Cavities All the Way Down: Precision Measurement for Gravitational Waves and Quantum Spacetime

Thesis by
Ian Alexander Ouellette MacMillan

In Partial Fulfillment of the Requirements for the
Degree of
Doctor of Philosophy



CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2027
Defended June 30, 2026

© 2027

Ian Alexander Ouellette MacMillan
ORCID: 0000-0002-6927-1031

All rights reserved except where otherwise noted

To Mom and Dad

ACKNOWLEDGEMENTS

Graduate school has a way of making very large projects feel, in retrospect, like a collection of very small moments. Years of work eventually collapse into a handful of memories, and almost all of mine are tied to the people who were there. This thesis bears their influence on every page, often in ways that would be impossible to trace. They made both me and this thesis better, and I am grateful for the ways they shaped the years behind this work.

I am especially grateful to Prof. Lee McCuller, who became my advisor three years into my PhD and took me on as the first graduate student in his group. The work that followed was shaped by Lee's close mentorship and willingness to work through problems at every level. His confidence, patience, and example have had a lasting influence on me. Alex Ramirez, Jeff Wack, Briana Chen, and Daniel Grass have made building GQuEST a genuinely shared project, taking ownership of major parts of the experiment and continuing to move it forward. Torrey Cullen and Sander Vermeulen brought experience and perspective to many of the problems we have encountered along the way. I am also grateful to Umran, Erin, Rafa, and Riley for the energy and thought they brought to their SURF projects. Working alongside this group made the challenges of building a new experiment and working in the laboratory both manageable and enjoyable.

My early experiences as a graduate student were shaped by the members of the LIGO experimental group. I am grateful to Shruti Maliakal, Radhika Bhatt, Peter Carney, Anchal Gupta, Duo Tao, Aaron Markowitz, Rajashik Tarafder, and Gautam Venugopalan for the discussions, assistance, and mentorship that filled daily life in the group. Helen Schwartz and Jancarlo Sanchez kept difficult experimental work moving and complemented it with practical knowledge. I also thank Paco Salces Carcoba, Jon Richardson, Yuta Michimura, Andrew Wade, Yehonathan Drori, Aaron Goodwin-Jones, Tega Edo, and Nathan Holland for their advice and generosity, and Prof. Rana Adhikari, Koji Arai, Christopher Wipf, and Aidan Brooks for sharing their vast experience and perspective.

Caltech's relativity and data-analysis group provided a strong sense of community while connecting the details of my experimental work to the objects and events being observed. Prof. Yanbei Chen, Prof. Alan Weinstein, and Prof. Katerina Chatziioannou were generous with their teaching and perspective. Sophie Hourihane, Joey

Payne, Rhiannon Udall, Simona Miller, Brian Seymour, James Gardner, Andrew Laeuger, Colin Weller, Su Direkçi, Isaac Legred, Jacob Golomb, Taylor Knapp, Derek Davis, Lucy Thomas, Sophie Bini, Jane Glanzer, Ryan Magee, Eliot Finch, and Jonah Kanner are, above all, my friends. Many of my best memories at Caltech are of the time spent talking, laughing, and putting off work. Their friendship made the long stretches of graduate school much easier and made the work more enjoyable.

The LIGO community supplied knowledge, mentorship, and experience that no single group could reproduce. I thank Prof. Gregg Harry and Garrett Cole for their contributions and guidance, and for giving me the opportunity to learn from the best. I also thank Gabriele Vajente, Elenna Capote, and Calum Torrie for sharing their scientific expertise. I am grateful to Ali James, Thomas Roocke, Thejas Seetharamu, Graeme McGhee, Arnaud Pele, Michael Coughlin, Sharon Brunett, Madeline Lesovsky, Stephen Appert, and Amanda Obregon for the expertise and practical help and advice. I am especially appreciative of Georgia Bolingbroke, who kept me going through the long hours, particularly in the final push. My thanks also go to Liz Natividad and Nately Reilly for the day-to-day help that made everything else possible, and to Dave Reitze for keeping LIGO running smoothly and, more importantly, funded. I also thank Prof. Barry Barish for the daily use of his reserved parking spot, which he unknowingly provided.

Many other students made Caltech a richer and more vibrant place to spend these years, including Elie Bataille, Stephen Naus, Dylan King, Nabha Shah, John Lathrop, Rahaf Youssef, Elina Sendonaris, Brandon Zhao, Michele Tarquini, Sophia Tevosyan, Eli Weissler, Mike McKeon, Maria Camarca, Costa Christopoulos, Jordan Benjamin, and Hannah Manetsch. I am particularly grateful to my fellow G1s, especially Yanky Landau, Ana Duarte, Chandler Conn, André Nicolov, Phelan Yu, Piero Chiappina, Ina Sørensen, Liam O'Brien, and Silvia Zhang, for sharing the journey with me.

The friends I found outside the laboratory gave me the balance and support needed to keep going. I am grateful to Miguel and Claudia Palos, Neil Ross, Val, Emmie Chipps, Sahidah Briggs, Melissa Galvin, Katie Rubbelke, Emma Walker, Persi Tam, Tyler Fleig, Jamie Tavares, Alicia Reed, Chase Little, Rafael Villalovos, Jessica Williams, Courtney Smith, Lindsay Tausch, Ana Valdes Curiel, Adrián Salces Valdes, and many others. Their friendship, hospitality, humor, and encouragement gave these years a life beyond the work.

My family remained a steadfast support system beyond the windowless laser labs. My siblings, Maddie and Ruairi, and my dogs, Jack and Luna, kept me connected to home and to the parts of life that cannot be measured by progress.

I could not have reached this point without both of my parents, whose support has never wavered. They gave me the freedom and confidence to pursue what mattered to me. Even though my dad did not get to see me finish my thesis, much of what eventually brought me to this point began with him. He found joy in understanding the world and the people in it, in a way that can only be called contagious. My mom has been a constant source of strength and reassurance, believing in me through every success and setback and always reminding me of the life that existed beyond the work. Together, their constant support kept me grounded and sane whenever the work threatened to consume everything else. Whatever is worthwhile in this thesis rests on the foundation they gave me.

ABSTRACT

Ground-based interferometers must aggressively filter noise while preserving their fundamental signal response. In the Laser Interferometer Gravitational-Wave Observatory (LIGO), this requires active alignment control that rejects low-frequency noise without reducing astrophysical sensitivity; in Gravity from the Quantum Entanglement of Space Time (GQuEST), the challenge is optical, requiring filter cavities to strip laser noise and carrier light before photon counting.

To improve LIGO alignment control, this thesis develops an optimal-control framework that balances residual angular motion against control noise coupled into the gravitational-wave readout. A generalized state-space model represents both effects and converts the strain-coupled noise into binary-neutron-star detection-range loss. An $\mathcal{H}_\infty$-bounded $\mathcal{H}_2$ synthesis minimizes the combined cost while enforcing robustness. Sweeping the performance weights and robustness bound produces a Pareto front among detection range, angular motion, and stability. Applied to a Hanford arm-cavity yaw loop, the method reduces modeled range loss by roughly an order of magnitude and angular root-mean-square motion by up to a factor of four relative to the current controller at comparable robustness.

Finite-precision arithmetic introduces rounding noise into digital feedback filters, and mathematically equivalent realizations can differ by orders of magnitude. The state-space conditioning used to make the high-dynamic-range $\mathcal{H}_2/\mathcal{H}_\infty$ synthesis numerically tractable can also reduce rounding noise in filter realizations. A proof-of-concept scan of Caltech 40 m filters finds that LIGO's Low-Noise Biquad Form I gives the lowest noise most often, while conditioned state-space forms perform better for many individual filters.

A new experiment, GQuEST, is a tabletop interferometer designed to search for quantum-spacetime fluctuations with a photon-counting readout. Laser phase-noise sidebands in the counting band would be indistinguishable from signal photons. This thesis details how the laser filter cavity (LFC) was designed, built, commissioned, and operated to suppress these sidebands before the interferometers while transmitting the carrier. Fast laser-frequency and slow cavity-length feedback keep the carrier resonant so the passive cavity response rejects the phase-noise sidebands. Vacuum commissioning demonstrates sub-nanometer length motion; ringdowns give finesse near 4000, corresponding to about 60 dB of suppression at the readout offset.

After the interferometers, the photon-counting readout presents the complementary problem: the selected signal sideband must be isolated from residual carrier light and extraneous classical noise. The output-filter design uses a four-cavity output filter cavity (OFC) chain to transmit this sideband while rejecting the carrier. Frequency-shifted second-harmonic fields provide auxiliary locks for the individual cavities, and their lengths are staggered so unwanted longitudinal and higher-order-mode resonances do not overlap across the chain. OFC 1 was built and commissioned, with ringdowns giving finesse near 3500. Its measured root-mean-square residual motion lies just within the alignment scale required to retain 90% of the signal power through the four-cavity passband. Modeling of the full chain predicts carrier rejection above 200 dB while preserving signal-sideband transmission.

PUBLISHED CONTENT AND CONTRIBUTIONS

[1] I. A. O. MacMillan and L. P. McCuller, "Robust bilinear-noise-optimal control for gravitational-wave detectors: a mixed lqg and $\mathcal{H}_\infty$ approach", Phys. Rev. D **113**, 082001 (2026),

Author contributions: I.A.O.M. led the writing of the manuscript, performed the analysis, developed the analysis code, and created the figures.

[2] S. M. Vermeulen, T. Cullen, D. Grass, I. A. O. MacMillan, A. J. Ramirez, J. Wack, B. Korzh, V. S. H. Lee, K. M. Zurek, C. Stoughton, et al., "Photon-counting interferometry to detect geontropic space-time fluctuations with GQuEST", Phys. Rev. X **15**, 011034 (2025),

Author contributions: I.A.O.M. participated in the final realization of GQuEST presented in this paper, participated in modeling, created figures, and participated in the writing of the manuscript.

[3] D. Grass, S. M. Vermeulen, I. A. O. MacMillan, and L. McCuller, "High-frequency thermal noise in michelson interferometers", arXiv preprint, 10.48550/arXiv.2605.16710 (2026),

Author contributions: I.A.O.M. contributed to the modeling, verified mathematical derivations, created selected figures, and provided feedback and editorial support on the manuscript.

TABLE OF CONTENTS

Acknowledgements . . . iv
Abstract . . . vii
Published Content and Contributions . . . ix
Table of Contents . . . x
List of Illustrations . . . xv
List of Tables . . . xviii
List of Acronyms . . . xix

Chapter I: Introduction . . . 1
1.1 Precision Measurement as Physics . . . 1
1.2 Small Signals and Large Instruments . . . 2
1.3 Interferometry and Quantum Spacetime . . . 3
1.4 Control, Digitization, and Trustworthy Readout . . . 4
1.5 Scope of This Thesis . . . 5

I Theoretical Background 6

Chapter II: Theory . . . 7
2.1 Space-time and General Relativity . . . 7
2.2 Gravitational Waves . . . 11
2.3 A Quantum Treatment of Spacetime . . . 17
2.3.1 The Quantum Gravity Gap . . . 17
2.3.2 The Standard Quantum Limit . . . 18
2.3.3 Gravitational Effects on Length Measurements . . . 20
2.3.4 Quantum Spacetime Fluctuations . . . 21
Chapter III: Background . . . 24
3.1 Interferometers as a Probe of Spacetime . . . 25
3.2 Laser Interferometer Gravitational-Wave Observatory (LIGO) . . . 28
3.2.1 Fabry-Pérot Cavities . . . 28
3.2.2 Traveling-Wave Cavities . . . 35
3.2.3 Pound-Drever-Hall Technique . . . 36
3.2.4 Power Recycling . . . 40
3.2.5 Signal Recycling . . . 41
3.3 GQuEST: Gravity from the Quantum Entanglement of Space-Time . . . 42
3.3.1 Experimental Design . . . 44
3.3.2 Laser Phase Noise . . . 49
3.3.3 High-Frequency Thermal Noise . . . 51
3.3.4 Photon Counting . . . 55

3.3.5 Twin-Interferometer Readout . . . 59
3.4 Noise Sources . . . 61
3.4.1 Seismic Noise . . . 61
3.4.2 Quantum Noise . . . 63
3.4.3 Technical Noise . . . 65
Chapter IV: Control Theory . . . 67
4.1 Introduction . . . 67
4.2 Control Systems . . . 69
4.2.1 Transfer Functions . . . 70
4.2.2 State-Space Representation . . . 71
4.2.3 Controllability . . . 72
4.2.4 Observability . . . 73
4.3 Feedback and Interconnected Systems . . . 73
4.3.1 Interconnected Systems . . . 73
4.3.2 Feedback Systems . . . 74
4.4 Bode Plots . . . 77
4.5 Stability . . . 79
4.5.1 Routh-Hurwitz Criterion . . . 80
4.5.2 Nyquist Criterion . . . 82
4.5.3 Stability Margins . . . 83
4.6 PID Controllers . . . 85

II Optimal Control 87

Chapter V: Quantization Noise and Filter Architecture . . . 88
5.1 Quantization Noise in Filters . . . 89
5.2 Direct Form II . . . 92
5.2.1 Measuring Noise in a DFII Filter . . . 93
5.3 Low-Noise Biquad Form I . . . 94
5.3.1 Estimating Noise in a Low-Noise Biquad Form I . . . 95
5.4 Low-Noise Biquad Form II . . . 97
5.4.1 Estimating Noise in Low-Noise Biquad Form II . . . 98
5.5 State-Space Realizations . . . 99
5.6 Measurement . . . 101
5.6.1 Example: Fourth-Order Notch Filter . . . 101
5.7 Procedure for a Generalized Measurement . . . 103
5.8 Quantization-Aware Filter Realization Selection . . . 105
5.9 Summary . . . 108
Chapter VI: Optimal Controls . . . 109
6.1 Introduction . . . 110
6.2 Noise in Feedback Systems . . . 110
6.3 Riccati Equation Approach to Mixed-Sensitivity Controls . . . 111
6.3.1 State-Space System Description . . . 111
6.3.2 Calculating the $\mathcal{H}_2$ Norm . . . 113

6.3.3 Optimal LQG Controllers . . . 114
6.3.4 Mathematical Assumptions and Their Physical Interpretation 116
6.3.5 System Augmentation . . . 118
6.3.6 Calculating the $\mathcal{H}_\infty$ Norm . . . 120
6.3.7 Optimal $\mathcal{H}_\infty$ Controllers . . . 122
6.4 Lyapunov Noise Derivation . . . 123
6.4.1 Relation Between Covariance and RMS . . . 123
6.4.2 Covariance of the Output of a State-Space System . . . 125
6.4.3 State Vector Covariance Dynamics . . . 126
6.4.4 Evaluation of the Cross Terms . . . 127
6.4.5 Continuous-Time Lyapunov Equation . . . 129
6.5 Optimal LQG Controllers With a Bounded $\mathcal{H}_\infty$ Norm . . . 129
6.5.1 Controller Stability and Implementation . . . 133
6.6 Translation of the Berstein-Haddad (BH) Equations . . . 133
6.7 Alternate Noise Arrangement . . . 134
6.7.1 Actuator-Centered form . . . 135
6.7.2 Total-Noise Optimization Form . . . 136
6.7.3 Measurement-Only Optimization Form . . . 137
Chapter VII: Optimal Control of LIGO's ASC DHARDY Loop . . . 138
7.1 Background . . . 140
7.1.1 Noise from Control Systems . . . 141
7.1.2 Application to Alignment Sensing Control System . . . 145
7.2 Problem Formulation . . . 146
7.2.1 System Overview . . . 147
7.2.2 Figures of Merit . . . 149
7.2.3 Pareto Front of Multiple FOMs for Bilinear Noise . . . 152
7.3 The Bilinear Noise Metric and Spectral Convolution . . . 154
7.4 LQG Optimization . . . 156
7.4.1 LQG Optimal Controllers . . . 157
7.4.2 Properties and Limitations Apparent in LQG Controllers . . 160
7.5 Solution With Custom Stability Margins . . . 160
7.5.1 $\mathcal{H}_\infty$ Optimization Problems . . . 161
7.5.2 LQG Optimal Solutions Given an $\mathcal{H}_\infty$ Performance Bound . 162
7.5.3 Properties and Limitations of $\mathcal{H}_\infty$ Norm Limited LQG Controllers . . . 165
7.6 Comparison with AI-Informed Controller Synthesis . . . 167
7.6.1 Relation to Convex Optimization . . . 169

III Commissioning for GQuEST 173

Chapter VIII: Optical Filter Cavity Chain for Photon-Counting . . . 174
8.1 Output Filtering for Photon-Counting Readout . . . 174
8.1.1 Signal Sideband Readout . . . 175
8.1.2 Carrier-Rejection Requirement . . . 176

8.1.3 Single-Cavity Finesse and Bandwidth . . . 177
8.1.4 Four-Cavity Pass-Band Alignment . . . 178
8.2 Frequency Alignment of the OFC Chain . . . 179
8.2.1 1550 nm Carrier Rejection and Sideband Transmission . . . 180
8.2.2 775 nm Auxiliary Resonances . . . 182
8.2.3 Length Offsets Between Cavities . . . 183
8.3 Optical and Mechanical Design . . . 184
8.3.1 Bow-Tie Geometry and Cavity Mode . . . 184
8.3.2 Mirror Coatings and Optical Functions . . . 186
8.3.3 Cavity Body, Rubber Isolation, and PZT Actuation . . . 188
8.3.4 Construction of the First OFC . . . 189
8.3.5 Alignment of The Output Filter Cavities . . . 192
8.4 Auxiliary 775 nm Control System . . . 198
8.4.1 SHG and AOM Sled . . . 198
8.4.2 PDH Sensing at 775 nm . . . 200
8.4.3 1550 nm Test Lock . . . 201
8.4.4 775 nm Operating Lock . . . 202
8.4.5 AOM-Assisted Calibration . . . 202
8.4.6 Control Configurations . . . 203
8.4.7 CDS and FPGA Control . . . 204
8.5 Suppression and Higher-Order-Mode Modeling . . . 205
8.5.1 Higher-Order-Mode Generation and Transmission . . . 205
8.5.2 Effect of Varying Mirror Radius of Curvature . . . 206
8.6 OFC 1 Characterization . . . 209
8.6.1 Finesse Measurements . . . 209
8.6.2 Peak Scans and Mode Matching . . . 211
8.6.3 Locked Noise Spectra . . . 213
8.7 Integration with the Photon-Counting Readout . . . 214
8.7.1 Cascaded OFC Operation . . . 215
8.7.2 Interface to the SNSPD Path . . . 215
8.7.3 Background and Leakage Checks . . . 216
8.7.4 Upgrades and Operating Upgrades . . . 216
8.8 Summary . . . 217
Chapter IX: Laser Filter Cavity . . . 219
9.1 Design Requirements . . . 220
9.2 Filter Response . . . 222
9.3 Optical Layout . . . 223
9.3.1 Input Optics . . . 224
9.3.2 Cavity Design . . . 226
9.3.3 Output Optics . . . 230
9.4 Vacuum System . . . 230
9.5 Cleaning and Vacuum Bakeout . . . 232
9.6 Spacer and Mechanical Design . . . 233
9.6.1 Spacer Material and Geometry . . . 234
9.6.2 Thermal Expansion and Drift . . . 236

