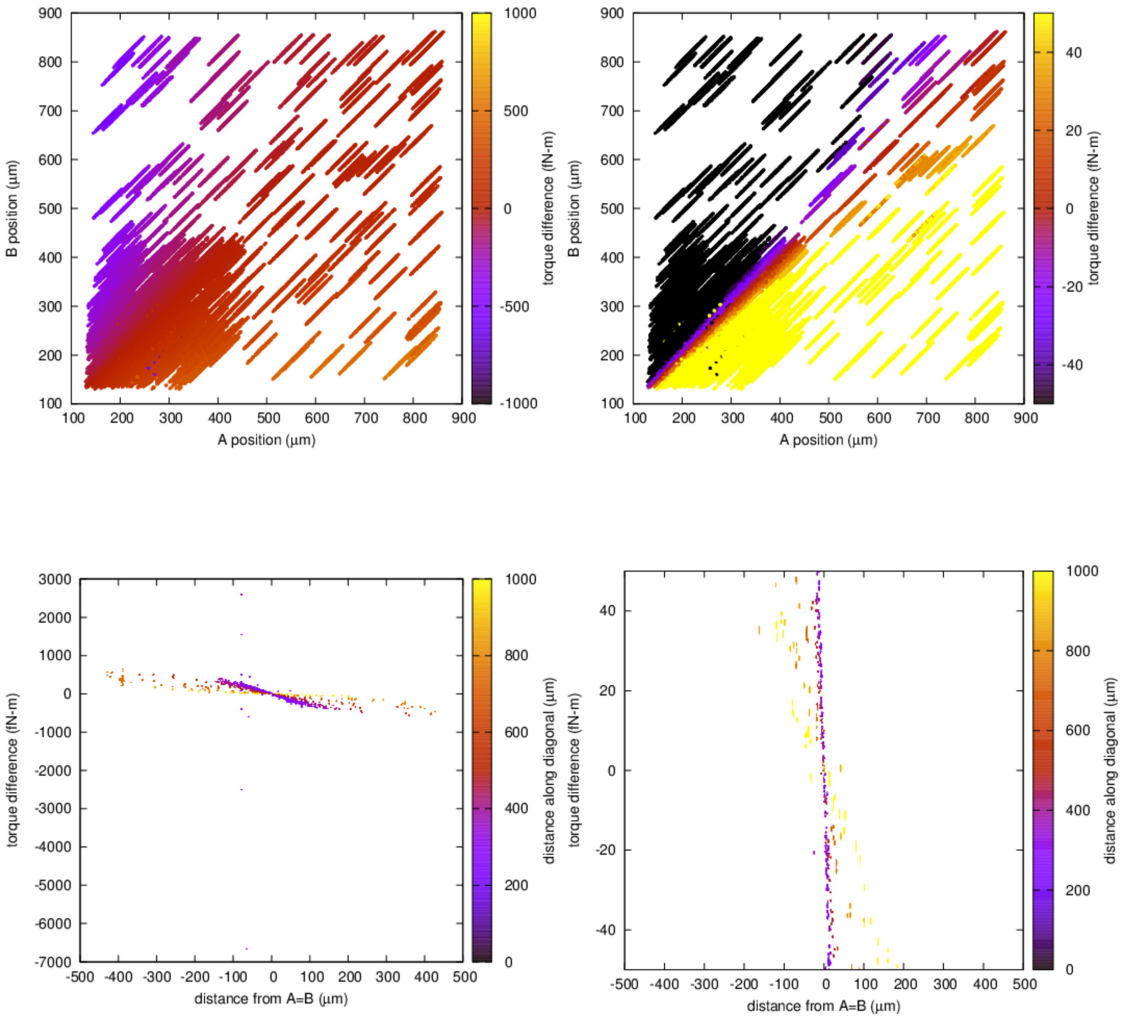

Figure B.38: Triangle plot. A signal was injected with linear slope 6×10^{-12} N-m/m, Yukawa interaction with $\alpha = -100$, $\lambda = 200$ μm. The foil systematic was subtracted from the original data