9.6.3 Spacer Mode Shapes . . . 238
9.7 Vibration-Isolation and Damping . . . 239
9.7.1 Mirror Mount Vibration Damping . . . 243
9.8 Finesse Measurements . . . 244
9.9 Cavity Scan Measurements . . . 248
9.9.1 Cavity Scan Calibration . . . 248
9.9.2 Cavity Scan for Higher-Order Modes . . . 251
9.9.3 Linewidth Measurements . . . 251
9.10 Control Scheme . . . 253
9.11 Commissioning Measurements . . . 257
9.11.1 Transfer Functions . . . 257
9.11.2 Noise Spectra . . . 260
9.12 Future Laser Phase-Noise Measurements . . . 261
9.13 Integration with GQuEST . . . 263
9.14 Summary . . . 264

Bibliography . . . 266

LIST OF ILLUSTRATIONS

Number *Page*

2.1 Gravitational-Wave Effect on a Ring of Particles 16
2.2 Geontropic Quantum Spacetime Fluctuation Signal 23
3.1 Basic Michelson Interferometer 25
3.2 Interferometer Configurations 29
3.3 Fabry-Pérot Cavity 30
3.4 Resonances of a Fabry-Pérot Cavity 31
3.5 Hermite-Gaussian Mode Cross Sections 34
3.6 Bow-Tie Cavity 35
3.7 PDH Locking Technique 37
3.8 PDH Locking Error Signal 40
3.9 Traveling-Wave GQuEST Layout 46
3.10 GQuEST Output Filter Cavity Chain 47
3.11 GQuEST Noise Budget 48
3.12 Laser Phase Noise from Spontaneous Emission 49
3.13 High-Frequency Thermal Noise in GQuEST 56
3.14 Traveling-Wave Twin-Interferometer Photon-Counting Readout . . . 60
3.15 LIGO's Noise Sources in O4 62
4.1 SISO System Block Diagram 70
4.2 Double Pendulum 75
4.3 Interconnected System Modeling a Double Pendulum 75
4.4 Basic Feedback System 76
4.5 Basic Feedback System with a Controller 76
4.6 Bode Plot for a Simple Harmonic Oscillator 78
4.7 Stable, Marginal, and Unstable Systems 81
4.8 Nyquist Plots for Stable, Marginal, and Unstable Systems 82
4.9 Bode Plot with Stability Margins 84
5.1 IEEE-754 Single-Precision Bit Layout 88
5.2 Quantizer Block Diagram 90
5.3 Rounding Quantizer Transfer Curve 90
5.4 Quantizer Squared Error Curve 91
5.5 Direct-Form II Filter Block Diagram 93

5.6 Low-Noise Biquad Form I Filter Block Diagram 95
5.7 Low-Noise Biquad Form I with Quantizers 96
5.8 Simplified Low-Noise Biquad Form I Quantizers 96
5.9 Equivalent Quantizer Placement at a Split 96
5.10 Low-Noise Biquad Form II Block Diagram 98
5.11 Low-Noise Biquad Form II Intermediate Quantizers 99
5.12 Low-Noise Biquad Form II Final Quantizers 99
5.13 Notch Filter Time-Domain Response 102
5.14 Notch Filter Output ASD 103
5.15 Notch Filter Quantization-Noise Spectrum 104
6.1 Input-Output Block Diagram for Optimal Control 112
6.2 Augmented System Layout with Modified Plant 118
6.3 Augmented System Layout with Actuator and TN Error Points 135
7.1 Augmented System Layout for Optimal Control 147
7.2 Measurement and Environmental Noise in ASC DHARDY DOF . . . 148
7.3 Figures of Merit for ASC DHARDY DOF 151
7.4 RMS Performance of Optimal LQG Controllers 158
7.5 Open-Loop Gain of Optimal LQG Controllers 159
7.6 Closed-Loop Gain of Optimal LQG Controllers 159
7.7 RMS Performance of $\mathcal{H}_\infty$-Bounded LQG Controllers 164
7.8 Open-Loop Gain of $\mathcal{H}_\infty$-Bounded LQG Controllers 165
7.9 Closed-Loop Gain of $\mathcal{H}_\infty$-Bounded LQG Controllers 166
7.10 Stability Margins of $\mathcal{H}_\infty$-Bounded LQG Controllers 166
7.11 Noise Performance of Optimal and DLS Controllers 168
7.12 Open-Loop Comparison of Convex and Algebraic Controllers 171
8.1 Signal-Sideband Transmission Through the OFC Chain 176
8.2 Frequency Alignment of the Output Filter Cavities 181
8.3 Single Output Filter Cavity Optical Layout 185
8.4 Output Filter Cavity Beam Radius 185
8.5 Output Filter Cavity Accumulated Gouy Phase 186
8.6 Output Filter Cavity Higher-Order-Mode Spacing 187
8.7 Output Filter Cavity 1 Diagnostic Optical Layout 190
8.8 SHG and AOM Preparation of the OFC Control Fields 199
8.9 OFC 1 Dual-Resonant Testing Control Layout 203
8.10 OFC 1 Operating Control Configuration 204
8.11 Higher-Order-Mode Rejection in the Output Filter Cavities 207

8.12 Radius-of-Curvature Variation of OFC HOM Leakage 208
8.13 OFC 1 1550 nm Ringdown Measurement 210
8.14 OFC 1 775 nm Auxiliary Linewidth Measurement 211
8.15 OFC 1 1550 nm Cavity Scan Transmission Peak 212
8.16 OFC 1 775 nm Locked Error Spectrum 214
9.1 Laser-Noise Suppression from the LFC and PRC 221
9.2 Laser Filter Cavity Optical Layout and Control Scheme 225
9.3 Laser Filter Cavity Beam Radius 227
9.4 Laser Filter Cavity Accumulated Gouy Phase 227
9.5 Laser Filter Cavity Higher-Order-Mode Spacing 228
9.6 Laser Filter Cavity Vacuum System 231
9.7 Laser Filter Cavity Spacer Render 234
9.8 FEA Spacer Low-Frequency Modes 238
9.9 Viton-Ball Damping for the Laser Filter Cavity 239
9.10 Laser Filter Cavity M4 Mirror-Mount Resonance 243
9.11 Laser Filter Cavity Ringdown Measurements 246
9.12 Laser Filter Cavity Ringdown Finesse Distribution 247
9.13 Laser Filter Cavity Nonlinear Scan Calibration 250
9.14 Laser Filter Cavity Mode Scans 252
9.15 Laser Filter Cavity Linewidth Scan 253
9.16 Laser Filter Cavity PZT Open-Loop Transfer Function 258
9.17 Laser Filter Cavity PZT Plant and Closed-Loop Suppression 259
9.18 Laser Filter Cavity Displacement Noise Spectra 261
9.19 Delay-Line Interferometer for Laser Phase-Noise Measurements . . . 262

LIST OF TABLES

Number *Page*

3.1 Current GQuEST Design Parameters 44
3.2 GQuEST Readout and Sensitivity Parameters 45
5.1 Lowest-Noise Realizations for the Caltech 40m Filters 107
8.1 Output Filter Cavity Design Values 180
8.2 Output Filter Cavity Length Offsets 183
8.3 Output Filter Cavity Mirror Coatings 187
8.4 Varying Radius-of-Curvature Gouy Phases 208
9.1 Laser Filter Cavity Reference Parameters 222
9.2 Laser Filter Cavity Optical Design Parameters 229
9.3 LFC Spacer Material and Geometric Properties 235
9.4 Viton-Ball Support Parameters 242
9.5 Summary of Laser Filter Cavity Finesse Derived Values 248
9.6 Laser Filter Cavity Piezo Loop Parameters 259

LIST OF ACRONYMS

ADC analog-to-digital converter

AOM acousto-optic modulator

ASC alignment sensing control

ASD amplitude spectral density

BBH binary black hole

BH Berstein-Haddad

BNS binary neutron star

BSBTN beam-splitter Brownian thermal noise

BTN Brownian thermal noise

CBTN coating Brownian thermal noise

CDS Control and Data acquisition System

DARM differential-arm length

DFII direct form II

DI deionized

DLS deep loop shaping

DOF degree of freedom

EOAM electro-optic amplitude modulator

EOM electro-optic modulator

EOPM electro-optic phase modulator

FDT fluctuation-dissipation theorem

FOM figure of merit

FPGA field-programmable gate array

FPMI Fabry-Pérot Michelson interferometer

FSR free spectral range

FWHM full width at half maximum

GQuEST Gravity from the Quantum Entanglement of Space Time

GW gravitational wave

HOM higher-order mode

LFC laser filter cavity

LIGO the Laser Interferometer Gravitational-Wave Observatory

LMI linear matrix inequality

LQG linear quadratic Gaussian

LTI linear time-invariant

MIMO multiple-input-multiple-output

MOTH Mode-matching Optimization THing

OFC output filter cavity

PBS polarizing beam splitter

PD photodiode

PDH Pound-Drever-Hall

PID proportional-integral-derivative

PRC power-recycling cavity

PSD power spectral density

PZT piezoelectric transducer

QST quantum spacetime

RMS root mean squared

ROC radius of curvature

SBTN substrate Brownian thermal noise

SHG second-harmonic generation

SISO single-input-single-output

SNR signal-to-noise ratio

SNSPD superconducting nanowire single-photon detector

SQL standard quantum limit

STR substrate thermorefractive noise

TEM transverse electromagnetic mode

UGF unity gain frequency

VCO voltage-controlled oscillator

ZPK zero-pole-gain

Chapter 1

INTRODUCTION

"To strive, to seek, to find, and not to yield"

— Alfred, Lord Tennyson, *Ulysses*

At the limits of precision measurement, distinguishing a physical signal from the response of the instrument is the central experimental problem. The target effect may be far smaller than disturbances within the apparatus, and each part of the experiment can alter what appears at the readout. Experiments at this level of precision often demand more than a single researcher or small team can provide. Collaborative teams make it possible to characterize the full instrument and check its response independently. A signal becomes credible only when its origin can be separated from the apparatus used to measure it.

Gravity from the Quantum Entanglement of Space Time (GQuEST) is a tabletop interferometric experiment designed to search for or constrain quantum spacetime (QST) fluctuations using a photon-counting readout [1]. the Laser Interferometer Gravitational-Wave Observatory (LIGO) is a kilometer-scale interferometer that made the first direct detections of gravitational waves (GWs) [2]. Both require the detector response and technical backgrounds to be understood before a small signal can be interpreted.

1.1 Precision Measurement as Physics

Physics advances by turning questions about nature into quantities that can be measured. That translation is rarely simple. A theory may predict a strain, a phase shift, a photon flux, or a displacement spectrum, but an experiment records data with which to test that prediction. The connection between the two is the measurement model. It describes how the physical quantity enters the apparatus, how the apparatus responds, how noise is budgeted, and how the recorded data is converted back into the units of the theory.

The careful construction of that measurement model is a subfield of physics in its own right. It includes metrology, calibration, noise modeling, system identification,

statistical inference, and the design of instruments whose response is consistent enough to support the desired conclusion. A measurement of length becomes meaningful after the length actuator has been calibrated. A measurement of phase depends on knowing the optical gain and demodulation phase. A measurement of photon flux depends on accounting for losses, backgrounds, and detector response.

This is not a modern departure from physics. Many early celebrated contributions were built around the invention of a measurement rather than the writing of a theoretical speculation. Michelson's Nobel Prize was awarded for optical precision instruments and the metrological investigations made with them [3]. Kamerlingh Onnes built the low-temperature apparatus that made liquid helium experimentally accessible [4], opening the regime in which superconductivity was first observed in mercury [5]. Millikan's measurement of the elementary charge depended on the controlled motion of charged drops in a known electric field [6, 7], and Wilson's cloud chamber turned the invisible paths of ionizing particles into visible tracks [8]. A quarter of the 2017 Nobel Prize in physics for the discovery of GWs was awarded for the theoretical work [9], while the rest was awarded for the experimental work that made the measurement possible. In each case, the apparatus was essential to making the phenomenon measurable.

The history of GW detection makes this point especially clear. A passing gravitational wave produces an extraordinarily small change in the relative lengths of the interferometer arms, and that change must be inferred through optical fields, suspended mirrors, feedback controls, and a calibrated readout [2]. Quantum fluctuations of light, thermal motion of the mirrors, and environmental motion all enter the same measurement model as the desired signal. LIGO's achievement was to reduce and characterize these effects well enough that a transient strain could be distinguished from the detector's own noise. The discovery was credible because the apparatus was not only an engineering achievement but also a carefully characterized physical instrument, operating near the limits of precision measurement.

1.2 Small Signals and Large Instruments

As experimental physics has moved toward weaker effects, instruments have become more layered. The number of subsystems grows because each dominant background requires its own mitigation strategy. Optical cavities increase phase sensitivity or reject unwanted frequencies. Vacuum systems suppress acoustic and refractive-index noise. Mechanical isolation reduces seismic coupling. Feedback loops hold

resonances in place. Low-noise electronics preserve signal information. Numerical models predict where the remaining backgrounds should appear. Each subsystem must work, but the calibration of each subsystem must also be good enough that its imperfections can be propagated into the final uncertainty.

This layered structure changes how a signal must be checked against its backgrounds. In a simple experiment, the result may be limited by whether an effect is visible above a background. In a precision experiment, the result is often limited by whether the background has been understood well enough to be subtracted, bounded, or distinguished from the signal. The question moves beyond "is there excess power?" to "is the excess power in the right channel, with the right coherence, with the right dependence on alignment, frequency, and operating point?" A measurement that answers these questions is much stronger than a measurement that records only a large response.

Instrument design also determines which signals reach the readout. A cavity can make one optical frequency visible while rejecting another. A feedback loop can hold a system near the operating point where its response is linear. A photon-counting readout can change the observable from a field quadrature to a flux of quanta. These choices determine which degrees of freedom are accessible, which backgrounds are rejected, and which physical limits remain.

Many modern experiments are designed to operate in this background-limited regime. Since some disturbances are set by fundamental physical limits, the goal is to arrange the experiment so that the remaining backgrounds are in known channels, at known frequencies, or below the scale required to test a theory. This allows a small effect to be detected or constrained.

1.3 Interferometry and Quantum Spacetime

Laser interferometry offers a particularly sensitive way to measure small displacements, and it is the basis for both LIGO and GQuEST. In LIGO, optical phase is used to infer the differential arm-length change produced by a passing gravitational wave [2]. In GQuEST, the same basic sensitivity to optical phase is used for a different target: stochastic QST fluctuations at MHz-scale frequencies in a tabletop apparatus [1]. In both cases, the physical effect is inferred from an optical signal that must be filtered, calibrated, and interpreted.

The GQuEST readout makes this especially demanding. Because the expected signal is stochastic rather than a deterministic waveform, the experiment uses photon

counting to measure signal-sideband power directly [1, 10]. That choice shifts much of the work onto a chain of optical filters. The laser filter cavity (LFC) suppresses laser noise before the interferometer, while the output filter cavities (OFCs) select the photon-counting band and reject carrier light and noise at the output. The signal becomes measurable only after the apparatus response has been measured, modeled, and checked against independent evidence. The same is true for LIGO, where the calibration of the interferometer and the control loops are essential to interpreting the data as a statement about astrophysical strain.

1.4 Control, Digitization, and Trustworthy Readout

Feedback control is another place where the instrument becomes part of the measurement. A loop holds an interferometer near its operating point, but it also suppresses disturbances, reinjects sensor noise, and couples auxiliary motion into the science readout. In LIGO, this matters most at low frequencies, where seismic motion and control noise can obscure astrophysical strain. This noise sets the low-frequency sensitivity of the instrument, which in turn sets the distance to which LIGO can detect mergers and the number of detections that can be made.

The design of better controllers can recover low-frequency sensitivity. Reducing control noise can increase the distance to which LIGO can detect mergers and, through the larger observable volume, increase the number of detections. It also improves access to low-frequency binary black hole (BBH) merger physics, where useful signal content can otherwise be hidden by seismic-noise-driven control artifacts. Optimal-control methods provide a way to compare designs against a model-based performance boundary and estimate how much astrophysical reach is being lost to these loops.

Digitization adds the same requirement in a different form. Finite precision, quantization, sampling, and filtering can set a floor in a small-signal measurement, so they belong in the measurement model. The common goal is a trustworthy readout: an instrument whose imperfections have been measured, modeled, or bounded well enough that the remaining data can be interpreted honestly.

These are real limitations in the current generation of GW detectors, and they will be more important for future instruments. The next generation of ultra sensitive interferometers will have improved sensitivity at low frequencies, but they will also have more complex control systems. Improvements in the control loops will be necessary to reach the full potential of these instruments, and the methods

implementing these improvements must be optimized to gain the most reach for the astrophysical signals.

1.5 Scope of This Thesis

The chapters that follow develop this measurement program from theory to instrument characterization. Chapter 2 reviews the theoretical setting: general relativity, interferometric response, and the motivation for possible QST fluctuations. Chapter 3 connects that theory to practical interferometers, with emphasis on the optical, thermal, and photon-counting ideas that motivate GQuEST. These chapters set the scale of the measurement problem and explain why the experiment is forced into a regime where filtering, calibration, and background control are central.

The middle part of the thesis addresses measurement limits and control. Chapter 5 treats quantization and numerical precision as physical limitations in digital measurement chains. Chapters 4, 6 and 7 develop and apply control methods for precision interferometry, including comparisons between optimal controllers and other approaches. They set a limit on how LIGO's detection range can be optimized by improvements to control loops and the method sets the limit for the low frequency performance of LIGO.

The final part of the thesis focuses on the GQuEST filtering systems. Chapter 9 describes the laser filter cavity, whose role is to suppress laser noise before it enters the interferometer. The chapter lays out the system's importance in the greater experiment and gives the measurements for its commissioning such as: its mode geometry, linewidth, vibration isolation, lock spectra, and mechanical resonances. Chapter 8 describes the output filter cavities, which define the photon-counting sideband and reject the carrier and then goes on to present the construction and commissioning of the first cavity.

Searches for GWs and possible QST fluctuations press interferometers against the limits of what they can resolve. At those limits, better control and optical filtering determine whether a faint physical signal can be distinguished from the instrument's own response. The methods developed here advance those detection efforts and apply more broadly wherever an experiment's sensitivity limits the physics it can test.

Part I

Theoretical Background

Chapter 2

THEORY

"Man must rise above the Earth, to the top of the atmosphere and beyond, for only thus will he fully understand the world in which he lives"

— Plato, *Phaedo*[a]

[a] Paraphrased.

2.1 Space-time and General Relativity

Civilization's understanding of gravity has evolved significantly over the centuries. Since before the dawn of modern science, gravity has been recognized as a fundamental force that governs the motion of objects both on Earth and in the cosmos. Humanity's earliest interactions with the effects of the force of gravity on a large scale can generally be traced back to our observations of the night sky and the motions of celestial bodies. Ancient civilizations such as the Babylonians, Greeks, and Mayans meticulously recorded the movements of the stars, planets, and the Moon, leading to early models of the cosmos that sought to explain these observations. These early models, whose goal was to predict celestial events such as eclipses and planetary positions, are considered some of the first steps toward a scientific understanding of gravity. Careful observations by astronomers such as Nicolaus Copernicus, Johannes Kepler, and Galileo Galilei in the 16th and 17th centuries laid the groundwork for a more systematic study of gravity. Copernicus proposed a heliocentric model of the solar system, challenging the long-held geocentric view. Kepler formulated his laws of planetary motion, describing how planets orbit the Sun in elliptical paths. Galileo's experiments with falling objects provided empirical evidence that contradicted Aristotelian physics, demonstrating that all objects fall at the same rate regardless of their mass.

The beginning of the more modern and quantitative description of gravity arrived with Isaac Newton. When Isaac Newton formulated his theory of gravity in the 17th century, he characterized gravity as a force that acted instantaneously at a distance

between two masses. His revolutionary theory successfully explained the motion of falling objects on Earth and interactions of celestial bodies. His equation of universal gravitation,

$$F = G\frac{m_1 m_2}{r^2}, \tag{2.1}$$

described the attractive force F between two masses m_1 and m_2 separated by a distance r, where G is the gravitational constant. It was remarkably successful in explaining the objects on the order of the mass of planets in the solar system and their motions.

Newton's theory remained the dominant description of gravity for over two centuries, until the early 20th century when Albert Einstein developed his special theory and then general theory of relativity [11, 12]. General relativity revolutionized our understanding of gravity by describing it not as a force, but as the curvature of spacetime caused by mass and energy. In this framework, massive objects cause spacetime to curve, and this curvature affects the motion of other objects, which move along the curved paths known as geodesics. While Newton's classical description of gravity was sufficient for many practical purposes of the time, general relativity provided a more accurate and comprehensive description of gravitational phenomena, especially in extreme conditions such as near massive objects like black holes or at cosmological scales.

The notation and conventions used in the following discussion follow the treatment of general relativity in Carroll [13]. In a four-dimensional coordinate system for spacetime, three coordinates represent space and one represents time. Using the Minkowski metric for flat spacetime, the spacetime interval between two events can be expressed as

$$\mathrm{d}s^2 = -c^2\mathrm{d}t^2 + \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2, \tag{2.2}$$

where $\mathrm{d}s^2$ is the spacetime interval, c is the speed of light in a vacuum, $\mathrm{d}t$ is the time difference between the two events, and $\mathrm{d}x$, $\mathrm{d}y$, and $\mathrm{d}z$ are the spatial separations in the three spatial dimensions. This formulation combines space and time into a single four-dimensional continuum, where the geometry of spacetime is influenced by the presence of mass and energy. Using tensors, the Minkowski metric tensor

$\eta_{\mu\nu}$ can be represented as

$$\eta_{\mu\nu} = \begin{pmatrix} -c^2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{2.3}$$

where the indices μ and ν run from 0 to 3, corresponding to the time and three spatial dimensions.

Geometric units are adopted here,[1] so that $c = 1$. The Einstein summation convention is also used: an index that appears once as a subscript and once as a superscript is summed over (e.g., $a_\mu b^\mu = \sum_\mu a_\mu b^\mu$). With these conventions, the spacetime interval can be written compactly using the metric tensor as

$$\mathrm{d}s^2 = g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu, \tag{2.4}$$

where $g_{\mu\nu}$ is the metric tensor that describes the geometry of spacetime in the presence of mass and energy.

One of the key insights of general relativity is that mass and energy tell spacetime how to curve, and the curvature of spacetime tells mass and energy how to move. Before general relativity, special relativity considered time and space were considered separate entities and thus, a free-falling mass would fall in a straight line through space while time progressed uniformly. However, in general relativity, the presence of mass and energy curves spacetime, and free-falling objects follow geodesics. Assuming that the mass does not accelerate along its path, the geodesic equation can be expressed as

$$\frac{\mathrm{d}^2 x^\rho}{\mathrm{d}\tau^2} + \Gamma^\rho_{\mu\nu} \frac{\mathrm{d}x^\mu}{\mathrm{d}\tau} \frac{\mathrm{d}x^\nu}{\mathrm{d}\tau} = 0, \tag{2.5}$$

where $\mathrm{d}\tau^2 = -\mathrm{d}s^2$, τ is the proper time along the object's worldline, and $\Gamma^\rho_{\mu\nu}$ are the Christoffel symbols that describe how the coordinate system changes in curved spacetime. Christoffel symbols are defined in terms of the metric tensor as

$$\Gamma^\rho_{\mu\nu} = \frac{1}{2} g^{\rho\sigma} \left(\partial_\mu g_{\nu\sigma} + \partial_\nu g_{\mu\sigma} - \partial_\sigma g_{\mu\nu} \right), \tag{2.6}$$

[1] Geometric units are a system of natural units in which the speed of light c and the gravitational constant G are set to 1. This simplifies many equations in general relativity by eliminating these constants. This section is the only one where geometric units are used.

where σ is a new tensor index, and $g^{\rho\sigma}$ is the inverse metric tensor $g_{\rho\sigma}$. For massless particles such as photons, which travel at the speed of light, the geodesic equation still applies, but becomes

$$\frac{\mathrm{d}^2 x^\rho}{\mathrm{d}\lambda^2} + \Gamma^\rho_{\mu\nu} \frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda} \frac{\mathrm{d}x^\nu}{\mathrm{d}\lambda} = 0, \tag{2.7}$$

where λ is an affine parameter along the null geodesic.

To describe the curvature of spacetime the Riemann curvature tensor is used. The Riemann curvature tensor assigns a tensor to each point in spacetime that describes how much spacetime is curved or deviates from the standard Euclidean space at that point. In terms of the Christoffel symbols, the Riemann curvature tensor is defined as

$$R^\sigma_{\rho\mu\nu} = \partial_\mu \Gamma^\sigma_{\nu\rho} - \partial_\nu \Gamma^\sigma_{\mu\rho} + \Gamma^\sigma_{\mu\gamma} \Gamma^\gamma_{\nu\rho} - \Gamma^\sigma_{\nu\gamma} \Gamma^\gamma_{\mu\rho}, \tag{2.8}$$

where γ is a new tensor index. The Riemann curvature tensor contains all the information about the curvature of spacetime and can be used to derive other important curvature tensors, such as the Ricci tensor and the Ricci scalar, which are contractions of the Riemann curvature tensor. The Ricci tensor is defined as

$$R_{\mu\nu} = R^\alpha_{\mu\alpha\nu}. \tag{2.9}$$

The Ricci scalar is defined as

$$R = g^{\mu\nu} R_{\mu\nu}. \tag{2.10}$$

Einstein's field equations describe this relationship between mass-energy and the curvature of spacetime as

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} = 8\pi T_{\mu\nu}, \tag{2.11}$$

where $G_{\mu\nu}$ is the Einstein tensor, and $T_{\mu\nu}$ is the stress-energy tensor that describes the distribution of momentum and energy in spacetime. Analytical solutions of the Einstein field equation exist for certain idealized scenarios, such as the Schwarzschild solution for non-rotating black holes [14], the Kerr solution for rotating black holes [15], and the Friedmann-Lemaître-Robertson-Walker metric for a homogeneous and isotropic expanding universe [16–19]. One such analytical solution is the Minkowski metric for flat spacetime used in special relativity. By leveraging the Minkowski metric as a background, small perturbations to the metric tensor can be studied to analyze weak gravitational fields and GWs.

General relativity has been extensively tested and confirmed through various experiments and observations, including the bending of light around massive objects (gravitational lensing) [20], the gravitational redshift of light [21], the perihelion precession of Mercury's orbit [22], and the first direct detection of gravitational waves by LIGO [2].

2.2 Gravitational Waves

Where Newtonian gravity describes gravity as a force acting at a distance, general relativity describes gravity as the curvature of spacetime caused by mass and energy. One of the consequences of the Newtonian description of gravity is that changes in the gravitational field propagate instantaneously. If the Sun were to suddenly disappear, the Earth would immediately feel the change in gravitational pull even though the light from the Sun would take approximately eight minutes to reach Earth. The inhabitants of Earth would instantly feel the effects of the Sun's disappearance but would not see the Sun go dark for eight minutes. This instantaneous action at a distance is a fundamental aspect of Newtonian gravity and is in direct contrast to the principles of special relativity, which state that no information or influence can travel faster than the speed of light. To resolve this inconsistency, Einstein's general theory of relativity predicts that changes in the gravitational field propagate through spacetime at the speed of light. These propagating changes manifest as GWs.

Spacetime can be conceptualized using a classic and intuitive model of a stretched rubber sheet representing spacetime. When a baseball, representing a massive astrophysical object such as a planet or star, is placed on the sheet, it creates a depression or curvature in the sheet. Other objects placed on the sheet will move along the curved paths created by this depression, which is analogous to how gravity influences the motion of objects in the universe. If the baseball is suddenly moved or disturbed, it creates ripples in the rubber sheet that propagate outward from the point of disturbance [23–25]. These ripples are analogous to gravitational waves.[2]

Gravitational waves are ripples in the fabric of spacetime that travel at the speed of light, carrying energy away from accelerating masses. They are produced by some

[2] One of the best demonstrations of GWs that resonates with the public leverages this rubber sheet analogy. The GWs are created by attaching two spheres about two inches apart on a holder that looks like an uppercase Y. The lower part of the holder fits into a cordless drill. When the drill is turned on, the two spheres spin around each other rapidly. When the two spheres are placed in contact with the rubber sheet, they create ripples that propagate outward on the sheet. This demonstration effectively illustrates how massive objects in motion can create disturbances in spacetime that propagate as gravitational waves.

of the most violent and energetic processes in the universe, such as the collision and merger of black holes or neutron stars, supernovae explosions, and even the rapid rotation of asymmetric neutron stars. As these massive objects accelerate, they create disturbances in spacetime that propagate outward as gravitational waves. These waves cause tiny oscillations in the distances between objects as they pass through space. The effect of a passing gravitational wave on a ring of test masses can be visualized as a stretching and squeezing of the ring in perpendicular directions, as illustrated in Fig. 2.1.

One way to develop a simplified model for gravitational waves is to utilize the most basic analytical solutions for the Einstein field equations presented at the end of Section 2.1 by considering small perturbations to the flat spacetime metric. This approach is known as the weak field approximation or linearized gravity.

In this weak field approximation, the metric tensor, $g_{\mu\nu}$, can be expressed as a small perturbation, $h_{\mu\nu}$, from the flat spacetime metric, $\eta_{\mu\nu}$, as

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \tag{2.12}$$

where $|h_{\mu\nu}| \ll 1$. This approximation is valid in regions of spacetime where the gravitational field is weak, such as far from massive objects. In these areas, the effects of gravity can be treated as small perturbations to the flat spacetime metric, allowing for simpler calculations and analyses of gravitational phenomena. This allows the simplification of the Christoffel symbols in Eq. 2.6 to

$$\Gamma^{\rho}_{\mu\nu} = \frac{1}{2}\left(\partial_\mu h^{\rho}_{\nu} + \partial_\nu h^{\rho}_{\mu} - \partial^{\rho} h_{\mu\nu}\right) + O(h^2), \tag{2.13}$$

where $O(h^2)$ represents higher-order terms in the perturbation $h_{\mu\nu}$. The Riemann curvature tensor in Eq. 2.8 can be approximated as

$$R^{\sigma}{}_{\rho\mu\nu} = \frac{1}{2}\left(\partial_\rho\partial_\mu h^{\sigma}_{\nu} + \partial^{\sigma}\partial_\nu h_{\rho\mu} - \partial_\rho\partial_\nu h^{\sigma}_{\mu} - \partial^{\sigma}\partial_\mu h_{\rho\nu}\right) + O(h^2). \tag{2.14}$$

The Ricci tensor in Eq. 2.9 can be approximated as

$$R_{\mu\nu} = \frac{1}{2}\left(\partial_\sigma\partial_\mu h^{\sigma}_{\nu} + \partial_\sigma\partial_\nu h^{\sigma}_{\mu} - \partial^{\sigma}\partial_\sigma h_{\mu\nu} - \partial_\mu\partial_\nu h^{\sigma}_{\sigma}\right) + O(h^2) \tag{2.15}$$

and the Ricci scalar in Eq. 2.10 becomes

$$R = \partial^{\mu}\partial^{\nu} h_{\mu\nu} - \partial^{\mu}\partial_\mu h^{\nu}_{\nu} + O(h^2). \tag{2.16}$$

With these approximations, the Einstein field equations, given in Eq. 2.11, can be linearized to describe the propagation of gravitational waves in weak gravitational fields. To the first order in $h_{\mu\nu}$, the Einstein field equations become

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} \tag{2.17}$$

$$= \frac{1}{2}\left(\partial_\sigma\partial_\mu h^\sigma_\nu + \partial_\sigma\partial_\nu h^\sigma_\mu - \partial_\mu\partial_\nu h - \Box h_{\mu\nu} - \eta_{\mu\nu}\partial_\rho\partial_\sigma h^{\rho\sigma} + \eta_{\mu\nu}\Box h\right), \tag{2.18}$$

where

$$h = \eta^{\mu\nu}h_{\mu\nu} = h^\mu_\mu \tag{2.19}$$

and

$$\Box = \eta_{\mu\nu}\partial_\mu\partial_\nu = \partial^\mu\partial_\mu \tag{2.20}$$

is the d'Alembertian operator. The linearized Einstein field equations reduce further by defining the trace-reversed perturbation

$$\bar{h}_{\mu\nu} = h_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}h, \tag{2.21}$$

leading to

$$G_{\mu\nu} = \frac{1}{2}\left(\partial^\sigma\partial_\nu\bar{h}_{\mu\sigma} + \partial^\sigma\partial_\mu\bar{h}_{\nu\sigma} - \Box\bar{h}_{\mu\nu} - \eta_{\mu\nu}\partial^\sigma\partial^\rho\bar{h}_{\sigma\rho}\right). \tag{2.22}$$

The next step in simplifying linearized Einstein field equations is to choose a gauge condition. A common choice, and most practical given the detection methods used today, is the Lorenz gauge, defined by

$$\partial^\nu\bar{h}_{\mu\nu} = 0, \tag{2.23}$$

which simplifies the linearized Einstein field equations to

$$G_{\mu\nu} = -\frac{1}{2}\Box\bar{h}_{\mu\nu}. \tag{2.24}$$

With the Lorenz gauge condition applied, the linearized Einstein field equations in Eq. 2.24 describe how small perturbations to the metric tensor propagate as GWs in weak gravitational fields. In vacuum, the stress-energy tensor vanishes:

$$T_{\mu\nu} = 0. \tag{2.25}$$

This further simplifies Eq. 2.24 to

$$\Box \bar{h}_{\mu\nu} = 0. \tag{2.26}$$

From Eq. 2.26, it is evident that the trace-reversed perturbation $\bar{h}_{\mu\nu}$ satisfies the wave equation in vacuum, indicating that GWs propagate at the speed of light. The wave equation in Eq. 2.26 can be solved using plane-wave solutions of the form

$$\bar{h}_{\mu\nu} = A_{\mu\nu} e^{ik^\alpha x_\alpha}, \tag{2.27}$$

where $A_{\mu\nu}$ is a constant 4×4 amplitude tensor, k_α is the wave vector which satisfies $k^\alpha k_\alpha = 0$, and x_α are the spacetime coordinates. In Lorenz gauge, the amplitude tensor is further constrained by $k^\mu A_{\mu\nu} = 0$. This solution describes GWs as transverse waves with two independent polarization states commonly referred to as the "plus" and "cross" polarizations.

To show these polarization states more explicitly, the analysis can further be simplified by choosing the transverse-traceless (TT) gauge, which imposes the additional conditions on the perturbation:

$$\bar{h}_{\mu 0} = 0, \tag{2.28}$$

$$\partial^\mu \bar{h}_{\mu\nu} = \partial^i \bar{h}_{ij} = 0, \tag{2.29}$$

$$\eta_{\mu\nu} \bar{h}_{\mu\nu} = \bar{h}^i_i = 0. \tag{2.30}$$

The perturbation $\bar{h}_{\mu\nu}$ reduces to a purely spatial tensor with only two independent components, corresponding to the two polarization states of gravitational waves. The metric perturbation in the TT gauge can be expressed as

$$h^{\mathrm{TT}}_{\mu\nu} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & h_+ & h_\times & 0 \\ 0 & h_\times & -h_+ & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{2.31}$$

Here, h_+ and $h_\times$ represent the amplitudes of the plus and cross polarizations, respectively. These polarization states describe how gravitational waves stretch and squeeze spacetime in perpendicular directions as they propagate.

Consider how gravitational waves interact with matter by predicting the local spacetime interval in Eq. 2.4 using the TT gauge metric perturbation. For a gravitational

wave propagating in the z-direction, the perturbed spacetime interval becomes

$$\begin{aligned} \mathrm{d}s^2 = -\mathrm{d}t^2 + \mathrm{d}z^2 + 2h_\times \cos(kz - \omega t)\mathrm{d}x\mathrm{d}y \\ + [1 + h_+ \cos(kz - \omega t)]\, \mathrm{d}x^2 \\ + [1 - h_+ \cos(kz - \omega t)]\, \mathrm{d}y^2. \end{aligned} \tag{2.32}$$

Consider a photon traveling between two test masses separated by a distance L along the x-axis. Since this photon is only probing the x-direction, set $\mathrm{d}y = \mathrm{d}z = 0$ and since the photon is massless it must travel on a null geodesic, thus $\mathrm{d}s^2 = 0$. This leads to the equation

$$\mathrm{d}t^2 = [1 + h_+ \cos(kz - \omega t)]\, \mathrm{d}x^2. \tag{2.33}$$

Here it is seen that if two test masses are placed along the x-axis only a plus-polarized gravitational wave will affect the distance between them. The same calculation can be done for a photon traveling between two test masses along the y-axis showing that only a plus-polarized gravitational wave will affect the distance between them as well. However, if the two test masses are placed along the line $y = x$, it can be shown that only a cross-polarized gravitational wave will affect the distance between them. As seen in Fig. 2.1, a passing gravitational wave causes the ring of test masses to alternately stretch and squeeze in perpendicular directions, demonstrating the characteristic effect of gravitational waves on spacetime.