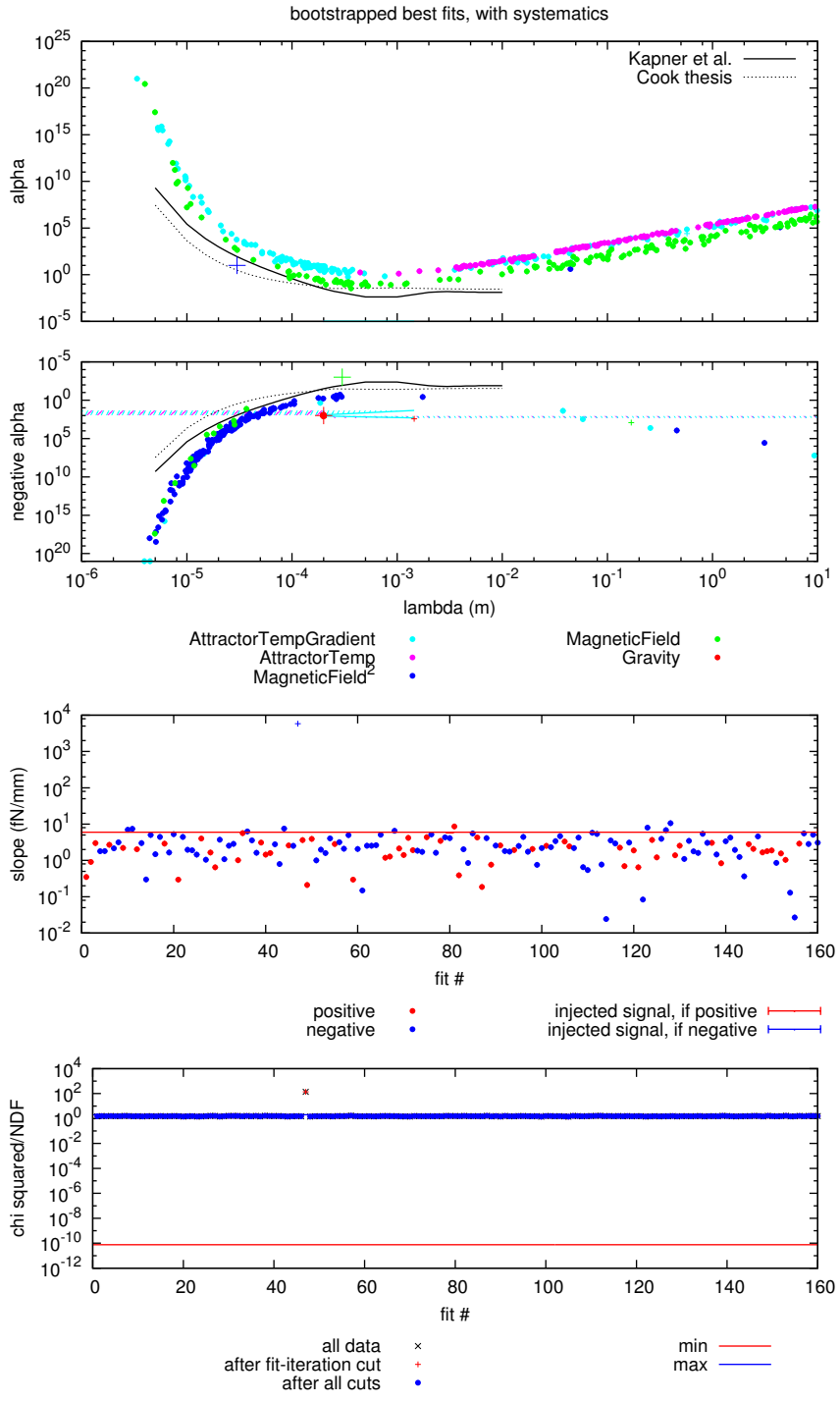

Figure B.39: Yukawa Plot, all systematics. A signal was injected with linear slope 6×10^{-12} N-m/m, Yukawa interaction with $\alpha = -100$, $\lambda = 200$ μm. The foil systematic was subtracted from the original data

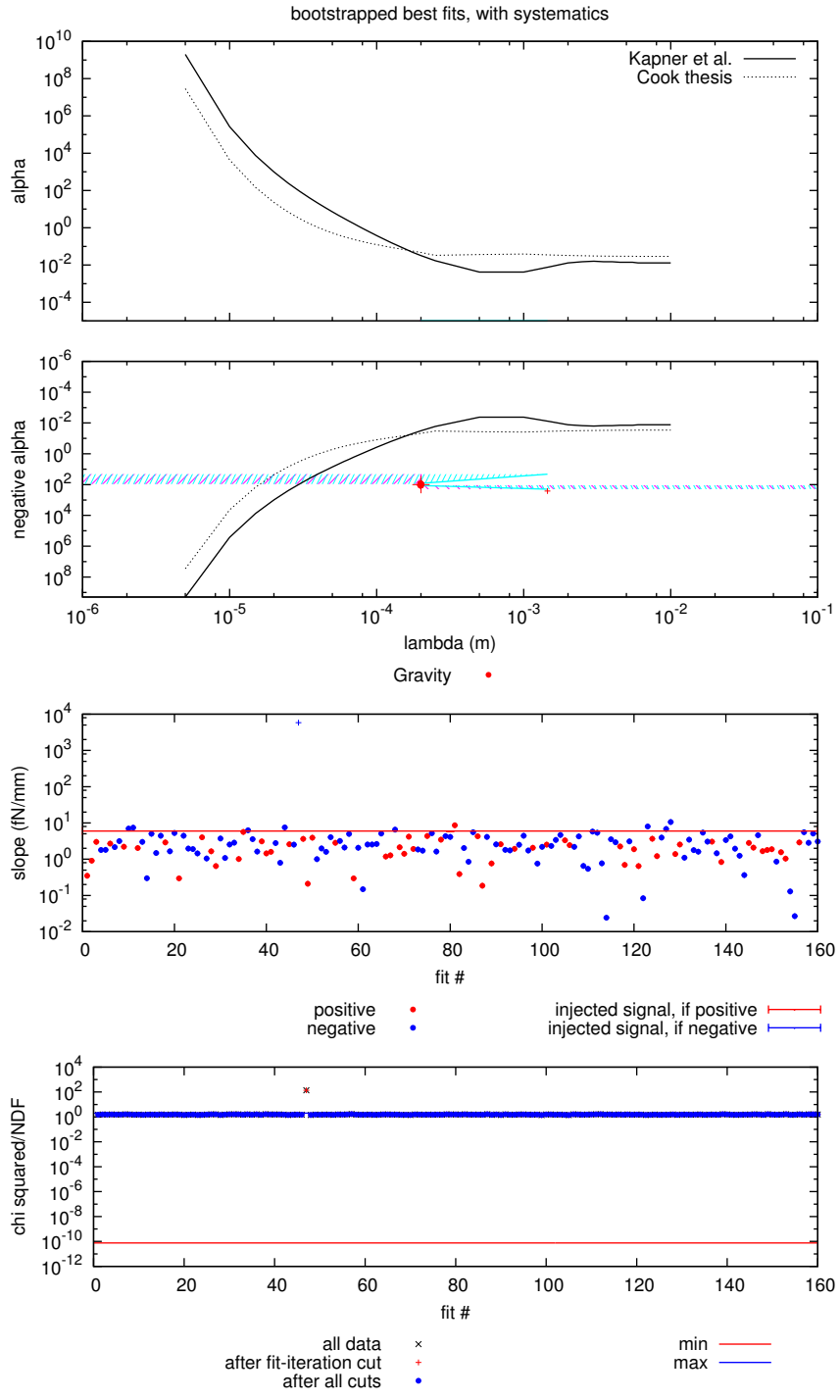

Figure B.40: Yukawa Plot, gravity only. A signal was injected with linear slope 6×10^{-12} N-m/m, Yukawa interaction with $\alpha = -100$, $\lambda = 200$ μm. The foil systematic was subtracted from the original data