Even with the mathematical framework in place, it might seem like their fluctuations are easily detectable. However, the effects of GWs were so small that almost everyone doubted they could ever be detected, and the technology to do so did not exist for many decades. The stretching and squeezing effect of gravitational waves on spacetime is incredibly small, making their detection extremely challenging. Although no direct source of the sentiment exists, it is widely reported that Einstein himself, who first assumed their existence was merely a crude first order approximation [24, 25], thought that the effects of the gravitational waves would be too small to ever be observed [26]. Now a name is given to h, the strain caused by a passing gravitational wave, which is discussed next. A very rough estimate of this effect can be found by approximating the strain of a passing GW. Strain is defined as the fractional change in length caused by the passing gravitational wave:

$$h = \frac{\Delta L}{L}, \tag{2.34}$$

where ΔL is the change in length caused by the gravitational wave, and L is the original length. In the ring of test masses shown in Fig. 2.1, the strain is the change

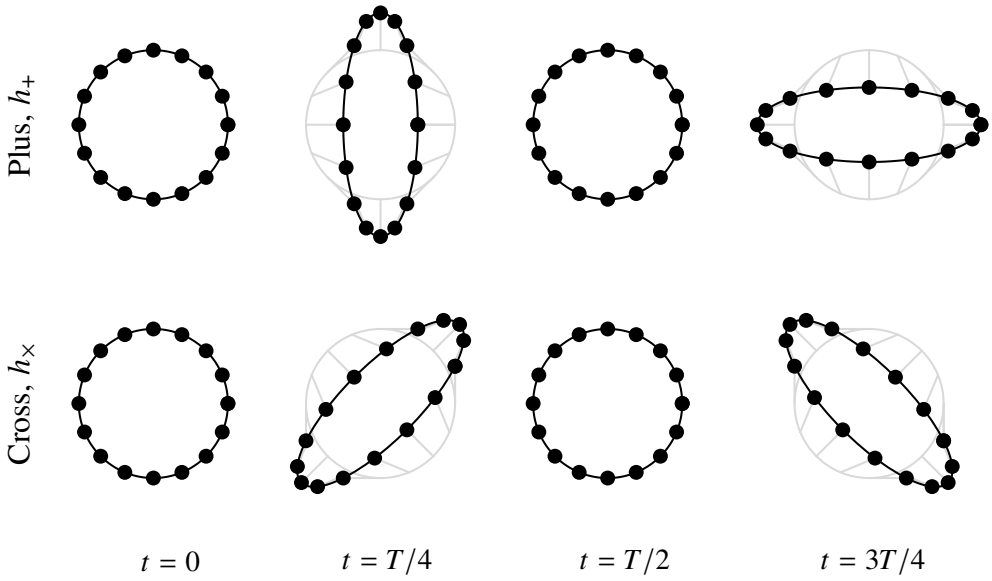


Figure 2.1: Illustration of the effect of a plus-polarized and cross-polarized gravitational wave passing through a ring of test masses. The gravitational wave causes the ring to alternately stretch and squeeze in perpendicular directions as it propagates through space. The top row shows the plus-polarization and the bottom row shows the cross-polarization. The grey lines show the displacement of the masses from their original position. Starting with the leftmost ring and moving to the right, the sequence shows the ring's deformation over one period of the gravitational wave, T.

in distance of one of the test masses from the center of the ring divided by the original radius of the ring. An order-of-magnitude calculation of the strain h caused by a passing GW can be made by considering two masses, m_1 and m_2, separated by a distance r and orbiting each other with an angular frequency ω. Using Kepler's third law, the angular frequency can be related to the masses and separation as

$$\omega^2 = \frac{G(m_1 + m_2)}{r^3}. \tag{2.35}$$

For a binary system in a circular orbit, the leading-order GW strain at a distance R from the source can be very roughly approximated as

$$h \sim \frac{4G^{5/3}}{c^4} \frac{\mathcal{M}^{5/3}\omega^{2/3}}{R} \tag{2.36}$$

where

$$\mathcal{M} = \frac{(m_1 m_2)^{3/5}}{(m_1 + m_2)^{1/5}} \tag{2.37}$$

is the chirp mass of the system [27, 28]. Substituting Eqs. 2.35 and 2.37 into Eq. 2.36 results in

$$h \sim \frac{4G^2}{c^4} \frac{m_1 m_2}{rR}. \tag{2.38}$$

For astrophysical sources such as binary neutron star (BNS) mergers, the masses involved are on the order of several times the mass of the Sun ($M_\odot \approx 2 \times 10^{30}$ kg), and the distances to Earth are typically millions to billions of light-years away. Even with these massive objects, the resulting strain h at Earth is on the order of 10^{-21} or smaller, which is an incredibly tiny effect.

A strain of 10^{-21} means that for every kilometer of distance, the change in length caused by the passing gravitational wave is only about one-thousandth the diameter of a proton (approximately 10^{-18} m). It is like seeing the distance from Earth to the nearest star (about 4.24 light-years or approximately 4×10^{16} m) change by less than the width of a human hair (approximately 10^{-5} m). Detecting such minuscule changes requires extremely sensitive instruments and sophisticated data analysis techniques.

2.3 A Quantum Treatment of Spacetime

Of the four fundamental interactions, electromagnetism and the weak and strong nuclear interactions have successful descriptions within quantum field theory. Gravity is the exception: general relativity, described in Section 2.1, is a classical theory of dynamical spacetime, and no experimentally confirmed quantum description of gravity is presently available [11, 12].

2.3.1 The Quantum Gravity Gap

To date there have been many notable attempts to quantize gravity, mainly string theory [29, 30] and loop quantum gravity[3] [31].

The difficulty is not simply that gravity has not yet been written in the same form as the other forces. Quantum field theory is normally formulated on a fixed spacetime background, while general relativity makes the geometry of spacetime itself dynamic. In most laboratory and astrophysical situations this tension is not visible because either quantum effects dominate while gravity is negligible, or gravitational effects dominate while the quantum state of spacetime can be ignored. A complete theory of quantum gravity must describe the regime where both assumptions fail at the same time.

The scale of that regime is set by the Planck length which is many orders of magnitude smaller than any distance that can be resolved directly. This creates an experimental gap as well as a theoretical one. A candidate theory may make clear statements

[3] Loop quantum gravity is very commonly abbreviated as LQG in the literature. Here, this abbreviation is avoided to prevent confusion with linear-quadratic-Gaussian control in Chapter 6.

about black holes, the early universe, or Planck-scale scattering, but those systems are not easily available as controlled laboratory experiments. The result is that many proposed theories can agree with all currently tested limits while differing in their predictions for the microscopic structure of spacetime. The Quantum Gravity gap is more of an experimental confirmation gap than a theoretical one.

For precision-measurement experiments, the useful question is not always whether a complete quantum-gravity theory can be tested directly. It is often whether broad low-energy consequences can be parameterized and constrained. Without an experimentally confirmed theory of quantum gravity, it is still possible to make progress by considering general principles and symmetries that many theories should satisfy, and then looking for phenomenological signatures that could be tested with current or near-future experiments.

2.3.2 The Standard Quantum Limit

To introduce the concept of quantum spacetime fluctuations, it is useful to first consider the standard quantum limit (SQL) in interferometric gravitational wave detectors [32]. As discussed in Chapter 3, interferometers very accurately measure changes in the distance between test masses caused by passing gravitational waves. This measurement is fundamentally a measurement of distance, which is subject to quantum mechanical limits.

Here a simplified derivation of the SQL for distance measurements is presented. It was developed by Salecker and Wigner in the late 1950s [33, 34], and expanded by Ng and Van Dam [35]. Consider a measurement of distance L with a clock at one end and a mirror at the other. If a photon is sent from the clock's position, at time t_0, to the mirror and reflected back at the clock, the time taken for the photon to complete a round trip, t_f, can be found to depend on the distance between the clock and mirror, L, as

$$t_\mathrm{f} = \frac{2L}{c} \implies L = \frac{ct_\mathrm{f}}{2}, \tag{2.39}$$

where c is the speed of light.

A logical next step for one in the field of metrology is to ask with what precision L can be measured. Quantum mechanics dictates that the clock can not be perfectly stationary and thus has an uncertainty in its momentum Δp conjugate to its uncertainty in position Δx. Here, these uncertainties are defined as the standard deviations of the variables. The product of these uncertainties is bounded by the Heisenberg

uncertainty principle as

$$\Delta x \Delta p \geq \frac{\hbar}{2}, \tag{2.40}$$

where $\hbar$ is the reduced Planck constant. This uncertainty in the agent making the measurement (the clock) at the time of the photon being emitted and the time of the photon's return means that the length becomes uncertain as well. This uncertainty in length can be expressed as

$$\delta L = \Delta x(t_0) + \Delta x(t_\mathrm{f}), \tag{2.41}$$

Since the clock has an uncertainty in its momentum Δp, it will have an uncertainty in its velocity Δv given by

$$\Delta v = \frac{\Delta p}{m}, \tag{2.42}$$

where m is the mass of the clock. This uncertainty in velocity means that over the time taken for the photon to complete its round trip, t_f, the latter half of Eq. 2.41, which adds the uncertainty in the clock's position at the time of the photon's return, can be expressed as

$$\Delta x(t_\mathrm{f}) = \Delta v t_\mathrm{f} = \frac{\Delta p}{m} t_\mathrm{f}. \tag{2.43}$$

Substituting this expression into Eq. 2.41 and then using the Heisenberg uncertainty principle to eliminate Δp results in

$$\delta L = \Delta x(t_0) + \frac{\Delta p}{m} t_\mathrm{f} = \Delta x(t_0) + \frac{\hbar t_\mathrm{f}}{2m\Delta x(t_0)}. \tag{2.44}$$

Minimizing this expression with respect to $\Delta x(t_0)$ and substituting back in the time taken for the photon to complete its round trip results in

$$\delta L_\mathrm{SQL} \geq \sqrt{\frac{\hbar L}{mc}}, \tag{2.45}$$

This scaling is the same back-action scaling that enters the SQL in laser interferometric measurements [32]. In an interferometer, the quantum limit is usually described as a balance between two noise terms. Shot noise is the measurement imprecision caused by photon-counting statistics at the readout, while quantum radiation pressure noise is the back action caused by photon-number fluctuations pushing on the test masses. The argument above isolates the back-action part of this tradeoff [33, 34]. Reducing the initial position uncertainty increases the momentum uncertainty of the measuring object, and that momentum uncertainty becomes displacement uncertainty during the light-travel time. The shot-noise term is not included explicitly in this thought experiment, but the combined imprecision and back-action tradeoff is the origin of the interferometric SQL.

2.3.3 Gravitational Effects on Length Measurements

If one wanted to measure a distance, L, to a high precision by minimizing Eq. 2.45, one would need to increase the mass of the clock m. However, increasing the mass of the clock unbounded would increasingly bend spacetime around it due to its gravitational field. This, in turn, would increasingly distort the path of the photon used to measure the distance L. While Section 2.3.2 derived a quantum limit to distance measurements, it did not consider the gravitational effects of the measuring device itself. Here a continuation of the Salecker-Wigner thought experiment is presented that considers these gravitational effects, following the treatment of Ng and Van Dam [35].

If it is assumed that the clock presented in Section 2.3.2 is spherically symmetric with a radius, r_{clock}, then to avoid the clock becoming a black hole, the radius must be larger than the Schwarzschild radius, r_{S}:

$$r_{\text{clock}} \geq r_{\text{S}} \equiv \frac{2Gm}{c^2}, \tag{2.46}$$

where G is the gravitational constant. The radial distance from the clock to the mirror must also be larger than the radius of the clock, $L \geq r_{\text{clock}}$. The proper radial distance expressed by the Schwarzschild metric is

$$\mathrm{d}l = \frac{\mathrm{d}r}{\sqrt{1 - \frac{r_{\text{S}}}{r}}}. \tag{2.47}$$

To measure the distance between the surface of the clock, $r = r_{\text{clock}}$, and the mirror, $r = r_{\text{clock}} + L$, the proper distance is integrated as

$$L_{\text{proper}} = \int_{r_{\text{clock}}}^{r_{\text{clock}}+L} \frac{\mathrm{d}r}{\sqrt{1 - \frac{r_{\text{S}}}{r}}}. \tag{2.48}$$

In a flat spacetime, $r_{\text{S}} \to 0$, and because of that $L_{\text{proper}} = L$. Thus, the static difference between the proper distance and the coordinate distance in this setup, δL_{GR}, is

$$\delta L_{\text{GR}} = L_{\text{proper}} - L = \int_{r_{\text{clock}}}^{r_{\text{clock}}+L} \frac{\mathrm{d}r}{\sqrt{1 - \frac{r_{\text{S}}}{r}}} - L. \tag{2.49}$$

For a realistic clock where $r_{\text{clock}} \gg r_{\text{S}}$, this expression may be approximated using a binomial expansion to first order as

$$\delta L_{\text{GR}} \approx \int_{r_{\text{clock}}}^{r_{\text{clock}}+L} \left(1 + \frac{r_{\text{S}}}{2r}\right) \mathrm{d}r - L = \frac{r_{\text{S}}}{2} \ln\left(\frac{r_{\text{clock}} + L}{r_{\text{clock}}}\right). \tag{2.50}$$

Given that this logarithm is on the order of one, the static difference caused by the curvature of spacetime around the clock scales as

$$\delta L_{\mathrm{GR}} \sim r_{\mathrm{S}} \sim \frac{Gm}{c^2}. \tag{2.51}$$

The next step in the argument laid out in [35] treats this static difference as an irreducible uncertainty and combines it with the SQL scale. Since the static term could in principle be calibrated or subtracted, this is a historical scaling argument, not a measurement-noise limit or a prediction for the fluctuations sought by GQuEST. Within that historical assumption, the sum is

$$\delta L_{\mathrm{heur}} \equiv \delta L_{\mathrm{SQL}} + \delta L_{\mathrm{GR}} \sim \sqrt{\frac{\hbar L}{mc}} + \frac{Gm}{c^2}. \tag{2.52}$$

This heuristic length scale is minimized when the mass of the clock is chosen as

$$m_{\mathrm{opt}} = \left(\frac{\hbar c^3 L}{4G^2}\right)^{1/3}. \tag{2.53}$$

Substituting this expression for m into Eq. 2.52 gives

$$\delta L_{\mathrm{heur}} \sim \left(l_{\mathrm{P}}^2 L\right)^{1/3}, \tag{2.54}$$

where $l_{\mathrm{P}} = \sqrt{\hbar G/c^3}$ is the Planck length. This scale depends on the Planck length and is far greater than that of other estimates such as those from loop quantum gravity [31] and string theory [29, 30]. It grows with the cube root of the distance being measured.

2.3.4 Quantum Spacetime Fluctuations

The framework presented in Sections 2.3.2 and 2.3.3 is not presented as a complete, comprehensive, or rigorous theory of quantum gravity. The geontropic fluctuations sought by GQuEST have a separately calculated interferometer response [1, 36]. Recent work has also analyzed QST fluctuation signals in interferometers for a number of quantum-gravity models [37]. A more general representation of the magnitude of root mean squared (RMS) uncertainty in length measurements caused by QST fluctuations may be expressed as

$$\delta L \propto (l_{\mathrm{P}})^{\alpha} (L)^{1-\alpha}, \tag{2.55}$$

where α is a constant between 0 and 1 that depends on the specific model of quantum gravity [38, 39]. Here α denotes the length-scaling exponent in Eq. 2.55, it should

not be confused with the geontropic signal-strength parameter α_ϕ used in Eq. 3.53 to scale the power spectral density of a specific QST signal model. The historical estimate in Section 2.3.3 has the same length scaling as older $\alpha = 2/3$ models [35, 40]. More modern theories classified as "random walk" models correspond to $\alpha = 1/2$, more recently "holographic" models have also proposed the same value [41, 42]. Models such as loop quantum gravity and string theory predict no dependence on L and thus a value of $\alpha = 1$, which corresponds to QST fluctuations on the order of the Planck length [29–31, 43–45].

While theories of quantum gravity treat α as a constant, experimentalists have often treated α as a free parameter to be constrained by experiment. And in this constraint effort it is possible to rule out certain models of quantum gravity. Experiments such as GQuEST, discussed in Chapter 3, seek to measure or constrain these QST fluctuations [36]. The lowest values of α predict the largest QST fluctuations for a given length L and are thus the easiest to measure or constrain. For this reason more recent experimental realizations [1, 46] have focused on $\alpha = 1/2$ models that include random walk [38], holographic, and pixellon models [36, 47–49]. It has also been predicted that these QST fluctuations may be measurable in the next generation of GW detectors such as Cosmic Explorer or the Einstein Telescope [50–52].

For $\alpha = 2/3$, Eq. 2.55 gives $\delta L \propto l_{\mathrm{P}}^{2/3} L^{1/3}$. This length scaling does not determine the frequency dependence or interferometer response. The geontropic spectrum in Fig. 2.2 follows from different physical assumptions [36] and therefore cannot be assigned to the $\alpha = 2/3$ scaling by an amplitude rescaling alone.

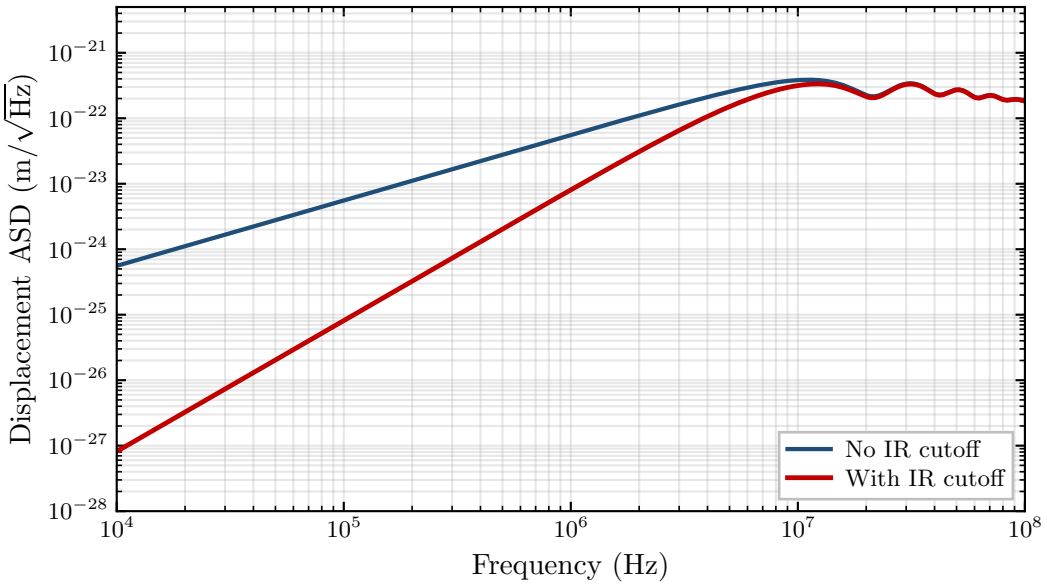


Figure 2.2: Displacement amplitude spectral density (ASD) for the geontropic (pixellon) QST model [36], calculated for GQuEST with arm length $L = 7$ m and signal-strength normalization $\alpha_\phi = 1$ as defined in Eq. 3.53. The red curve includes the infrared cutoff in the response, and the blue curve shows the response without it.