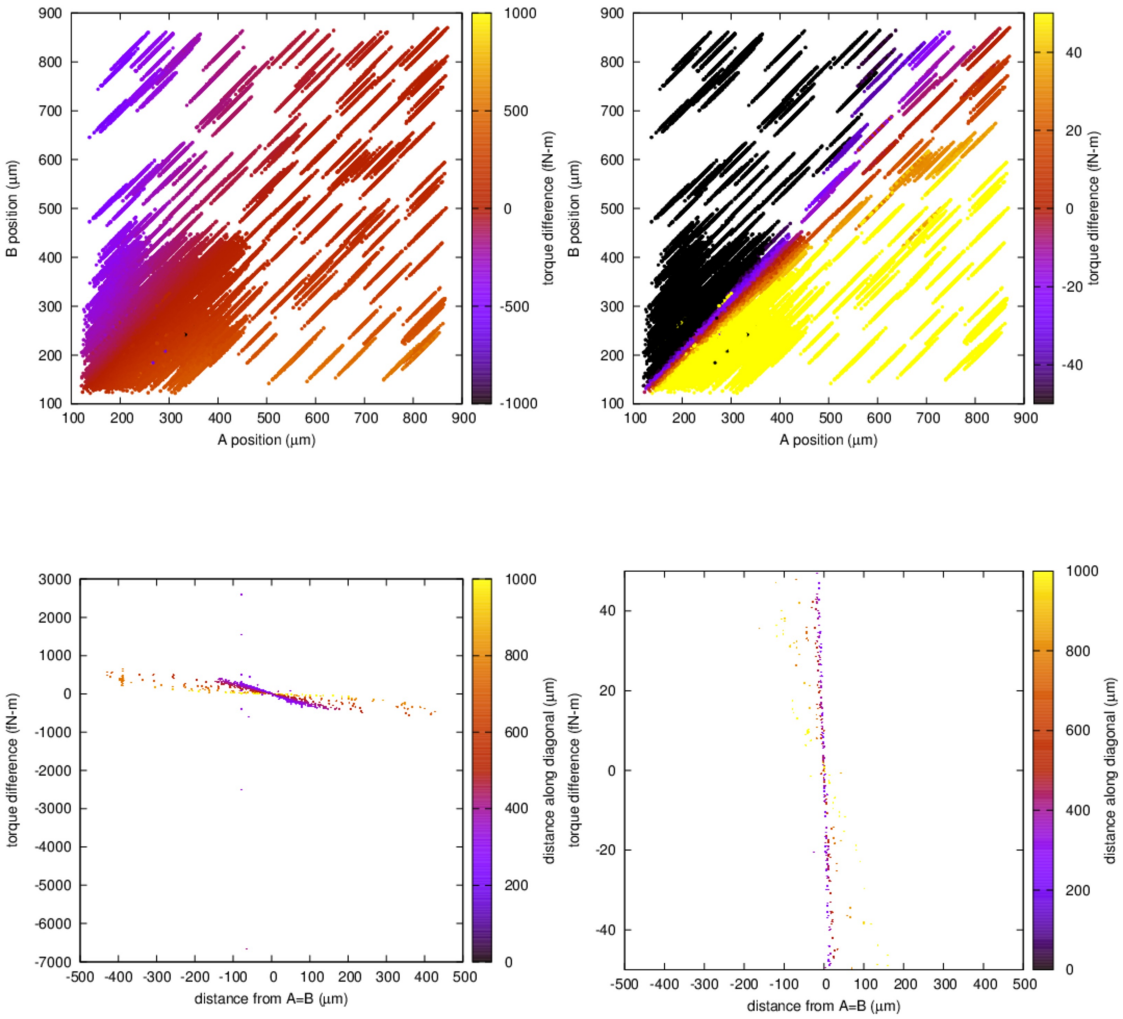

Figure B.41: Triangle plot. A signal was injected with linear slope 6×10^{-12} N-m/m, Yukawa interaction with $\alpha = -100$, $\lambda = 200$ μm. The foil systematic was not corrected.

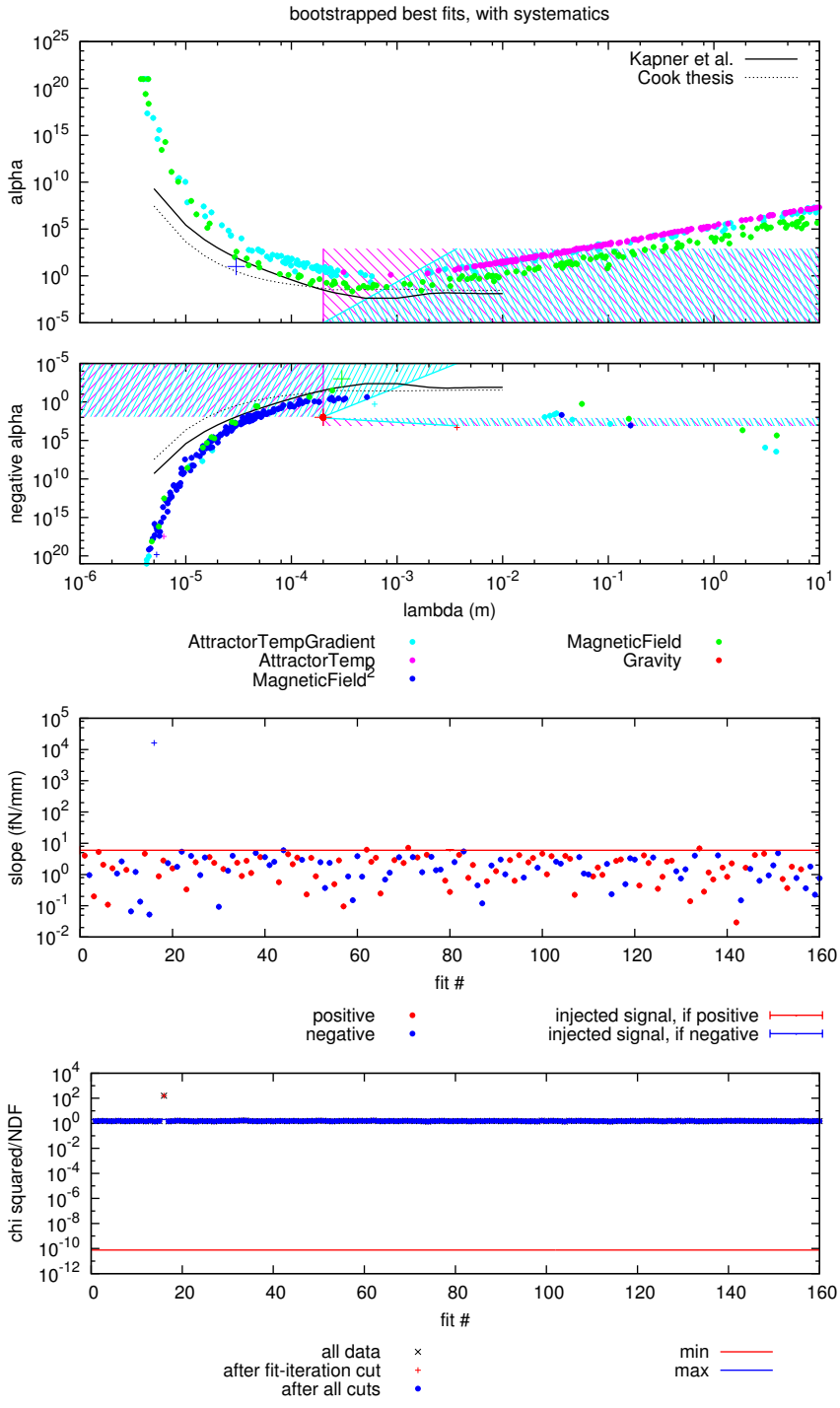

Figure B.42: Yukawa Plot, all systematics. A signal was injected with linear slope 6×10^{-12} N-m/m, Yukawa interaction with $\alpha = -100$, $\lambda = 200$ μm. The foil systematic was not corrected.

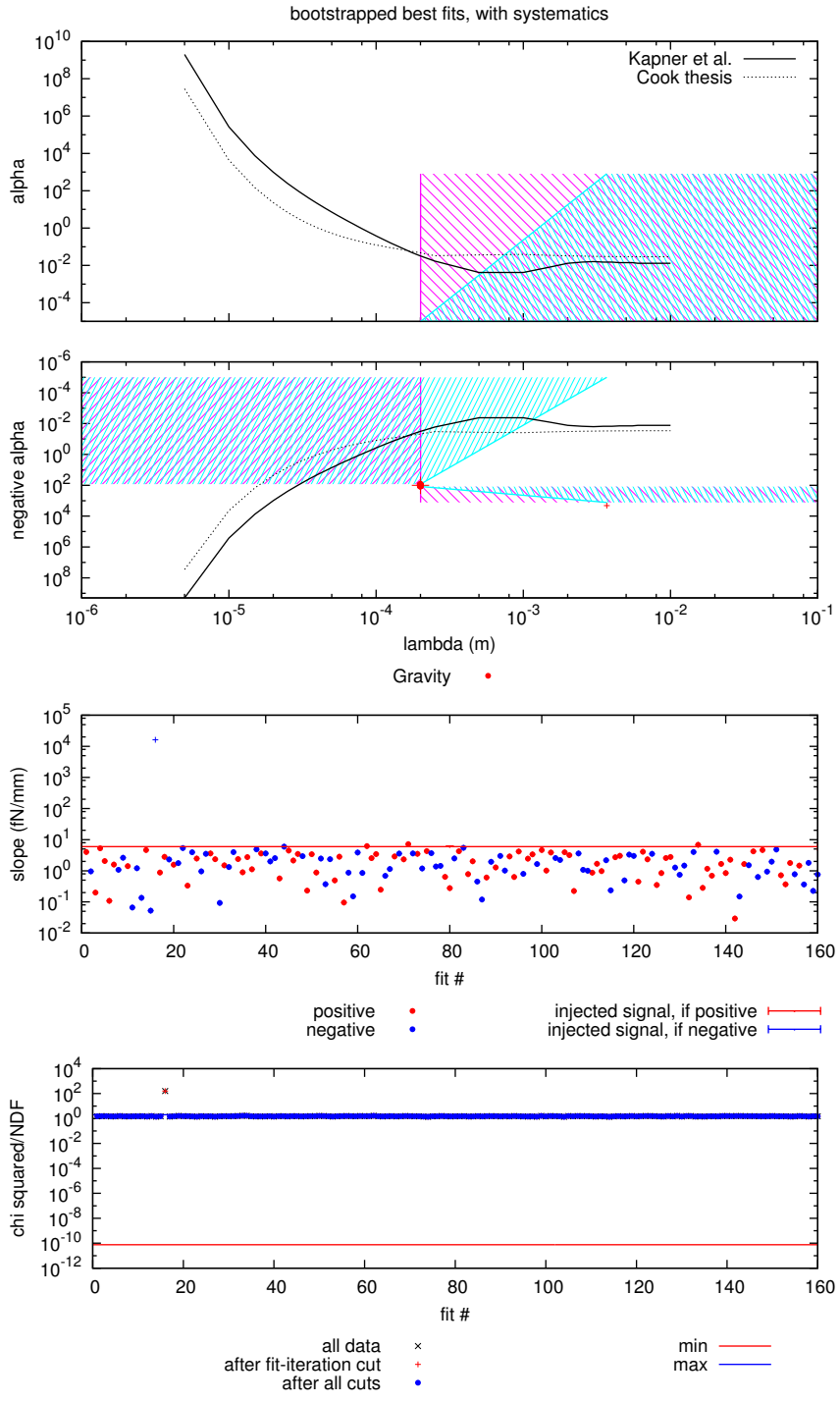

Figure B.43: Yukawa Plot, gravity only. A signal was injected with linear slope 6×10^{-12} N-m/m, Yukawa interaction with $\alpha = -100$, $\lambda = 200$ μm. The foil systematic was not corrected.

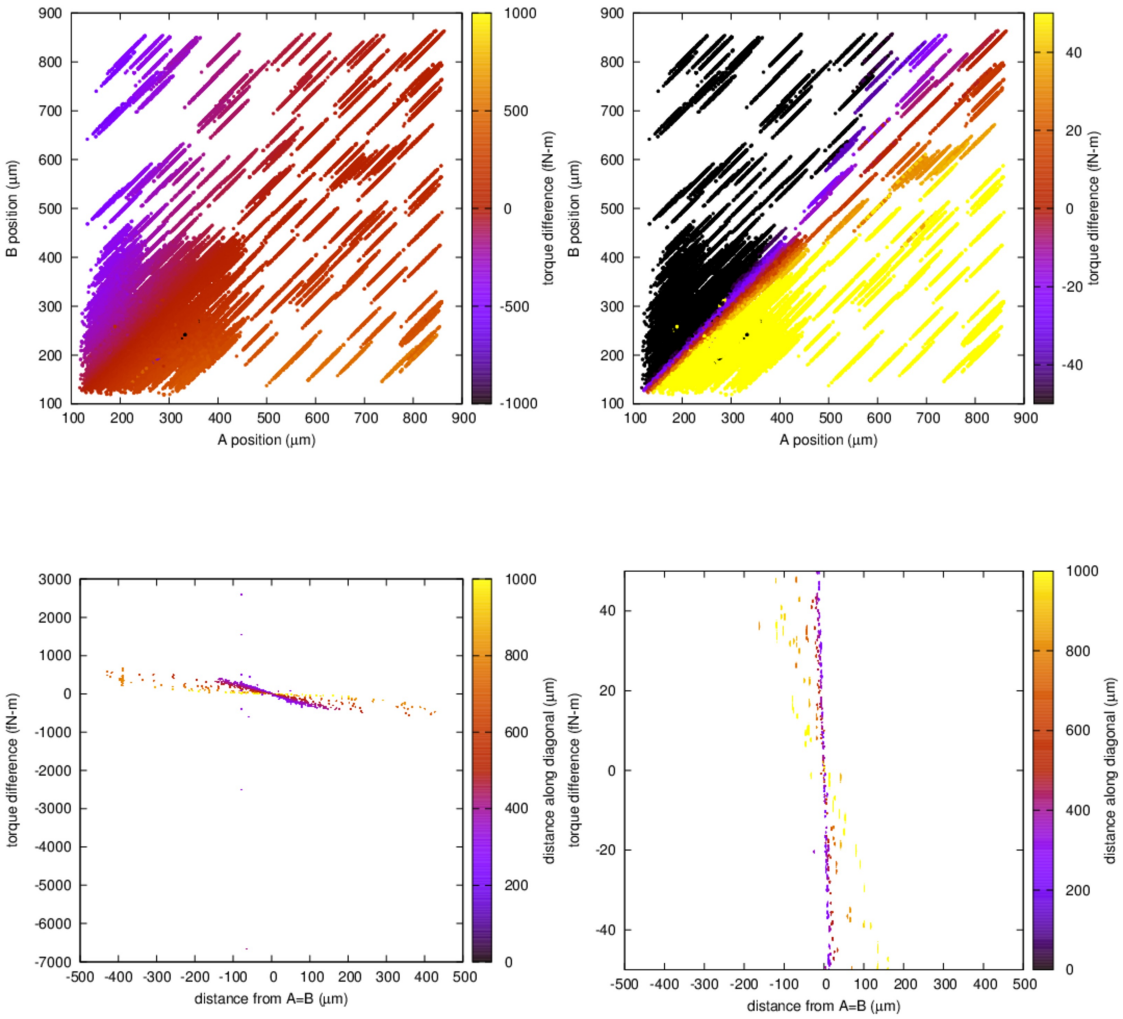

Figure B.44: Triangle plot. A signal was injected with linear slope 6×10^{-12} N-m/m, Yukawa interaction with $\alpha = -100$, $\lambda = 200$ μm. The foil systematic was added to the original data.