Chapter 3

BACKGROUND

"The wide world is all about you: you can fence yourselves in,
but you cannot for ever fence it out"

— J. R. R. Tolkien, *The Fellowship of the Ring*

Gravitational waves exist on an almost incomprehensibly small scale. As discussed in Section 2.2, the strains produced by astrophysical sources are on the order of 10^{-21} or smaller, meaning that to be detected a clever method of measurement is required. To provide evidence for Einstein's prediction of GWs, direct measurement of these tiny strains needed to be achieved. Indirect evidence for GWs was first observed in the Hulse-Taylor binary pulsar system in the 1970's [53–55], when the orbital decay of the system matched the predictions made by general relativity for energy loss due to GW emission. But direct detection of GWs required the development of new technologies and experimental techniques.

One of the first proposed methods to detect GWs was through the use of resonant mass detectors, also known as Weber bars after their inventor Joseph Weber at the University of Maryland in the 1960's [56]. The idea behind these detectors was to use a large, solid cylinder of metal that would resonate when a GW passed through it. By measuring the longitudinal vibrations of the cylinder using piezoelectric transducers, one could infer the presence of a GW. While Weber claimed that the bars detected gravitational waves [57, 58], his results were never replicated and today the scientific community does not believe that his signals were GWs [59]. While Weber bars were not successful in detecting GWs, they did pave the way for future detectors, mainly laser interferometers.

The idea of using interferometers to detect GWs was first proposed by Rainer Weiss at MIT in 1972 [60]. He suggested that by using a Michelson interferometer with long arms, one could measure the tiny changes in length caused by passing GWs. In addition, he also analyzed the various noise sources that would limit the sensitivity of such a detector, such as radiation-pressure noise, seismic noise, laser phase noise, and mechanical thermal noise, which are the most dominant noise sources in the

detector today [60]. The foresight and careful consideration of the noise sources in interferometric GW detectors laid the groundwork for its eventual success. This proposal led to the eventual success of LIGO and was the motivation for Weiss receiving half of the 2017 Nobel Prize in Physics, shared with Barry Barish and Kip Thorne at Caltech [9]. Along with Thorne and Ronald Drever at Caltech, Weiss helped to establish the LIGO project in the 1980's [61].

3.1 Interferometers as a Probe of Spacetime

Laser interferometers have become the most successful method for detecting GWs to date [2]. Interferometers work by splitting a laser beam into two perpendicular arms, reflecting the beams off mirrors at the ends of the arms, and then recombining the beams at the beam splitter. If the lengths of the arms change relative to each other, such as when a GW passes through the detector, the interference pattern of the recombined beams will change, allowing for the detection of the GW. This sensitivity to changes in length is what makes interferometers such powerful tools for probing spacetime. The basic principle of a Michelson interferometer is illustrated in Fig. 3.1, and takes advantage of the differential effect a GW has on the two arms of the interferometer as described in Section 2.2. Consider Fig. 2.1, where a GW passes through the plane of the interferometer in the z-direction. If the gravitational wave is + polarized, then the arm aligned with the x-axis will be stretched while the arm aligned with the y-axis will be compressed, and vice versa half a cycle later.

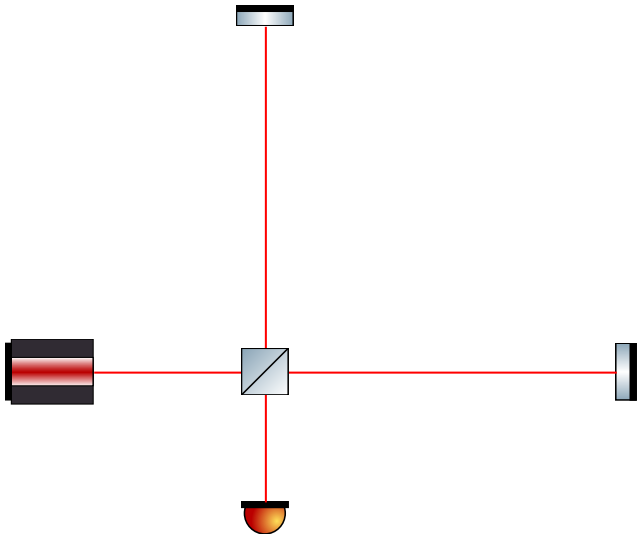


Figure 3.1: The basic layout of a Michelson interferometer. Changes in the lengths of the arms, such as those caused by passing gravitational waves, will alter the interference pattern observed at the photodetector.

In a Michelson interferometer, the differential optical phase accumulated between the two arms is given by

$$\Delta\phi = \frac{4\pi}{\lambda}(L_x - L_y) = \frac{4\pi\nu}{c}(L_x - L_y) = 2k(L_x - L_y), \tag{3.1}$$

where λ is the laser wavelength, c is the speed of light, ν is the laser frequency, $k = 2\pi/\lambda$ is the laser wavenumber, and L_x and L_y are the arm lengths. The factor of two arises from the round-trip propagation of the light in each arm.

When a GW passes through the interferometer, it induces differential changes in the arm lengths. Similar differential changes may also be produced by actuators that displace the end mirrors. The total differential phase may therefore be written as

$$\Delta\phi = \Delta\phi_0 + \delta\phi_{\mathrm{GW}} = 2k\,\Delta L_0 + 2k\,\delta L_{\mathrm{GW}}, \tag{3.2}$$

where $\Delta L_0 = L_{x,0} - L_{y,0}$ is the static differential arm-length offset defining the operating point, and δL_{GW} is the time-dependent differential arm-length change induced by the GW.

In LIGO, the interferometer is nominally operated near a dark fringe, corresponding to

$$\Delta\phi_0 = (2N + 1)\pi, \tag{3.3}$$

with integer N. For this reason, the output port of the interferometer is commonly referred to as the dark port or the antisymmetric port while the input port is called the bright port or the symmetric port.

The optical power at the dark port is recorded on a photodetector and may be expressed as

$$P = P_0 \frac{1}{4}\left(r_{\mathrm{x}}^2 + r_{\mathrm{y}}^2\right)\left(1 - C_{\mathrm{IFO}}\cos\Delta\phi\right), \tag{3.4}$$

where P_0 is the input laser power, r_{x} and r_{y} are the *field* reflectivities of the end mirrors in the x and y arms, and C_{IFO} is the interferometer contrast,

$$C_{\mathrm{IFO}} = \frac{2r_{\mathrm{x}}r_{\mathrm{y}}}{r_{\mathrm{x}}^2 + r_{\mathrm{y}}^2}. \tag{3.5}$$

This expression shows that the detected power depends on the relative optical phase between the two arms. Maximum power occurs for constructive interference, while minimum power occurs for destructive interference. In a realistic interferometer,

the minimum power is not strictly zero due to imperfections such as optical losses, contrast defects, and mismatched arm reflectivities.

Considering the idealized case $r_x = r_y = 1$ and linearizing to first order in the small phase perturbation $\delta\phi_{GW}$, the dark-port power becomes

$$P = \frac{P_0}{2}\left[1 - \cos(\Delta\phi_0 + \delta\phi_{GW})\right] \tag{3.6}$$

$$\simeq \frac{P_0}{2}\left[1 - \cos(\Delta\phi_0) + \sin(\Delta\phi_0)\ \delta\phi_{GW}\right]. \tag{3.7}$$

The power fluctuation induced by a passing GW is therefore

$$\delta P_{GW} = \frac{P_0}{2}\sin(\Delta\phi_0)\ \delta\phi_{GW} \tag{3.8}$$

$$= \frac{P_0}{2}\sin(\Delta\phi_0)\ 2khL, \tag{3.9}$$

where h is the gravitational-wave strain defined in Eq. 2.34, and L is the arm length. For a gravitational wave with plus polarization aligned to the interferometer arms, the differential arm-length change satisfies $\delta(L_x - L_y) = hL$.

This expression shows that the *linear* response of the interferometer is maximized when it is operated at mid-fringe,

$$\Delta\phi_0 = \left(N + \frac{1}{2}\right)\pi, \tag{3.10}$$

for an integer N where $\sin(\Delta\phi_0) = \pm 1$. While LIGO operates nominally near a dark fringe, auxiliary techniques such as DC and RF readout are employed to recover a linear response to phase fluctuations. While operating at mid-fringe maximizes the linear response, it also increases susceptibility to laser power fluctuations. Therefore, a trade-off exists between maximizing sensitivity to gravitational waves and minimizing noise from laser power fluctuations.

While Michelson interferometers are excellent detectors because of their sensitivity to differential arm-length changes, a standard Michelson design, depicted in Fig. 3.1, is not sufficient to achieve the sensitivity required to detect astrophysical sources of GWs or QSTs fluctuations. To enhance the sensitivity of interferometric GW detectors, several advanced techniques are employed, including the use of Fabry-Pérot cavities in the arms, power recycling, and signal recycling. These techniques effectively increase the interaction time of the laser light with the gravitational wave signal, thereby improving the detector's sensitivity.

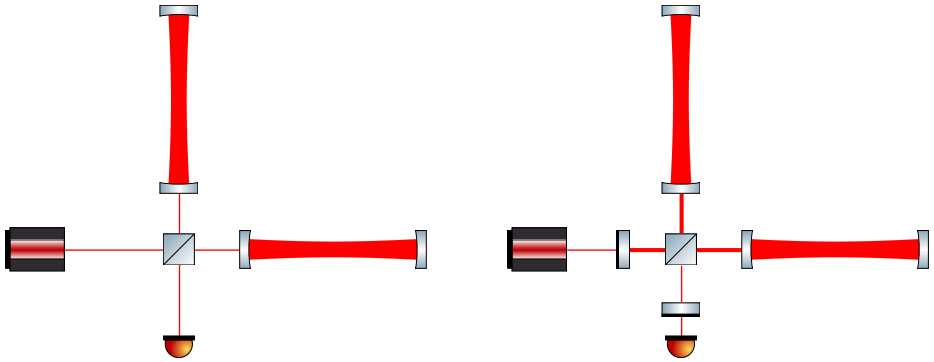
3.2 Laser Interferometer Gravitational-Wave Observatory (LIGO)

Interferometers can probe spacetime, as demonstrated by the LIGO detectors. The LIGO detectors are the most successful example of this to date. The LIGO detectors are large-scale Michelson interferometers with 4-kilometer-long arms, located in Hanford, Washington and Livingston, Louisiana. They are designed to detect gravitational waves from astrophysical sources such as merging black holes and neutron stars. The LIGO detectors have made groundbreaking discoveries since their first detection of gravitational waves in 2015 [2], and continue to be at the forefront of gravitational wave astronomy. The LIGO detectors are the most sensitive instruments ever built for detecting gravitational waves, and their success has paved the way for future detectors such as the Einstein Telescope [52] and Cosmic Explorer [51], which will further enhance our ability to probe spacetime and understand the universe. Each LIGO detector uses many advanced interferometric techniques to achieve the sensitivity required to detect astrophysical sources of GWs. As discussed in Section 2.2, the signals from GWs are extremely weak. These weak signals mean that the basic Michelson design from Fig. 3.1 is not sufficient to achieve the required sensitivity, so LIGO incorporates several enhancements to the basic design, such as Fabry-Pérot cavities in the arms, power recycling, and signal recycling as carefully budgeting and reducing noise from various sources. The resulting interferometer design is far more complicated than the basic Michelson design, but is necessary to achieve the sensitivity required to detect astrophysical sources of GWs.

3.2.1 Fabry-Pérot Cavities

One of the most common enhancements to the basic Michelson interferometer design is the incorporation of Fabry-Pérot cavities in each arm. While the basic Michelson interferometer is sensitive to differential arm-length changes, its sensitivity can be significantly improved by optimizing its parameters. As described in Eq. 3.9, the power fluctuation induced by a passing GW is proportional to the arm length L. Therefore, increasing the effective arm length of the interferometer enhances its sensitivity to small fluctuations in phase. Note that while this is true for low-frequency gravitational-wave signals, high-frequency signals do not always benefit from longer arms due to the finite travel time of light in the arms. As the light travel time becomes comparable to the period of the gravitational wave, the response of the interferometer can be reduced. For this reason, GQuEST, described in Section 3.3, does not employ a Fabry-Pérot Michelson interferometer (FPMI). On the other

hand, LIGO does see the benefit, so Fabry-Pérot cavities are used to increase the effective arm length without physically extending the arms.



(a) Fabry-Pérot Michelson

(b) Signal-Recycled Power-Recycled Fabry-Pérot Michelson

Figure 3.2: Two common Michelson interferometer configurations that build on the basic Michelson design shown in Fig. 3.1. (a) A Fabry-Pérot Michelson interferometer (FPMI), which uses Fabry-Pérot cavities in each arm to increase the effective path length of the light, thereby enhancing sensitivity to gravitational waves. (b) A Signal-Recycled Power-Recycled Fabry-Pérot Michelson interferometer (SR-PRFPMI), which incorporates both power recycling and signal recycling techniques to further improve sensitivity and bandwidth.

Depicted in Fig. 3.2a, a FPMI adds two mirrors to the arms of the interferometer making each arm a Fabry-Pérot cavity. A Fabry-Pérot cavity consists of two highly reflective mirrors facing each other, forming a resonant optical cavity. When laser light enters the cavity, it undergoes multiple reflections between the mirrors, effectively increasing the optical path length. This increased path length enhances the sensitivity of the interferometer to changes in arm length caused by passing gravitational waves.

In addition to being used in the arms of the interferometer, Fabry-Pérot cavities as well as other types of optical cavities are also employed as important optical tools for laser frequency stabilization, mode cleaning, and filtering of squeezed states of light [62]. For this reason it is important to understand the basic properties of optical cavities.

Consider two mirrors, M_1 and M_2, with respective amplitude reflectivities r_1 and r_2, amplitude transmissivities t_1 and t_2, and the distance between them L, forming a Fabry-Pérot cavity as shown in Fig. 3.3. The two reflective coatings on the mirrors face each other, creating a resonant optical cavity.

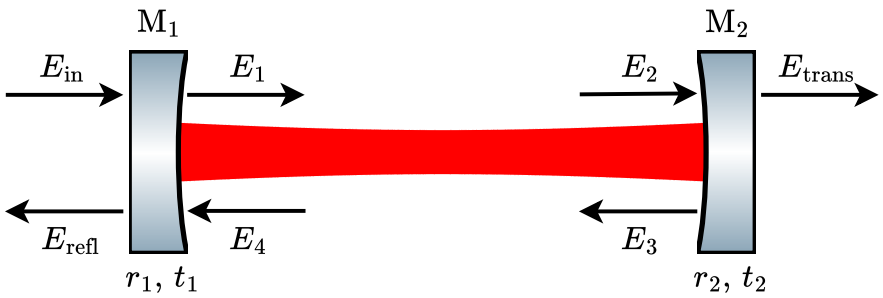


Figure 3.3: A Fabry-Pérot cavity formed by two mirrors, M_1 and M_2, separated by a distance L. The cavity enhances the effective optical path length through multiple reflections of the light between the mirrors.

When a laser beam with electric field amplitude E_{in} enters the cavity through mirror M_1, it undergoes multiple reflections between the two mirrors. The electric fields inside the cavity, E_1, E_2, E_3, and E_4, as well as the transmitted field, E_{trans}, and reflected field, E_{refl}, can be expressed in terms of the input field and the cavity parameters as follows:

$$E_1 = it_1 E_{\text{in}} + r_1 E_4, \tag{3.11}$$

$$E_2 = E_1 e^{-ikL}, \tag{3.12}$$

$$E_3 = r_2 E_2, \tag{3.13}$$

$$E_4 = E_3 e^{-ikL}, \tag{3.14}$$

$$E_{\text{trans}} = it_2 E_2, \tag{3.15}$$

$$E_{\text{refl}} = r_1 E_{\text{in}} + it_1 E_4, \tag{3.16}$$

where $k = 2\pi/\lambda$ is the wavenumber of the laser light; the mirrors are assumed lossless, so that $|r_i|^2 + |t_i|^2 = 1$. From these equations, one can solve for the cavity's transmitted and reflected power. The transmitted power is given by

$$P_{\text{trans}} = |E_{\text{trans}}|^2 = P_{\text{in}} \frac{T_1 T_2}{(1 - r_1 r_2)^2 + 4 r_1 r_2 \sin^2(kL)}, \tag{3.17}$$

where $P_{\text{in}} = |E_{\text{in}}|^2$ is the input power, and $T_1 = |t_1|^2$ and $T_2 = |t_2|^2$ are the power transmissivities of mirrors M_1 and M_2, respectively. The reflected power is given by

$$P_{\text{refl}} = |E_{\text{refl}}|^2 = P_{\text{in}} \frac{(r_1 - r_2)^2 + 4 r_1 r_2 \sin^2(kL)}{(1 - r_1 r_2)^2 + 4 r_1 r_2 \sin^2(kL)}. \tag{3.18}$$

Both of these equations can be expressed as frequency-dependent functions by substituting $k = 2\pi\nu/c$, where ν is the frequency of the laser light and c is the

speed of light. In this way it is possible to write the transmitted and reflected power as frequency dependent factors that multiply the input power, which are commonly referred to as the cavity's transmission and reflection coefficients, respectively. The cavity is said to be on resonance when the round-trip phase shift satisfies $2kL = 2m\pi$ for an integer m, which corresponds to constructive interference of the circulating light. On resonance, the transmitted power is maximized and the reflected power is minimized.

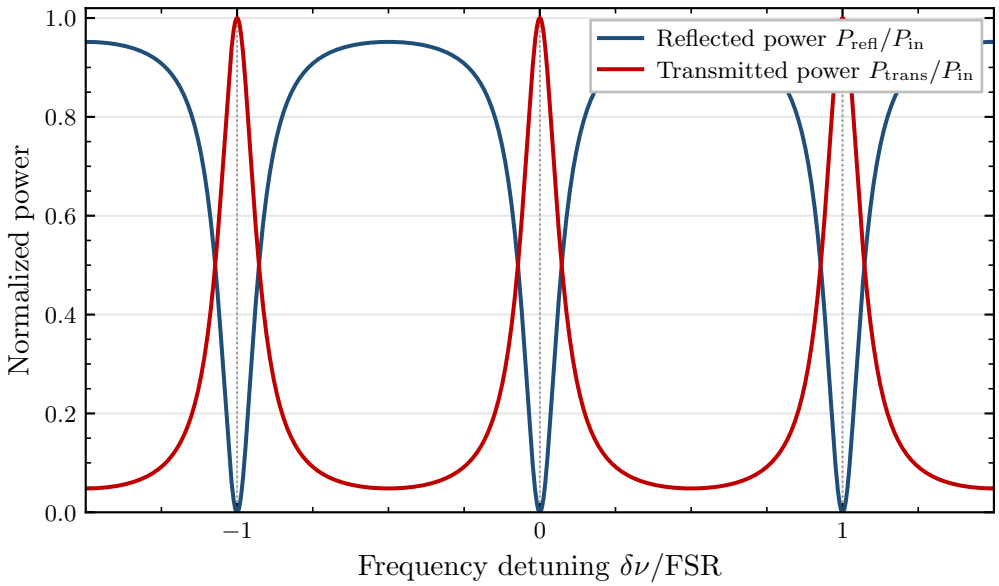


Figure 3.4: Normalized reflected and transmitted powers, $P_{\text{refl}}/P_{\text{in}}$ and $P_{\text{trans}}/P_{\text{in}}$, for a Fabry-Pérot cavity plotted as a function of frequency detuning in units of the cavity free spectral range. Resonances occur at integer multiples of the free spectral range (FSR), where the transmitted power is maximized and the reflected power is minimized.