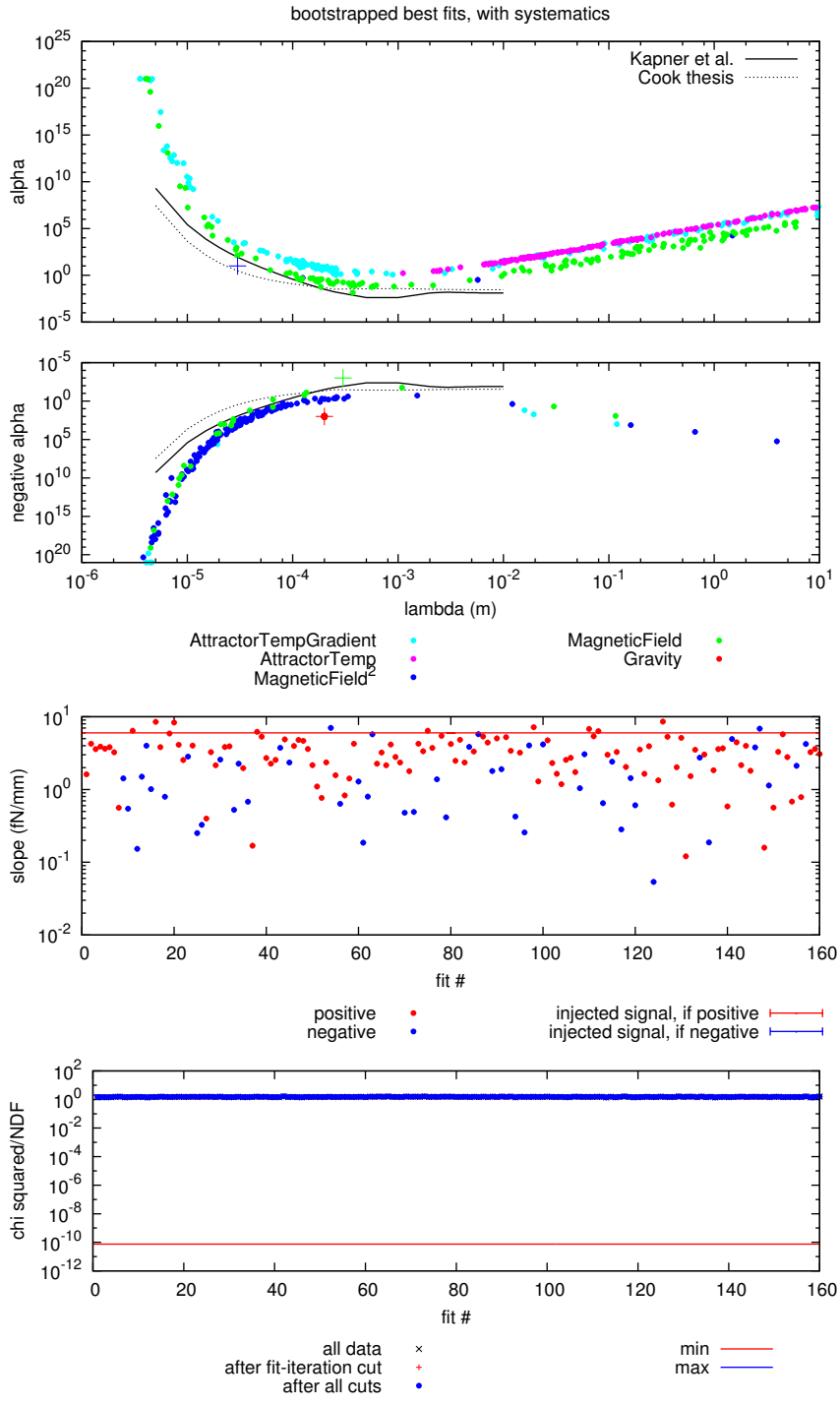

Figure B.45: Yukawa Plot, all systematics. A signal was injected with linear slope 6×10^{-12} N-m/m, Yukawa interaction with $\alpha = -100$, $\lambda = 200$ μm. The foil systematic was added to the original data.

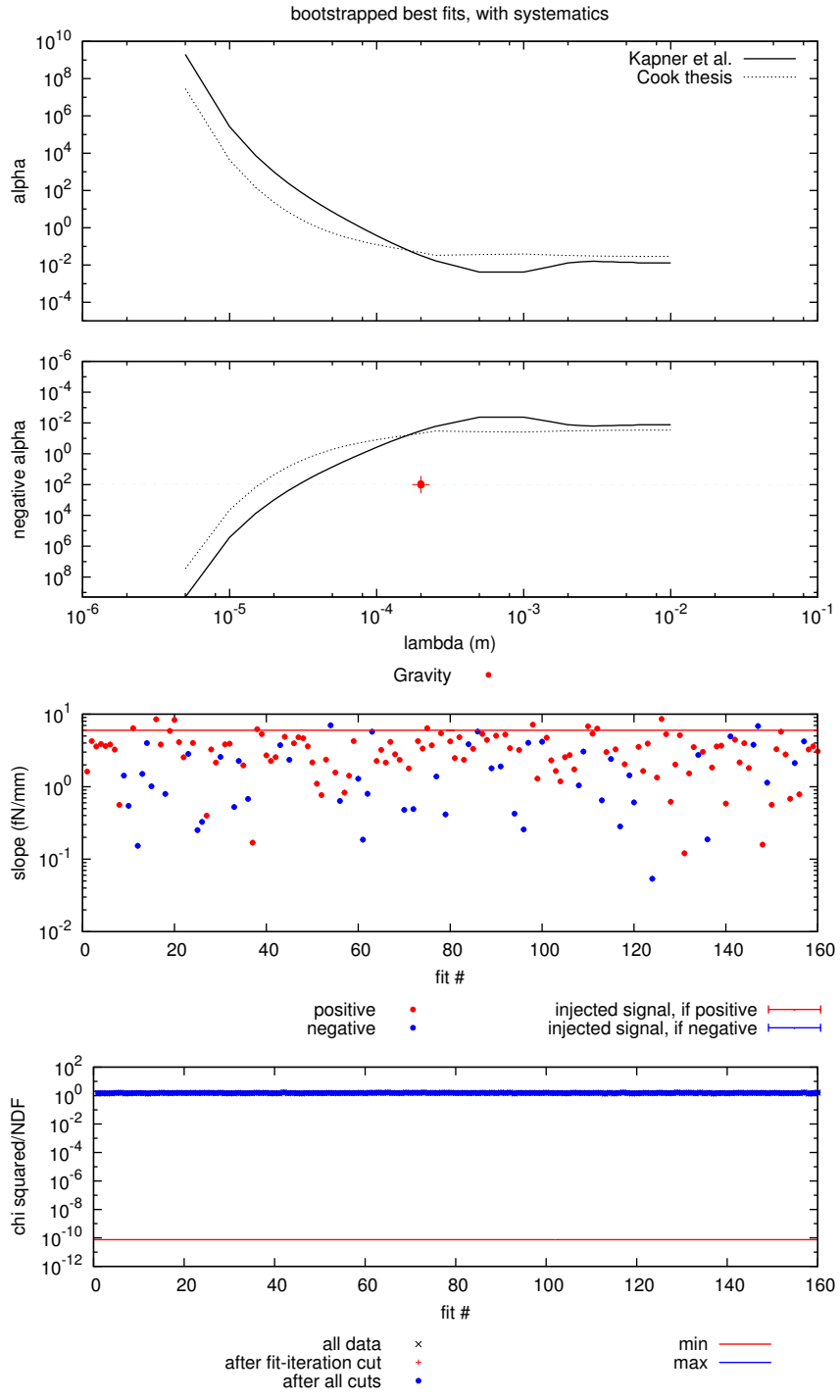

Figure B.46: Yukawa Plot, gravity only. A signal was injected with linear slope 6×10^{-12} N-m/m, Yukawa interaction with $\alpha = -100$, $\lambda = 200$ μm. The foil systematic was added to the original data.

Appendix C

Build System Configuration

The names of all packages installed on the Debian Linux computer used to generate this analysis and text are recorded in the project git repository. Furthermore, the SHA256 hash of the list, as evaluated by `sha256sum debianPackagesThesisBuild.txt` is `115f9bbfb23c0c49644c1f85b2db5f1b8ad968dcd09d94ec2d79a201d01d414c`.

Bibliography

[1] *Press release: The 1979 nobel prize in physics* (2013), URL `Nobelprize.org`.

[2] ATLAS Collaboration, Physics Letters B **716**, 1 (2012), ISSN 0370-2693.

[3] CMS Collaboration, Physics Letters B **716**, 30 (2012), ISSN 0370-2693.

[4] C. M. Will, Living Reviews in Relativity **9** (2006), URL `http://www.livingreviews.org/lrr-2006-3`.

[5] R. J. Adler, B. Casey, and O. C. Jacob, American Journal of Physics **63**, 620 (1995).

[6] J. Beringer, J. F. Arguin, R. M. Barnett, K. Copic, O. Dahl, D. E. Groom, C. J. Lin, J. Lys, H. Murayama, C. G. Wohl, et al. (Particle Data Group), Phys. Rev. D **86**, 010001 (2012).

[7] J. Silk, *Measuring and Modelling the Universe* (Cambridge University Press, Oxford, 2004), chap. Dark Matter Theory.

[8] G. Bennett, B. Bousquet, H. Brown, G. Bunce, R. Carey, P. Cushman, G. Danby, P. Debevec, M. Deile, H. Deng, et al., Physical Review D **73**, 072003 (2006).

[9] K. Fisher, R. Gerig, D. Green, J. Harkins, J. Kao, R. Kass, R. Lutha, P. Ostroumov, S. Prestemon, C. Rode, et al., Tech. Rep., Department of Energy (2014).

[10] P. Ade et al. (Planck Collaboration) (2013), `1303.5062`.

[11] P. Ade et al. (Planck Collaboration) (2013), `1303.5076`.

[12] I. Newton, *Philosophiae Naturalis Principia Mathematica* (Royal Society, 1687).

[13] H. Cavendish, Philosophical Transactions of the Royal Society of London **88**, 469 (1798).

[14] E. Fischbach and C. L. Talmadge, *The Search for Non-Newtonian Gravity* (1999).

[15] E. Adelberger, B. Heckel, and A. Nelson, Annual Review of Nuclear and Particle Science **53**, 77 (2003).

[16] T. Cook, Ph.D. thesis, University of Washington/CENPA (2013).

[17] R. Bonicalzi, P. Boynton, M. Moore, R. Newman, and E. Berg, in *Proceedings of the 12th Marcel Grossmann Meeting on General Relativity, Paris 2009.*, edited by R. J. T. Damour and R. Ruffini (World Scientific, 2011).

[18] L. Tu, S. Guan, J. Luo, C. Shao, and L. Liu, Physical Review Letters **98**, 201101 (2007).

[19] S.-Q. Yang, B.-F. Zhan, Q.-L. Wang, C.-G. Shao, L.-C. Tu, W.-H. Tan, and J. Luo, Phys. Rev. Lett. **108**, 081101 (2012).

[20] H. Yan, E. Housworth, H. Meyer, G. Visser, E. Weisman, and J. Long, arXiv preprint arXiv:1402.0145 (2014).

[21] K. Wagoner, Ph.D. thesis, Washington University in St. Louis (2010).

[22] S. J. Smullin, A. A. Geraci, D. M. Weld, J. Chiaverini, S. Holmes, and A. Kapitulnik, Physical Review D **72**, 122001 (2005).

[23] V. Nesterov, S. Buetefisch, and L. Koenders, Annalen der Physik **525**, 728 (2013).

[24] J. Murphy, T. W., E. G. Adelberger, J. B. R. Battat, L. N. Carey, C. D. Hoyle, P. LeBlanc, E. L. Michelsen, K. Nordtvedt, A. E. Orin, J. D. Strasburg, et al., Publications of the Astronomical Society of the Pacific **120**, pp. 20 (2008), ISSN 00046280, URL `http://www.jstor.org/stable/10.1086/526428`.

[25] K. Venkateswara, Ph.D. thesis, University of Maryland (2010).

[26] V. V. Nesvizhevsky, G. Pignol, and K. V. Protasov, Phys. Rev. D **77**, 034020 (2008).

[27] E. Chalkley, S. Aston, C. Collins, M. Nelson, and C. Speake (Rencontres de Moriond and GPhyS: Gravitational Waves and Experimental Gravity, 2011), URL `http://moriond.in2p3.fr/J11/transparents/chalkley.pdf`.

[28] S. Schiller, Ph.D. thesis, Heinrich-Heine-Universitat Dusseldorf (2006).

[29] R. Shankar, *Principles of Quantum Mechanics* (Plenum Press, 1994).

[30] N. Arkani-Hamed, S. Dimopoulos, and G. Dvali, Phys.Lett. **B429**, 263 (1998), `hep-ph/9803315`.

[31] N. Arkani-Hamed, S. Dimopoulos, and G. Dvali, Phys.Rev. **D59**, 086004 (1999), `hep-ph/9807344`.

[32] D. J. Kapner, T. S. Cook, E. G. Adelberger, J. H. Gundlach, B. R. Heckel, C. D. Hoyle, and H. E. Swanson, Physical Review Letters **98**, 021101 (2007), `arXiv:hep-ph/0611184`.