Fabry-Pérot cavities are fundamental building blocks of most systems in interferometric GW detectors, and are used in various configurations to enhance sensitivity. A cavity has a series of resonances as a function of frequency, as shown in Fig. 3.4, where the transmitted power is maximized and the reflected power is minimized at the resonant frequencies. The properties of these resonances are determined by the cavity parameters, such as the mirror reflectivities and the cavity length. The frequency spacing between adjacent resonances is determined by the cavity length and is known as the cavity's free spectral range (FSR), which is expressed as

$$\text{FSR} = \frac{c}{2L}. \tag{3.19}$$

The finesse, which quantifies the sharpness of the resonance, is given by

$$\mathcal{F} = \frac{\pi\sqrt{r_1 r_2}}{1 - r_1 r_2}. \tag{3.20}$$

Finesse is a measure of how many times the light bounces back and forth between the mirrors before it is transmitted or reflected, and higher finesse corresponds to a narrower resonance. This is closely related to the cavity's quality factor, Q, which is a measure of how underdamped the cavity is. The quality factor is related to the finesse and the FSR by

$$Q = \mathcal{F}\frac{\nu}{\mathrm{FSR}}, \tag{3.21}$$

where ν is the frequency of the laser light. The quality factor quantifies how long the light can circulate in the cavity before it is lost, and higher Q corresponds to a longer storage time for the light in the cavity. The storage time for a cavity is given by

$$\tau_{\mathrm{photon}} = \frac{Q}{2\pi\nu} = \frac{\mathcal{F}}{2\pi\mathrm{FSR}}. \tag{3.22}$$

From the storage time, one can also define the cavity's pole frequency, which is the frequency at which the cavity's response to a signal begins to roll off, as

$$f_{\mathrm{pole}} = \frac{1}{2\pi\tau_{\mathrm{photon}}} = \frac{\mathrm{FSR}}{2\mathcal{F}}. \tag{3.23}$$

The pole frequency is an important parameter for interferometric GW detectors, as it determines the bandwidth of the detector and its sensitivity to phase modulations at different frequencies. The cavity acts as a first order low-pass filter for signals, with the pole frequency setting the cutoff frequency above which the response begins to roll off. For frequencies much lower than the pole frequency, the cavity provides a significant enhancement to the signal, while for frequencies much higher than the pole frequency, the cavity's response is significantly reduced. The f_{pole} is the 3dB point of the cavity's response, meaning that at this frequency the response is reduced to half of its maximum value. For example the LIGO arm cavities have a pole frequency of approximately around 42 Hz, using the values of FSR $\approx$ 37.5 kHz and $\mathcal{F} \approx 450$. Note that this is different from the coupled cavity pole frequency, f_{cc}, which is the pole frequency of the entire interferometer and is determined by the arm cavity parameters as well as the power recycling and signal recycling parameters. In LIGO the coupled cavity pole frequency is around 95 Hz, which is higher than

the arm cavity pole frequency due to the presence of the power recycling and signal recycling cavities.

The cavity's linewidth, which is the full width at half maximum (FWHM) of the resonance and is determined by the mirror reflectivities, is given by

$$\Delta\nu = \frac{\mathrm{FSR}}{\mathcal{F}}. \tag{3.24}$$

The optical gain of the cavity, which quantifies the enhancement of the effective optical path length, is given by

$$G = \frac{2\mathcal{F}}{\pi}. \tag{3.25}$$

The optical gain is the factor by which the effective arm length is increased due to the presence of the cavity, and is directly proportional to the finesse. For example, a cavity with a finesse of 450 would have an optical gain of approximately 286, meaning that the effective arm length is increased by a factor of 286 compared to a simple Michelson interferometer without cavities. This enhancement in effective arm length is what allows interferometric GW detectors like LIGO to achieve the sensitivity required to detect astrophysical sources of GWs.

The stability of the cavity is determined by the geometry of the mirrors and the cavity length. For each mirror one can define a g-parameter as

$$g_i = 1 - \frac{L}{R_i}, \tag{3.26}$$

where R_i is the radius of curvature of the i-th mirror. For a Fabry-Pérot cavity to be stable, the g-parameters must satisfy the following condition:

$$0 < g_1 g_2 < 1. \tag{3.27}$$

The product of the g-parameters, often called the g-factor, must be positive and less than one for the cavity to be stable. If the product is greater than or equal to one, the cavity is unstable and will not support stable resonant modes. If the product is less than or equal to zero, the cavity is also unstable and will not support stable resonant modes. The stability condition ensures that the light can be confined within the cavity and that the resonant modes are well-defined.

For a general folded or traveling-wave cavity, the accumulated round-trip Gouy phase can be calculated from the round-trip ABCD matrix [63, 64]. If

$$M_{\mathrm{rt}} = \begin{pmatrix} A & B \\ C & D \end{pmatrix}, \tag{3.28}$$

then the accumulated round-trip Gouy phase is

$$\zeta = \mathrm{sgn}(B) \cdot \cos^{-1}\left(\frac{A+D}{2}\right), \tag{3.29}$$

where the inverse-cosine branch is chosen such that $0 \leq \cos^{-1} x < 2\pi$ for $-1 < x \leq 1$.

The modes of a Fabry-Pérot cavity are determined by the cavity geometry and the boundary conditions imposed by the mirrors. The fundamental mode of the cavity is the Gaussian beam mode, which has a specific spatial profile and is characterized by its waist size and location. Shown in Fig. 3.5 are the fundamental Gaussian mode and the first few higher-order modes of a Fabry-Pérot cavity. Higher-order modes can also exist in the cavity, which have more complex spatial profiles and are characterized by their mode indices. The mode structure of the cavity can affect the sensitivity of the interferometer, as different modes may have different coupling to noise sources like mirror imperfections and thermal noise [65].

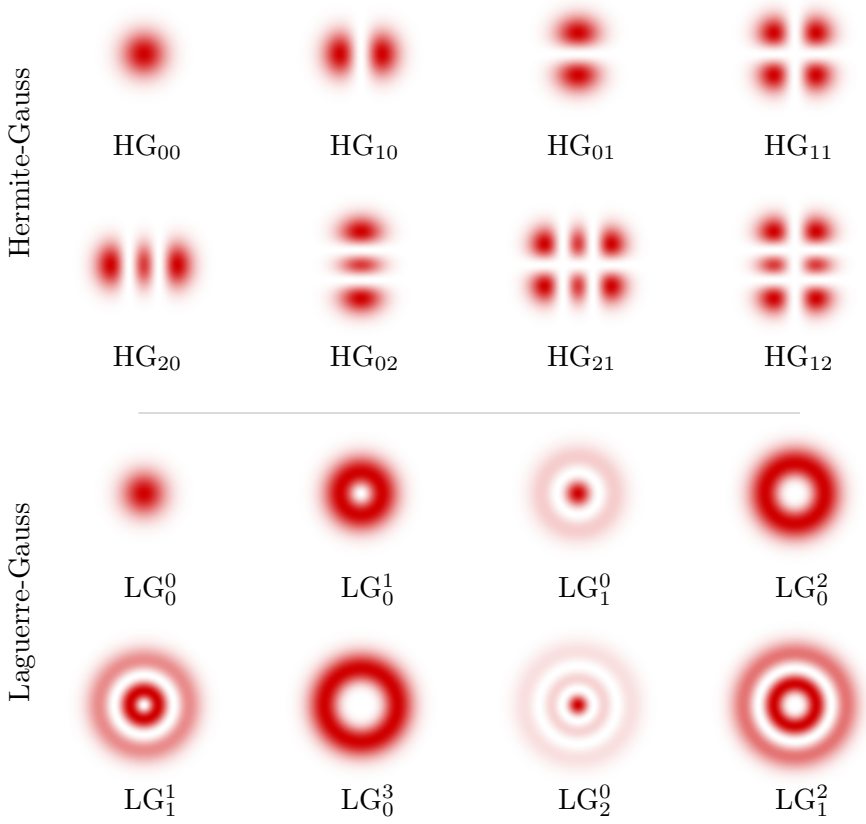


Figure 3.5: Cross sections of the fundamental Gaussian mode and the first few higher-order modes of a Fabry-Pérot cavity.

3.2.2 Traveling-Wave Cavities

In addition to Fabry-Pérot cavities, which are standing-wave cavities, traveling-wave cavities are also commonly used in interferometric GW detectors. Traveling-wave cavities, such as ring cavities and bowtie cavities, allow the light to circulate in a single direction, i.e., there is never light going in both directions at the same time. A bowtie cavity is a specific type of traveling-wave cavity that has a distinctive shape resembling a bow tie and is shown in Fig. 3.6. This can be advantageous for certain applications, such as mode cleaning, frequency stabilization, and nonlinear optics, where it is desirable to have a unidirectional flow of light. One of the benefits is reducing back reflections, which can cause interference and noise in the system, and allowing for more efficient coupling of light into and out of the cavity. Another benefit is traveling-waves don't have standing-wave patterns, i.e., they don't have nodes and anti-nodes, which can be advantageous when using nonlinear optical materials in the cavity, as it can help to reduce the effects of spatial hole burning and improve the efficiency of nonlinear processes. Traveling-wave cavities can also be used to create more complex mode structures, which can be beneficial for certain applications such as mode cleaning and frequency stabilization. Traveling-wave cavities can be more complex to design and align compared to standing-wave cavities, but they offer unique advantages that make them valuable tools in interferometric GW detectors, including LIGO.

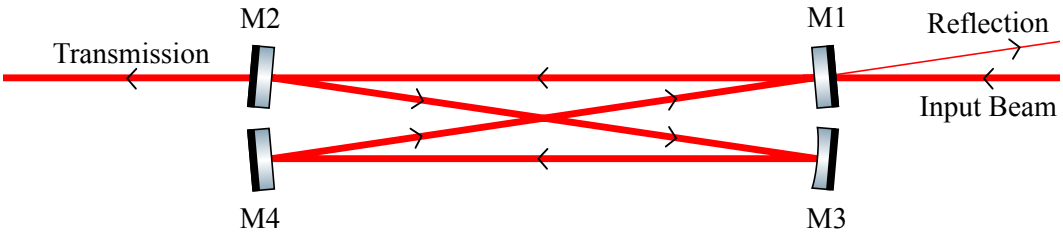


Figure 3.6: Single bow-tie traveling-wave cavity. The input beam enters through M1, while the promptly reflected field leaves through a separate reflection port. A resonant field circulates in one direction around the four-mirror path M1–M2–M3–M4, and the transmitted field exits through M2. This geometry separates the input, reflection, and transmission beams spatially, making it useful when different optical ports must be monitored or controlled independently. Because the light travels around the cavity rather than forming a standing wave between two mirrors, bow-tie cavities are also useful for mode cleaning, frequency filtering, and nonlinear-optics applications where back-reflections must be reduced.

3.2.3 Pound-Drever-Hall Technique

The optical cavities described in Sections 3.2.1 and 3.2.2 are fundamental building blocks of high-sensitivity interferometric detectors. Their resonant properties are crucial for enhancing the sensitivity of the detector and filtering out unwanted noise. However, to take advantage of these properties, it is necessary to keep the cavities on resonance with the laser light. The Pound-Drever-Hall (PDH) technique is a widely used method for locking the laser frequency to a cavity resonance, locking the cavity length to the laser frequency, or locking either degree of freedom to a stable frequency reference such as an atomic transition [66, 67]. Achieving and maintaining resonance is critical for the performance of any cavity, i.e., maximizing the optical buildup of the input laser field inside the cavity, the corresponding transmitted power, and the phase response used for sensing.

The PDH technique works by creating an error signal that is proportional to the detuning of the cavity from resonance. In the most common use case, the goal is to maximize transmission through the cavity. This is a subtle sensing problem because the transmitted and reflected powers of a cavity are symmetric about resonance. If the cavity is detuned above or below resonance by the same amount, the reflected and transmitted powers are the same, so a direct power measurement alone does not indicate which direction the laser frequency or cavity length should be changed. The PDH technique solves this problem by producing a signal that is antisymmetric about resonance. The sign of this signal gives the sign of the detuning, while its magnitude gives a measure of how far the cavity is from resonance.

To create the PDH error signal, the input light is phase modulated before it enters the cavity, and the light reflected from the cavity is monitored on a photodetector, as shown in Fig. 3.7. The phase modulation produces a carrier field at the laser frequency and two radio-frequency sidebands. The modulation frequency is chosen so that the sideband frequencies, $\omega \pm \omega_m$, are many cavity linewidths away from the carrier resonance and are not resonant with any other longitudinal mode of the cavity. As a result, the sidebands do not build up in the cavity or get filtered by the same narrow resonance as the carrier. Instead, they are promptly reflected with an approximately fixed phase and act as a stable phase reference.

For concreteness, consider the two-mirror Fabry-Pérot cavity from Section 3.2.1. The same derivation applies to other cavity geometries by replacing the Fabry-Pérot reflection coefficient with the appropriate complex reflection coefficient. Using the same mirror convention as before, the field reflection coefficient for a lossless

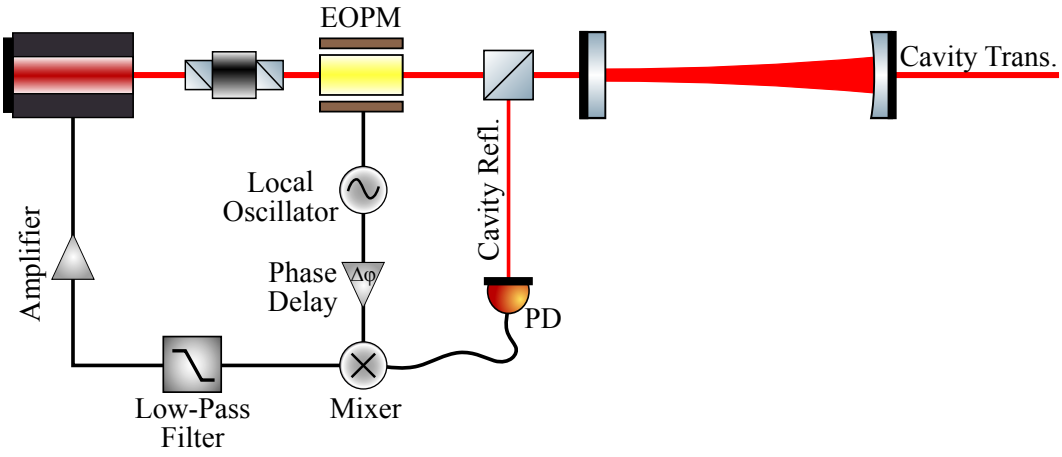


Figure 3.7: Schematic of the PDH locking technique. The input laser field is phase modulated to create sidebands around the carrier. The reflected light from the cavity is detected on a photodetector and demodulated at the modulation frequency to produce an error signal proportional to the detuning from resonance.

Fabry-Pérot cavity is

$$R(\omega) \equiv \frac{E_{\text{refl}}}{E_{\text{in}}} = \frac{r_1 - \left(r_1^2 + t_1^2\right) r_2 e^{-2i\omega L/c}}{1 - r_1 r_2 e^{-2i\omega L/c}} \tag{3.30}$$

$$= \frac{r_1 - r_2 e^{-2i\omega L/c}}{1 - r_1 r_2 e^{-2i\omega L/c}}, \tag{3.31}$$

where the final equality uses $r_1^2 + t_1^2 = 1$. The reflected power from an unmodulated beam would be $P_{\text{refl}} = P_0 |R(\omega)|^2$, which is symmetric about resonance and therefore cannot by itself provide a signed error signal.

The input field after phase modulation can be written as

$$E_{\text{in}}(t) = E_0 e^{i\omega t} e^{i\beta \sin(\omega_m t)} \tag{3.32}$$

$$= E_0 e^{i\omega t} \sum_{n=-\infty}^{\infty} J_n(\beta) e^{in\omega_m t}, \tag{3.33}$$

where ω is the laser angular frequency, ω_m is the modulation angular frequency, β is the modulation index, and J_n is the n-th Bessel function. For small modulation depth, $\beta \ll 1$, the carrier and first-order sidebands dominate, so

$$E_{\text{in}}(t) \simeq E_0 e^{i\omega t} \left[1 + \frac{\beta}{2} e^{i\omega_m t} - \frac{\beta}{2} e^{-i\omega_m t}\right] \tag{3.34}$$

$$= E_0 \left[e^{i\omega t} + \frac{\beta}{2} e^{i(\omega+\omega_m)t} - \frac{\beta}{2} e^{i(\omega-\omega_m)t}\right]. \tag{3.35}$$

Using a cosine rather than a sine modulation only changes the phase convention of the sidebands, which is absorbed into the electronic demodulation phase below. For larger modulation depths the higher-order Bessel terms must be retained, producing additional sidebands at $\omega \pm n\omega_m$.

Each frequency component reflects from the cavity with the corresponding value of the complex reflection coefficient. Defining

$$R_0 \equiv R(\omega), \qquad R_+ \equiv R(\omega + \omega_m), \qquad R_- \equiv R(\omega - \omega_m), \tag{3.36}$$

the reflected field is

$$E_{\text{refl}}(t) \simeq E_0 \left[R_0 e^{i\omega t} + \frac{\beta}{2} R_+ e^{i(\omega+\omega_m)t} - \frac{\beta}{2} R_- e^{i(\omega-\omega_m)t} \right]. \tag{3.37}$$

The photodetector measures power, so the relevant quantity is $P_{\text{refl}}(t) = |E_{\text{refl}}(t)|^2$. Keeping the carrier power, the sideband powers, and the carrier-sideband beat terms gives

$$\begin{aligned} P_{\text{refl}}(t) = P_0 &\left[|R_0|^2 + \frac{\beta^2}{4} \left(|R_+|^2 + |R_-|^2 \right) \right] \\ &+ P_0 \beta \left[\text{Re}[\chi(\omega)] \cos(\omega_m t) + \text{Im}[\chi(\omega)] \sin(\omega_m t) \right] + \text{terms at } 2\omega_m, \end{aligned} \tag{3.38}$$

where the complex quantity containing the useful detuning information is

$$\chi(\omega) = R_0 R_+^* - R_0^* R_-. \tag{3.39}$$

The first line of Eq. 3.38 contains the ordinary reflected carrier and sideband powers. These terms are symmetric about the cavity resonance and do not provide a signed measure of detuning. The second line contains beat notes between the carrier and the sidebands at the modulation frequency. These terms contain the antisymmetric information needed for locking.