[33] CMS Collaboration (CMS Collaboration), Phys. Rev. Lett. **108**, 111801 (2012).

[34] I. Antoniadis, S. Baessler, M. Büchner, V. Fedorov, S. Hoedl, A. Lambrecht, V. Nesvizhevsky, G. Pignol, K. Protasov, S. Reynaud, et al., Comptes Rendus Physique **12**, 755 (2011).

[35] G. Hinshaw, D. Larson, E. Komatsu, D. Spergel, C. Bennett, J. Dunkley, M. Nolta, M. Halpern, R. Hill, N. Odegard, et al., arXiv:1212.5226 (2012).

[36] C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (W H Freeman and Company, 1973), 2nd ed.

[37] R. Sundrum, Phys. Rev. D **69**, 044014 (2004).

[38] E. G. Adelberger, B. R. Heckel, S. Hoedl, C. D. Hoyle, D. J. Kapner, and A. Upadhye, Physical Review Letters **98**, 131104 (2007), `arXiv:hep-ph/0611223`.

[39] J. Khoury and A. Weltman, Phys. Rev. D **69**, 044026 (2004).

[40] J. Khoury and A. Weltman, Phys. Rev. Lett. **93**, 171104 (2004).

[41] B. Feldman and A. E. Nelson, Journal of High Energy Physics **2006**, 002 (2006).

[42] A. E. Nelson and J. Walsh, Phys. Rev. D **77**, 095006 (2008).

[43] A. Upadhye, Phys. Rev. D **86**, 102003 (2012).

[44] A. Upadhye, Phys. Rev. Lett. **110**, 031301 (2013).

[45] B. Jain, A. Joyce, R. Thompson, A. Upadhye, J. Battat, P. Brax, A.-C. Davis, C. de Rham, S. Dodelson, A. Erickcek, et al., ArXiv e-prints (2013), `1309.5389`.

[46] J. Khoury, Classical and Quantum Gravity **30**, 214004 (2013).

[47] H. Georgi, Phys. Rev. Lett. **98**, 221601 (2007).

[48] H. Georgi, Physics Letters B **650**, 275 (2007), ISSN 0370-2693.

[49] H. Goldberg and P. Nath, Phys. Rev. Lett. **100**, 031803 (2008).

[50] N. Deshpande, S. D. Hsu, and J. Jiang, Physics Letters B **659**, 888 (2008).

[51] R. Newman, E. Berg, and P. Boynton, Space Science Reviews **148** (2009).

[52] J. Murata and S. Tanaka, Classical and Quantum Gravity **32**, 033001 (2015).

[53] J. Kepler, *Astronomia Nova* (1609).

[54] J. Kepler, *Harmonices Mundi* (1619).

[55] S. G. Turyshev, V. T. Toth, G. Kinsella, S.-C. Lee, S. M. Lok, and J. Ellis, Phys. Rev. Lett. **108**, 241101 (2012).

[56] C. Talmadge, J. P. Berthias, R. W. Hellings, and E. M. Standish, Phys. Rev. Lett. **61**, 1159 (1988).

[57] T. W. Murphy, Reports on Progress in Physics **76**, 076901 (2013).

[58] R. Spero, J. K. Hoskins, R. Newman, J. Pellam, and J. Schultz, Phys. Rev. Lett. **44**, 1645 (1980).

[59] M. V. Moody and H. J. Paik, Phys. Rev. Lett. **70**, 1195 (1993).

[60] C. D. Hoyle, U. Schmidt, B. R. Heckel, E. G. Adelberger, J. H. Gundlach, D. J. Kapner, and H. E. Swanson, Physical Review Letters **86**, 1418 (2001), `arXiv:hep-ph/0011014`.

[61] C. D. Hoyle, D. J. Kapner, B. R. Heckel, E. G. Adelberger, J. H. Gundlach, U. Schmidt, and H. E. Swanson, Physical Review Letters **70**, 042004 (2004), `arXiv:hep-ph/0405262`.

[62] J. C. Long, H. W. Chan, A. B. Churnside, E. A. Gulbis, M. C. M. Varney, and J. C. Price, Nature **421**, 922 (2003), `arXiv:hep-ph/0210004`.

[63] A. A. Geraci, S. J. Smullin, D. M. Weld, J. Chiaverini, and A. Kapitulnik, Phys. Rev. D **78**, 022002 (2008).

[64] D. M. Weld, J. Xia, B. Cabrera, and A. Kapitulnik, Physical Review D **77**, 062006 (2008).

[65] C. A. Gerig, Undergraduate thesis, Cornell University, Cornell University (2011), URL `https://math.berkeley.edu/~cgerig/physthesis.pdf`.

[66] L. Haiberger, D. Jäger, and S. Schiller, Review of scientific instruments **76**, 045106 (2005).

[67] H. Leopardi, 2012 NCUR (2013).

[68] A. O. Sushkov, W. J. Kim, D. A. R. Dalvit, and S. K. Lamoreaux, Phys. Rev. Lett. **107**, 171101 (2011).

[69] NIST, *2013 NIST precision measurement grants* (2014), URL `http://www.nist.gov/pml/div684/fcdc/upload/current-grantees.pdf`.

[70] R. Horvat, D. Kekez, M. Krcmar, Z. Krecak, and A. Ljubicic, Physics Letters B **699**, 21 (2011), ISSN 0370-2693.

[71] J. H. Simonetti, M. Kavic, D. Minic, U. Surani, and V. Vejayan, Astrophys.J. **737**, L28 (2011), `1010.5245`.

[72] M. Kavic, J. H. Simonetti, S. E. Cutchin, S. W. Ellingson, and C. D. Patterson, JCAP **0811**, 017 (2008), `0801.4023`.

[73] T. Arp, C. Hagedorn, S. Schlamminger, and J. Gundlach, Review of Scientific Instruments **84**, 095007 (2013).

[74] M. D. Turner, C. A. Hagedorn, S. Schlamminger, and J. H. Gundlach, Optics letters **36**, 1479 (2011).

[75] A. Fetter and J. Walecka, *Theoretical Mechanics of Particles and Continua* (Dover Publications, 2003).

[76] G. Arfken, *Mathematical Methods for Physicists* (Academic Press, 1970), 2nd ed.

[77] D. R. Lide, ed., *Handbook of Chemistry and Physics* (CRC Press, 2000), 81st ed.

[78] S. Schlamminger, C. A. Hagedorn, and J. H. Gundlach, Physical Review D **81**, 123008 (2010).

[79] H. Fizeau, Ann. Chim. Phys **66**, 429 (1862).

[80] D. Rugar, H. Mamin, R. Erlandsson, J. Stern, and B. Terris, Review of Scientific Instruments **59**, 2337 (1988).

[81] D. Rugar, H. Mamin, and P. Guethner, Applied Physics Letters **55**, 2588 (1989).

[82] Thorlabs, *Manual fiber polarization controllers*, Online (2013).

[83] R. Ulrich, S. Rashleigh, and W. Eickhoff, Optics letters **5**, 273 (1980).

[84] J. Ross, Optical and Quantum electronics **16**, 455 (1984).

[85] H. W. Ott, *Noise Reduction Techniques in Electronic Systems* (Wiley-Interscience, 1988).

[86] P. R. Saulson, Physical Review D **42**, 2437 (1990).

[87] G. I. Gonzalez and P. R. Saulson, Physics Letters A **201**, 12 (1995), ISSN 0375-9601.

[88] G. I. Gonzalez, Ph.D. thesis, Syracuse University. (1995).

[89] A. Cavalleri, G. Ciani, R. Dolesi, A. Heptonstall, M. Hueller, D. Nicolodi, S. Rowan, D. Tombolato, S. Vitale, P. Wass, et al., Physical review letters **103**, 140601 (2009).

[90] R. Dolesi, M. Hueller, D. Nicolodi, D. Tombolato, S. Vitale, P. Wass, W. Weber, M. Evans, P. Fritschel, R. Weiss, et al., Physical Review D **84**, 063007 (2011).

[91] A. Safavi-Naini, P. Rabl, P. F. Weck, and H. R. Sadeghpour, Phys. Rev. A **84**, 023412 (2011).

[92] D. A. Hite, Y. Colombe, A. C. Wilson, K. R. Brown, U. Warring, R. Jördens, J. D. Jost, K. S. McKay, D. P. Pappas, D. Leibfried, et al., Phys. Rev. Lett. **109**, 103001 (2012).

[93] J. Lecoeur, J. Bellier, and C. Koehler, Electrochimica Acta **35**, 1383 (1990), ISSN 0013-4686.

[94] S. K. Lamoreaux, Physical Review Letters **78**, 5 (1997).

[95] J. Campbell, The Quarterly Journal of Mechanics and Applied Mathematics **9**, 84 (1956).

[96] M. Suijlen, J. Koning, M. van Gils, and H. Beijerinck, Sensors and Actuators A: Physical **156**, 171 (2009).

[97] Krayden Incorporated, *Hysol eccobond 24 datasheet*, URL `krayden.com/tds/henk\_eccobond\_24\_tds.pdf`.

[98] T. Varpula and T. Poutanen, Journal of Applied Physics **55**, 4015 (1984).

[99] S.-K. Lee and M. Romalis, Journal of Applied Physics **103**, 084904 (2008).

[100] J. D. Jackson, *Classical Electrodynamics* (Wiley and Sons, 1999), 3rd ed.

[101] J. R. Davis et al., *Copper and copper alloys* (ASM international, 2001).

[102] A. C. Jones, B. T. O'Callahan, H. U. Yang, and M. B. Raschke, Progress in Surface Science **88**, 349 (2013).

[103] O. Loebich, Gold Bulletin **5**, 2 (1972).

[104] M. Bantel and R. Newman, Journal of Alloys and Compounds **310**, 233 (2000).

[105] S. Schlamminger, *Thermal Noise Wallet Card* (2009).

[106] *NTP: The network time protocol* (2015), URL `http://www.ntp.org/`.

[107] Microsoft, *Support boundary to configure the windows time service for high-accuracy environments* (2015), URL `http://support.microsoft.com/kb/939322`.

[108] Octave community, *GNU/Octave* (2012), URL `www.gnu.org/software/octave/`.

[109] A. V. Oppenheim, R. W. Schafer, and J. R. Buck, *Discrete-time Signal Processing* (Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1999), 2nd ed., ISBN 0-13-754920-2.

[110] S. Orfandis, *Introduction To Signal Processing* (Prentice-Hall, Inc., 1996).

[111] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery, *Numerical Recipes in C++* (Cambridge University Press, 2002).

[112] J. R. Taylor, *An Introduction To Error Analysis: The Study Of Uncertainties In Physical Measurements Author: John R. Taylor, Publisher* (University Science Books, 1996).

[113] D. J. Kapner, Ph.D. thesis, University of Washington (2005).

[114] L. Torvalds, J. C. Hamano, and many others, *Git –everything-is-local*, `http://git-scm.com` (2014).

[115] R. M. R.M. Stallman and P. Smith, *GNU Make Version 3.81* (Free Software Foundation, 2013), URL `http://www.gnu.org/software/make/manual/`.

[116] T. Williams, C. Kelley, and many others, *Gnuplot 4.6: an interactive plotting program*, `http://gnuplot.info/` (2014).

[117] D. Gruber, *Do we really need quaternions?* (2000), URL `http://www.gamedev.net/page/resources/\_/technical/math-and-physics/do-we-really-need-quaternions-r1199`.

[118] J. H. Gundlach, C. A. Hagedorn, K. K. Krause, S. Pollack, and S. Schlamminger, CENPA Annual Report (2007).

[119] S. Caley, P. Emani, J.H.Gundlach, C. Hagedorn, K. K. Krause, S. E. Pollack, S. Schlamminger, and M. D. Turner, CENPA Annual Report (2008).

[120] J. H. Gundlach, C. A. Hagedorn, S. Schlamminger, and M. D. Turner, CENPA Annual Report (2009).

[121] J. H. Gundlach, C. A. Hagedorn, S. Schlamminger, and M. Turner, CENPA Annual Report (2010).

[122] J. H. Gundlach, C. A. Hagedorn, S. Schlamminger, and M. D. Turner, CENPA Annual Report (2011).

[123] J. H. Gundlach, C. Hagedorn, and M. Turner, CENPA Annual Report (2012).

[124] J. H. Gundlach, C. A. Hagedorn, M. Turner, and K. Venkateswara, CENPA Annual Report (2013).

[125] AsahiKasei, *AK8975/AK8975C Datasheet* (2010).

[126] I. McCammon, in *Proceedings ISSW* (2002), pp. 244–251.

[127] D. Atkins, in *International Snow Science Workshop* (2000), pp. 46–51.

[128] S. Schlamminger, K.-Y. Choi, T. Wagner, J. Gundlach, and E. Adelberger, Physical review letters **100**, 041101 (2008).

[129] K. Snover, Private Communication.

[130] S. Pollack, M. Turner, S. Schlamminger, C. Hagedorn, and J. Gundlach, Physical Review D **81**, 021101 (2010).

[131] J. Baron, W. C. Campbell, D. DeMille, J. M. Doyle, G. Gabrielse, Y. V. Gurevich, P. W. Hess, N. R. Hutzler, E. Kirilov, I. Kozyryev, et al., Science **343**, 269 (2014).

[132] C. Hagedorn, *Defense 1-6*, Video; Youtube (2015), URL `https://www.youtube.com/watch?v=tkunLtv4GRE`.

[133] C. Hagedorn, *Unblinding tweet*, Twitter (2015), URL `https://twitter.com/CharlieHagedorn/status/591782911860285440`.