The photodetector voltage is proportional to reflected power, $V_{\text{PD}}(t) = G_{\text{PD}} P_{\text{refl}}(t)$, where G_{PD} is the photodetector gain. The signal is then mixed with a phase-shifted oscillator at the modulation frequency,

$$V_{\text{mix}}(t) = V_{\text{PD}}(t) \cos(\omega_m t + \phi), \tag{3.40}$$

and low-pass filtered to remove terms oscillating at ω_m, $2\omega_m$, and higher frequencies. Absorbing the photodetector gain, mixer gain, and local oscillator amplitude into an overall gain G_{PDH}, the low-pass filtered signal is

$$V_{\text{err}}(\omega) = G_{\text{PDH}} \frac{P_0 \beta}{2} \left[\text{Re}[\chi(\omega)] \cos\phi - \text{Im}[\chi(\omega)] \sin\phi \right]. \tag{3.41}$$

The demodulation phase ϕ is chosen so that the antisymmetric quadrature of $\chi(\omega)$ appears in the output. With this phase choice, one can define the real PDH response function

$$\chi_{\mathrm{PDH}}(\omega) \equiv \mathrm{Re}[\chi(\omega)] \cos\phi - \mathrm{Im}[\chi(\omega)] \sin\phi, \tag{3.42}$$

so that the error signal takes the simple form

$$V_{\mathrm{err}}(\omega) = G_{\mathrm{PDH}} \frac{P_0 \beta}{2} \chi_{\mathrm{PDH}}(\omega). \tag{3.43}$$

Equivalently, after the demodulation phase has been chosen, the relevant part of the mixed signal before the low-pass filter can be written as

$$V_{\mathrm{mix}}(t) = C_{\mathrm{sym}}(\omega) \cos(\omega_m t) + G_{\mathrm{PDH}} P_0 \beta \, \chi_{\mathrm{PDH}}(\omega) \cos^2(\omega_m t) + \cdots \tag{3.44}$$

$$= C_{\mathrm{sym}}(\omega) \cos(\omega_m t) + G_{\mathrm{PDH}} \frac{P_0 \beta}{2} \chi_{\mathrm{PDH}}(\omega) \, [1 + \cos(2\omega_m t)] + \cdots, \tag{3.45}$$

where $C_{\mathrm{sym}}(\omega)$ contains the symmetric carrier and sideband power terms, and the ellipsis denotes oscillatory terms rejected by the low-pass filter. The remaining DC term is precisely Eq. 3.43.

Near a cavity resonance at ω_{res}, the detuning is

$$\Delta \equiv \omega - \omega_{\mathrm{res}}. \tag{3.46}$$

The PDH response is antisymmetric about resonance and crosses zero at $\Delta = 0$. For small detunings it can therefore be linearized as

$$V_{\mathrm{err}}(\Delta) \simeq G_{\mathrm{PDH}} \frac{P_0 \beta}{2} \left. \frac{d\chi_{\mathrm{PDH}}}{d\omega} \right|_{\omega=\omega_{\mathrm{res}}} \Delta. \tag{3.47}$$

This linear region around the zero crossing is the operating point of the lock. The slope determines the sensing gain, while the sign of the slope determines the feedback's direction. The overall sign of the error signal is convention-dependent and can be changed by shifting the demodulation phase or reversing the sign of the feedback actuator.

In practice, the low-pass filtered PDH error signal is fed back either to the laser frequency actuator or to an actuator that changes the cavity length. The feedback loop drives the error signal to zero, thereby holding the carrier field on resonance with the cavity while the off-resonant sidebands provide the phase reference needed to determine the sign of the detuning.

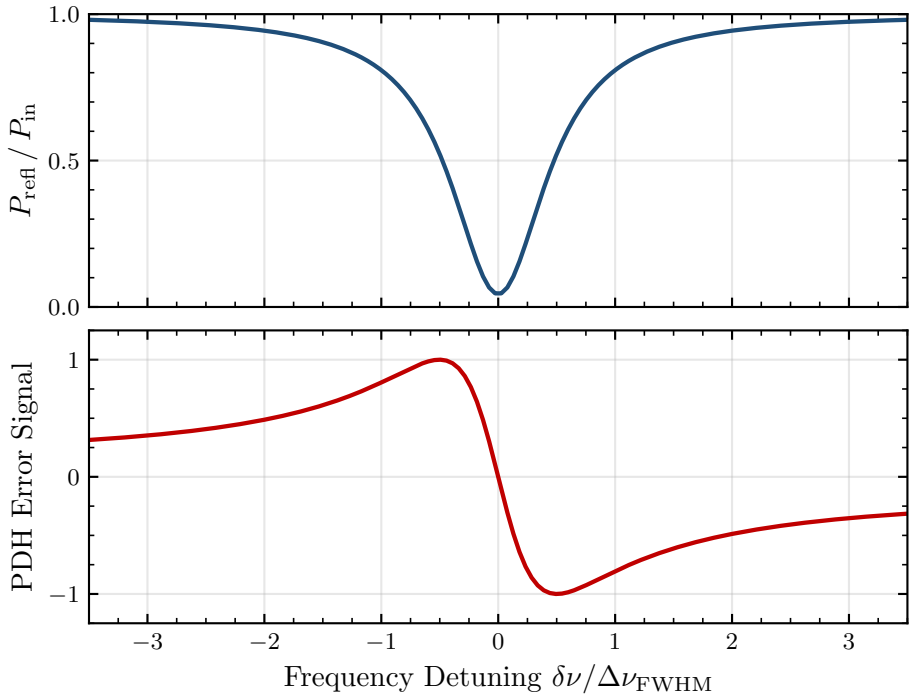


Figure 3.8: Representative PDH error signal as a function of cavity detuning. Unlike the reflected or transmitted cavity power, the PDH signal is antisymmetric about resonance and crosses zero when the cavity is resonant with the carrier field. The sign of the signal indicates which direction the laser frequency or cavity length must be actuated to return to resonance.

3.2.4 Power Recycling

In addition to the use of Fabry-Pérot cavities in the arms, the most notable alterations LIGO makes to the basic Michelson design are the incorporation of power recycling and signal recycling techniques. One can understand the purpose of power recycling rather intuitively by first remembering that the more light injected into the interferometer, the more light is available for detection. Since the interferometer is operated near a dark fringe almost all of the light is reflected back towards the laser. The interferometer is acting like a mirror, and reflecting the light back towards the laser. Once the basics of this are understood, the solution is the same as it is for the arms of the interferometer: add a mirror to form a cavity.[1] By adding a mirror between the laser and the beam splitter, as shown in Fig. 3.2b, one can form a power recycling cavity that allows the light reflected back from the interferometer to constructively interfere and build up in power, thereby increasing the effective

[1]Most of the problems in LIGO seem to have the same solution: a cavity. Not enough light in the arms? Add a cavity. Too much light being lost to the dark port? Add a cavity. Need to rotate the squeezing quadrature depending on frequency? Add a cavity. Need to filter the light? Add a cavity. Need higher efficiency from your nonlinear crystal? Add a cavity. It's a dentist's worst nightmare.

input power and improving the sensitivity of the detector.

The sensitivity improvement from power recycling can be quantified by the power recycling gain, which is given by

$$G_{\mathrm{PR}} = \left(\frac{1}{1 - r_{\mathrm{PR}} r_{\mathrm{IFO}}} \right)^2, \tag{3.48}$$

where r_{PR} is the amplitude reflectivity of the power recycling mirror and r_{IFO} is the effective amplitude reflectivity of the interferometer as seen from the power recycling cavity. The effective reflectivity of the interferometer is a result of the interference of the light returning from the two arms of the interferometer and can be written as

$$r_{\mathrm{IFO}} = r_{\mathrm{BS}}^2 r_x + t_{\mathrm{BS}}^2 r_y, \tag{3.49}$$

where r_{BS} and t_{BS} are the amplitude reflectivity and transmissivity of the beam splitter, and r_x and r_y are the effective amplitude reflectivities of the arm cavities in the two interferometer arms. When the interferometer is operated at the dark fringe and the two arm cavities are identical, $r_x = r_y = r_{\mathrm{arm}}$, and the effective reflectivity simplifies to

$$r_{\mathrm{IFO}} = r_{\mathrm{arm}}. \tag{3.50}$$

The power recycling gain can be quite large, with typical values on the order of 10 to 100, meaning that the effective input power can be increased by similar factors. This enhancement in effective input power is crucial for achieving the sensitivity required to detect astrophysical sources of gravitational waves.

3.2.5 Signal Recycling

Signal recycling is another technique used in LIGO to enhance the detectors' sensitivity, particularly at specific frequencies. It is an often misunderstood enhancement made to interferometers but it can be quite powerful. Signal recycling is similar to power recycling in that it is simply adding a mirror on one side of the beam splitter, as shown in Fig. 3.2b. The main difference is that the dark port does not receive the carrier light, since the carrier light destructively interferes at the dark port, while gravitational-wave signal sidebands do not. Therefore, when a GW passes through the interferometer, there is light at the dark port and now one can form a cavity that allows these sidebands to be partially stored in the signal recycling cavity, thereby increasing the signal strength and improving the sensitivity of the detector to GWs

at those frequencies. The signal recycling cavity can be tuned to enhance sensitivity at specific frequencies by adjusting its length, which allows for targeted searches for specific astrophysical sources of GWs. The response of the signal recycling cavity can be quantified by the frequency dependent signal recycling gain, which is given by

$$G_{\mathrm{SR}}(\omega_{\mathrm{sb}}) = \left| \frac{1}{1 - r_{\mathrm{SR}} r_{\mathrm{IFO}}(\omega_{\mathrm{sb}}) e^{2i\omega_{\mathrm{sb}} L_{\mathrm{SR}}/c}} \right|^2, \tag{3.51}$$

where r_{SR} is the amplitude reflectivity of the signal recycling mirror, L_{SR} is the length of the signal recycling cavity, and $r_{\mathrm{IFO}}(\omega_{\mathrm{sb}})$ is the effective amplitude reflectivity of the interferometer as seen from the signal recycling cavity for signal sidebands at angular frequency ω_{sb}. The effective reflectivity of the interferometer for the signal sidebands is a result of the interference of the light returning from the two arms of the interferometer and can be written as

$$r_{\mathrm{IFO}}(\omega_{\mathrm{sb}}) = r_{\mathrm{BS}}^2 r_x(\omega_{\mathrm{sb}}) e^{i\phi_x(\omega_{\mathrm{sb}})} + t_{\mathrm{BS}}^2 r_y(\omega_{\mathrm{sb}}) e^{i\phi_y(\omega_{\mathrm{sb}})}, \tag{3.52}$$

where $r_x(\omega_{\mathrm{sb}})$ and $r_y(\omega_{\mathrm{sb}})$ are the effective amplitude reflectivities of the arm cavities for the signal sidebands, and $\phi_x(\omega_{\mathrm{sb}})$ and $\phi_y(\omega_{\mathrm{sb}})$ are the phase shifts acquired by the signal sidebands in the two arms of the interferometer. By tuning the length of the signal recycling cavity, one can control the frequency response of the detector and enhance sensitivity at specific frequencies. This is particularly useful for targeting specific astrophysical sources of gravitational waves, such as binary neutron star mergers. The combination of power recycling and signal recycling allows interferometric GW detectors like LIGO to achieve the sensitivity required to detect a wide range of astrophysical sources of gravitational waves across a broad frequency band.

3.3 GQuEST: Gravity from the Quantum Entanglement of Space-Time

The LIGO detectors are designed to measure astrophysical GWs in the audio band, from a few tens of hertz to a few kilohertz. In that band, the detector output is usually treated as a calibrated time series. The phase evolution of that time series carries the waveform information.

The signals discussed in Section 2.3.4 are different. Geontropic QST fluctuations are expected to appear as a broadband stochastic signal, as shown in Fig. 2.2. The characteristic frequency is set by the light-crossing time of the experiment, so a meter-scale interferometer is naturally more sensitive in the MHz band. Since the

signal is stochastic, the measurement is not a phase-coherent waveform search like the searches performed by LIGO. A homodyne readout can still measure this type of signal [46], but a readout that measures signal power directly is better matched to the problem. The GQuEST experiment is being built around this power-measurement technique [1].

For an interferometer with arm length L, the peak geontropic displacement power spectral density (PSD) is

$$\overline{S}_L^{\phi} \equiv S_L^{\phi}(f_{\mathrm{pk}}) = \alpha_{\phi} \frac{\ell_{\mathrm{P}} L^2}{8\pi^2 c}, \tag{3.53}$$

where ℓ_{P} is the Planck length convention used in the geontropic model, and α_{ϕ} is the dimensionless geontropic signal-strength parameter. This is distinct from the length-scaling exponent α in Eq. 2.55; here $\alpha_{\phi} \sim 1$ means an order-unity geontropic signal amplitude for the $\alpha = 1/2$ model. The peak frequency of the signal is defined by

$$f_{\mathrm{pk}} \equiv \arg\max_{f} S_L^{\phi}(f) \tag{3.54}$$

and the effective bandwidth is defined as

$$\Delta f \equiv \frac{1}{\left(\overline{S}_L^{\phi}\right)^2} \int_0^{\infty} \left[S_L^{\phi}(f)\right]^2 df. \tag{3.55}$$

The numerical values used in this section are collected in Tables 3.1 and 3.2. These tables use the updated traveling-wave design parameters from the high-frequency thermal-noise model rather than the nominal 5 m reference design from [1].

The GQuEST experiment is a tabletop[2] Michelson interferometer. It has a family resemblance to LIGO, but the design target is different. The arms are short, there are no Fabry-Pérot arm cavities, and the optics are not hung from large seismic isolation chains. Those choices fit a MHz-band stochastic search. Long arm cavities would store light for too long and would roll off the response in the same frequency band where the geontropic signal is expected to be largest.

The readout is also different. GQuEST uses photon counting rather than a conventional homodyne readout. Photon counting measures signal-sideband energy directly. The dark-port light is filtered so that photons in a narrow band around the chosen sideband frequency reach the detector. The experiment then counts those photons and uses the count statistics to estimate the signal power [1, 10].

[2] A table fit for a king.

Quantity	Symbol	Current value
Geontropic signal-strength parameter	α_ϕ	$O(1)$
Interferometer arm length	L	7 m
Interarm angle	Θ	90°
Laser wavelength	λ	1550 nm
Laser frequency	ν	193.4 THz
Power on beam splitter	P_{BS}	$O(10\ \mathrm{kW})$
Arm power	P_{arm}	5000 W
Contrast-defect carrier power	P_{cd}	0.1 W
Power-recycling cavity effective path length	$L_{\mathrm{eff,PRC}}$	10 m
Power-recycling cavity finesse	$\mathcal{F}_{\mathrm{PRC}}$	3000
Laser filter cavity length	L_{LFC}	4.45 m
Laser filter cavity coupler transmissivity	T_{1550}	883 ppm
Laser filter cavity optics-derived finesse	$\mathcal{F}_{\mathrm{LFC,opt}}$	3550
Laser filter cavity measured finesse	$\mathcal{F}_{\mathrm{LFC}}$	4024.5 ± 8.6
Laser filter cavity pole inferred from ringdown	$f_{\mathrm{pole,LFC}}$	8.370 ± 0.018 kHz
Input laser white phase-noise PSD	$S^{\mathrm{LP}}_{\phi,\mathrm{in}}$	$\left(10^{-7}\ \mathrm{rad}/\sqrt{\mathrm{Hz}}\right)^2$
Interferometer DC offset	Λ_{DC}	20 ppm
Interferometer contrast defect	Λ_{CD}	40 ppm
Arm-length difference	L_{Schnupp}	< 0.01 m
Practical output-filter tuning range	ϵ_r	8-40 MHz
Target readout band	ϵ_r	near 17.6 MHz
Effective output-filter bandwidth	$\Delta\epsilon$	~ 25 kHz

Table 3.1: Current interferometer, filtering, and operating-point design parameters for GQuEST.

3.3.1 Experimental Design

The current GQuEST design uses two power-recycled, traveling-wave Michelson interferometers operated close to destructive interference at the output port. Most of the carrier light is reflected out of the input port. A small amount of light exits the dark port because of the intentional DC offset, residual differential motion, contrast defect, and signal sidebands. A power-recycling cavity raises the injected laser power to a beam-splitter power of order 10 kW [1].

The traveling-wave geometry is a design change from a conventional Michelson layout and a slight modification from the main body of [1]. In a conventional layout, the forward and return fields can form a standing optical wave in the beam splitter. That standing wave raises substrate thermorefractive noise (STR) noise. In the traveling-wave layout, the optical paths through the beam splitter are spatially separated, so the standing-wave term is removed [1, 68].

Quantity	Symbol	Current value
Vacuum pressure		10 μPa
Temperature	T	293 K
End-mirror beam radius	w	3 mm
End-mirror radius	a	12.7 mm
End-mirror thickness	h	2 mm
End-mirror substrate		crystalline silicon
End-mirror coating		Ta_2O_5-SiO_2
End-mirror coating thickness	h_c	10 μm
Beam-splitter beam radius	w_{BS}	3 mm
Beam-splitter radius	a_{BS}	19 mm
Beam-splitter thickness	h_{BS}	2 mm
Beam-splitter substrate		crystalline silicon
Signal peak frequency	f_{pk}	$\sim$ 11 MHz
Effective signal bandwidth	Δf	$\sim$ 26 MHz
Peak geontropic PSD	$\overline{S}_L^{\phi}$	$\alpha_\phi \left(4.1 \times 10^{-22}\ \mathrm{m}/\sqrt{\mathrm{Hz}}\right)^2$
Classical-noise ASD near 17.6 MHz	$\sqrt{\overline{S}_L^{\mathrm{c}}}$	$8 \times 10^{-22}\ \mathrm{m}/\sqrt{\mathrm{Hz}}$
Homodyne shot-noise PSD	$\overline{S}_L^{\mathrm{q}}$	$\left(6.2 \times 10^{-19}\ \mathrm{m}/\sqrt{\mathrm{Hz}}\right)^2$
Filtered signal photon flux	$\dot{N}_{\mathrm{pass}}^{\phi}$	$\sim 2.7 \times 10^{-3} \alpha_\phi$ Hz
Filtered classical-noise photon flux	$\dot{N}_{\mathrm{pass}}^{\mathrm{c}}$	$\sim 1.0 \times 10^{-2}$ Hz
Detector dark count rate	$\dot{N}^{\mathrm{d}}$	$< 10^{-3}$ Hz
Twin-IFO separation (for coherence)	L_{sep}/L	0.3
Signal coherence for $L_{\mathrm{sep}}/L = 0.3$	$\overline{C}^{\phi}$	0.88

Table 3.2: Current thermal-noise, readout, and reference sensitivity parameters for GQuEST. The photon-flux values are peak-band estimates obtained from the photon-counting formulas in Section 3.3.4.

The output light from the two interferometers is combined with a beam splitter after one of the beams undergoes a controllable phase delay. This phase delay allows the experiment to measure the coherence of the signal between the two interferometers, which is a key signature of geontropic fluctuations since other noise sources are expected to be uncorrelated. By adjusting the phase delay, the experiment can cancel the signal in the combined output and measure the background noise alone, or it can maximize the signal by making the two interferometer outputs constructively interfere, effectively turning the interferometer on and off. The coherence measurement is important for distinguishing a real geontropic signal from uncorrelated noise sources.

After the phase delay and combination of the two interferometer (IFO) outputs, the

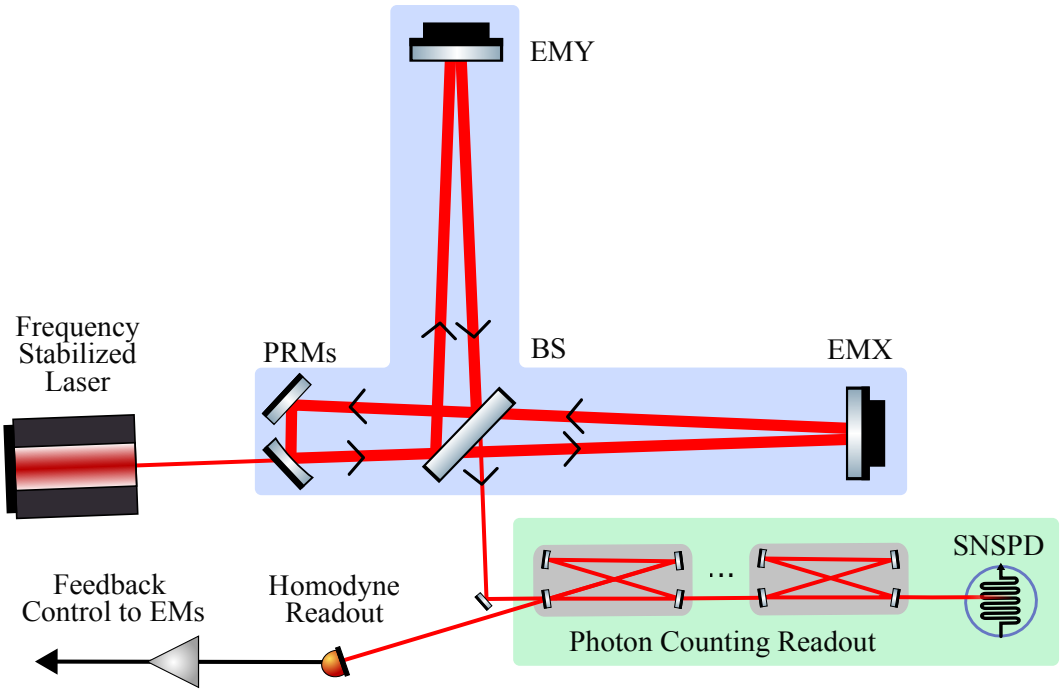


Figure 3.9: Planned traveling-wave layout for a single GQuEST interferometer. The traveling-wave geometry avoids a standing optical wave in the beam splitter, while the output filter cavities isolate the MHz signal sideband before photon counting. Adapted from [1].

light is passed through a chain of narrowband optical filter cavities called OFCs. These cavities transmit light near the optical frequency $\nu + \epsilon_r$, where ν is the laser carrier frequency and ϵ_r is the selected signal-sideband offset. They reject the carrier and most other sideband frequencies to reject the overpowering carrier light as well as sidebands generated from other sources. The filter cavities are tunable over the practical range shown in Table 3.1. While the geontropic signal itself remains broadband, the OFC offset frequency is tuned to target the area with the highest signal-to-noise ratio (SNR), as seen in Fig. 3.11 [1].

The output filter chain uses four bow-tie cavities in series. For a cavity pole near 21 kHz, one cavity gives roughly 50-60 dB of carrier-power suppression across the target MHz band. Four cavities therefore give more than 200 dB of carrier suppression at design frequency [1]. This extreme filtering is needed because contrast-defect carrier light can be many orders of magnitude larger than the expected signal photon flux. The more detailed design, control scheme, and commissioning of these output filter cavities are discussed in Chapter 8.

Photons transmitted through the output filters are detected with a superconducting

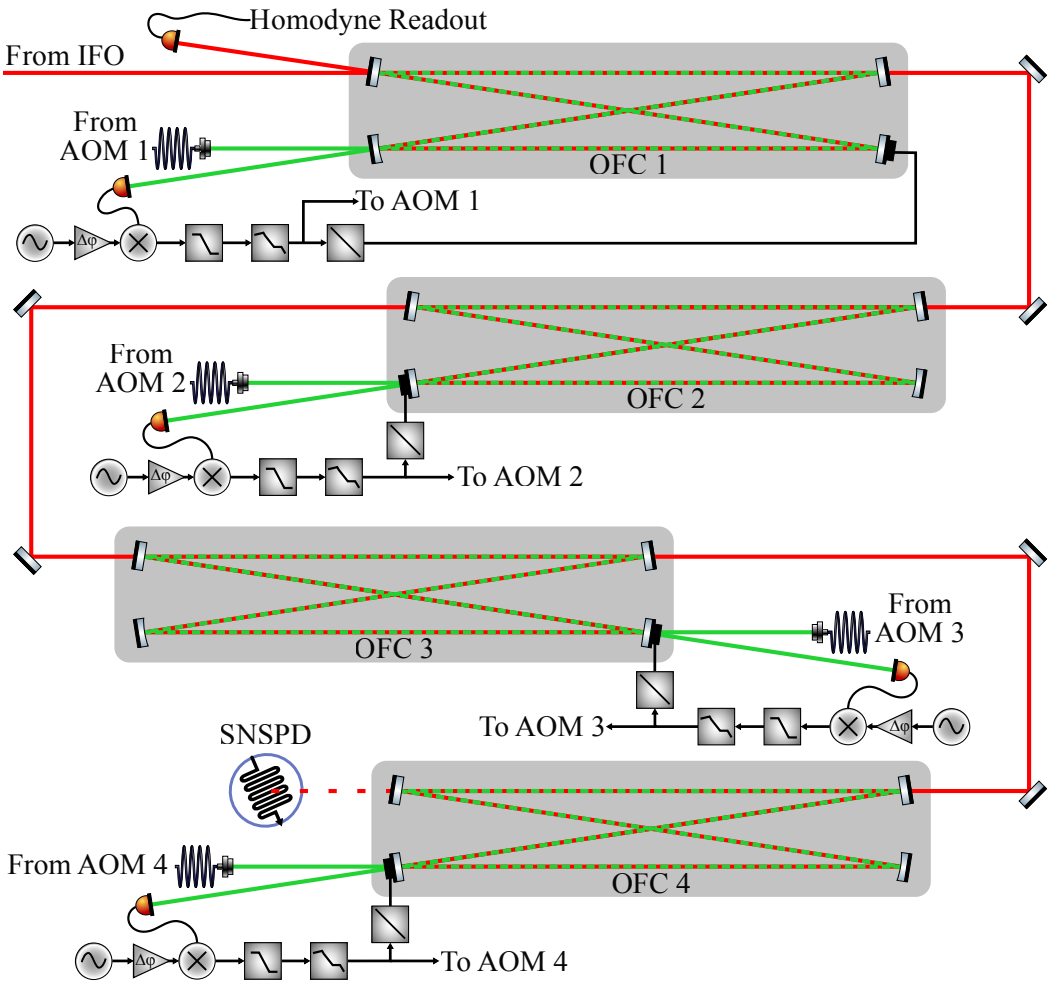


Figure 3.10: Photon-counting readout path with four chained OFCs. The red path represents the 1550 nm light from GQuEST's combined interferometers' dark ports. It is passed through four bow-tie cavities in series so that only the narrow band around the selected detuning frequency reaches the superconducting nanowire single-photon detector (SNSPD). Carrier-frequency light and off-resonant sidebands are rejected by the product of the four cavity transmissions. The green paths are the auxiliary 775 nm control fields, described in Section 8.4, used to hold each cavity on its selected lock point. The control blocks indicate the PDH demodulation, filtering, and feedback used to generate the length-control signals, while the cavity-specific acousto-optic modulator (AOM) offsets set the auxiliary lock points that align the 1550 nm pass bands. The control scheme shown generates an error signal that controls each cavity's length through the piezoelectric transducer (PZT) in a cavity mirror and adjusts the auxiliary laser's frequency by feeding back to the acousto-optic modulator (AOM) through voltage-controlled oscillator (VCO) modulation. In OFCs 2–4 the green 775 nm control beam counter-propagates with respect to the red 1550 nm signal path, reducing the chance that auxiliary-control light or back-scatter is directed along the photon-counting output path.

nanowire single-photon detector (SNSPD), as so much light is filtered out of the interferometers' output that only single photons are left. These detectors register individual photons as digital counts. The interferometer control system uses a different path. Light reflected from the first output filter cavity is used as a homodyne signal for low-frequency length control. That control loop keeps the interferometer near its operating point, while the photon-counting path measures the MHz-band sideband power.

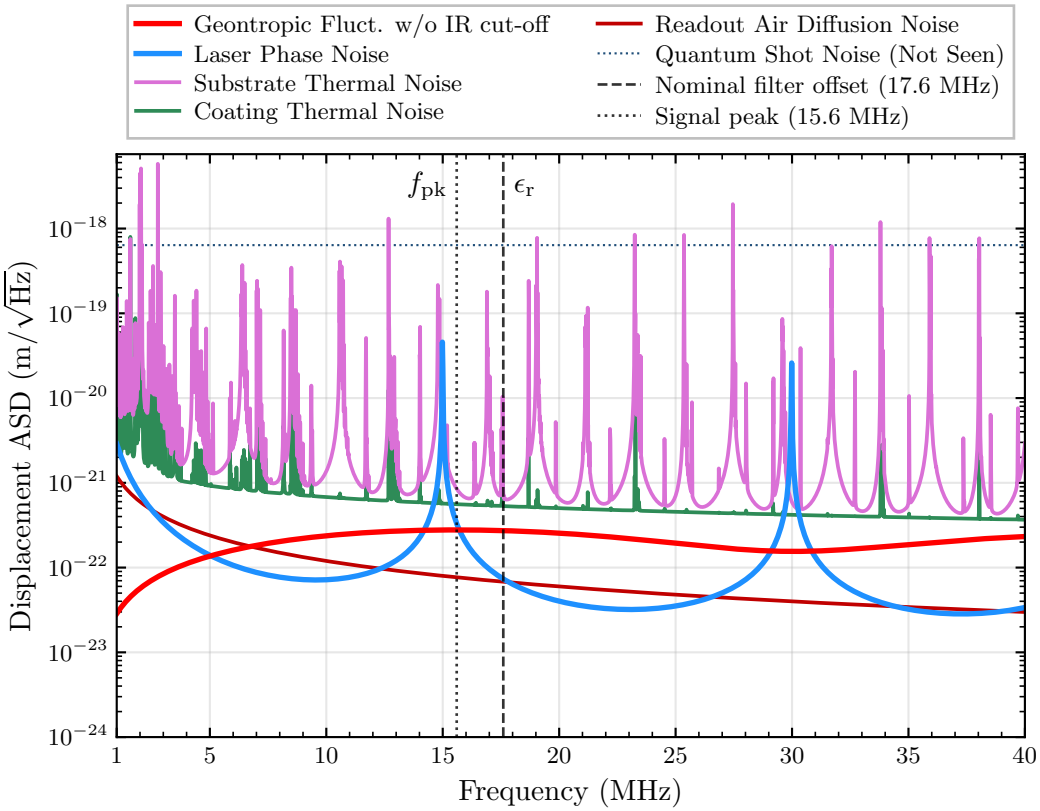


Figure 3.11: Modeled displacement amplitude spectral densities (ASDs) for GQuEST, shown with the expected geontropic signal. The conventional homodyne shot-noise level is shown as a reference, but the photon-counting readout is not limited by that curve. The main classical background near the readout band is Brownian thermal noise (BTN) in the interferometer optics. This is the thermally driven elastic motion of the substrates, coatings, and beam splitter. The thermoelastic and thermorefractive traces are kept separate because they come from temperature fluctuations rather than elastic Brownian motion.

The arm length is a compromise. Longer arms raise the geontropic signal power, since Eq. 3.53 gives $\overline{S}_L^{\phi} \propto L^2$. Longer arms also move the signal peak to lower frequency, which is not always better. At lower frequency, thermal noise is slightly larger, the output filters provide less carrier rejection, and the Brownian thermal

noise (BTN) spectrum has fewer clean local minima where the readout band can be placed. The current 7 m design keeps the signal in the tens of MHz band while allowing the readout to be placed in a low-noise thermal bin.

3.3.2 Laser Phase Noise

The output filter cavities reject carrier light at optical frequency ν, but they do not reject broadband laser noise that falls inside the photon-counting pass band near $\nu + \epsilon_r$. Laser phase noise at this offset produces real sideband photons. Once those photons pass through the output filter cavities, they are indistinguishable from signal photons at the SNSPD. This is the laser-noise problem for photon counting. The relevant quantity is therefore the laser phase-noise PSD at MHz offsets from the carrier, not just the integrated laser linewidth.

The physical origin of this noise is shown in Fig. 3.12. Stimulated emission adds photons to the intracavity field with the same phase, frequency, polarization, and spatial mode as the existing laser field. This is what builds the coherent carrier. Spontaneous emission is different. It occurs at random times and with random phase. Most spontaneous photons leave in the wrong direction or in the wrong spatial mode and do not contribute to the useful laser output. A small fraction, however, couples into the lasing mode or into nearby modes that are later amplified. These photons add a random complex field to the coherent carrier.

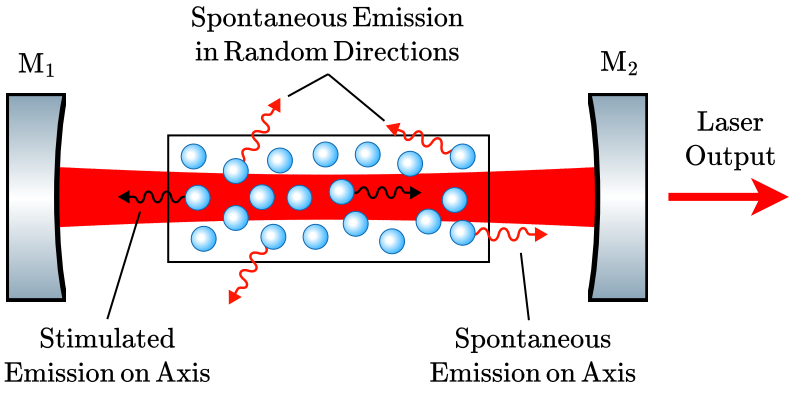


Figure 3.12: Simplified laser-cavity picture used to motivate laser phase noise. Stimulated emission builds the coherent on-axis carrier field. Spontaneous emission occurs with random phase and direction. Most spontaneous photons are emitted away from the lasing mode, but the fraction that couples into the on-axis mode or is amplified into the laser output adds random optical sidebands around the carrier.

This random field can be decomposed into amplitude and phase quadratures relative

to the carrier. If the output field is written as

$$E(t) = [E_0 + \delta E(t)]\, e^{i\omega t}, \tag{3.56}$$

then the phase fluctuation is approximately

$$\delta\phi(t) \simeq \frac{\mathrm{Im}[\delta E(t)]}{E_0}, \tag{3.57}$$

for $|\delta E| \ll E_0$. Spontaneous emission therefore gives random phase kicks to the laser field. The random walk of these phase kicks is the microscopic origin of the laser linewidth. In the MHz band relevant for GQuEST, amplified spontaneous emission also appears as a broadband sideband background around the carrier. Over the narrow readout band, this background is well modeled and measured as a white phase-noise floor [69].

The laser phase noise before filtering is expected to be of order

$$S^{\mathrm{LP}}_{\phi,\mathrm{in}} \sim \left(10^{-7}\,\frac{\mathrm{rad}}{\sqrt{\mathrm{Hz}}}\right)^2, \tag{3.58}$$

near the MHz offsets of interest [1, 69]. This phase noise corresponds to a photon-flux spectral density in the optical sidebands. After the output filter cavities, the laser-noise count rate can be written as

$$\dot{N}^{\mathrm{LP}}_{\mathrm{pass}} = \int_{-\infty}^{\infty} S^{\mathrm{LP}}_{\dot{N}}(\epsilon) F(\epsilon - \epsilon_r) d\epsilon, \tag{3.59}$$

where $S^{\mathrm{LP}}_{\dot{N}}$ is the photon-flux spectral density associated with laser phase noise, and F is the power transmission of the output-filter chain centered at ϵ_r.

For the GQuEST design values, amplified spontaneous emission would produce an output-filtered photon flux of about 2×10^8 Hz if the laser light were not filtered before entering the interferometer [1]. This is far above the expected geontropic signal photon flux. The mitigation is optical filtering upstream of the interferometer output. The seed laser and amplifier are followed by the LFC, and the power-recycling cavity provides a second resonant filtering stage. The LFC pole inferred from the ringdown finesse is 8.370 ± 0.018 kHz, corresponding to 66.5 dB of power suppression at the nominal 17.6 MHz operating frequency. With the 10 m effective power-recycling cavity (PRC) path length and $\mathcal{F}_{\mathrm{PRC}} \simeq 3000$ used in the laser-noise budget, the PRC gives about 60 dB of additional power suppression at the nominal operating frequency. Together these filters suppress the laser amplitude and phase noise by roughly 125 dB in power near the operating band.

After this filtering, the remaining laser phase-noise photon flux is expected to be about 1×10^{-4} Hz, below the signal count rate for α_ϕ of order unity [1]. The suppression is not uniform at all frequencies, since optical filter cavities transmit again at multiples of their FSR. The photon-counting readout band must therefore avoid those residual laser-noise peaks. The detailed LFC design and analysis are given in Chapter 9.

3.3.3 High-Frequency Thermal Noise

Photon counting removes the usual homodyne shot-noise background from the measurement, so the classical noise model becomes the limiting noise. In GQuEST, the largest classical background near the readout band is expected to come from thermal fluctuations in the optics. The thermally driven elastic part of this noise is referred to as BTN. This is the same class of noise that is often called thermal noise, mechanical Brownian noise, or mechanical noise in interferometer literature [1, 68] but should not be confused with Johnson-Nyquist thermal noise.

Brownian thermal noise refers to random thermal motion of dissipative elastic degrees of freedom. The measured optical path length is a mechanical degree of freedom coupled to a thermal bath, and the fluctuation-dissipation theorem (FDT) fixes the spectrum of those random displacement fluctuations from the same mechanical loss that dissipates energy.

The low-frequency thermal-noise formulas used for gravitational-wave detectors are valid in the quasistatic limit. In that limit, the readout frequency is far below the first mechanical eigenfrequency of the optic. This modeling is valid at low frequencies where LIGO is most sensitive. The first resonance of the LIGO test masses is about 6 kHz. However this modeling fails in the MHz band. Millimeter-thick optics have acoustic modes in or near the GQuEST band. The high-frequency model explained in [68] must therefore be used to accurately predict the classical noise in GQuEST. The high-frequency model includes the full acoustic response of the optic, as well as corrections to the thermo-optic noise models that become important when the thermal diffusion length, optical penetration depth, coating thickness, and individual layer thicknesses all become comparable in the MHz band.

The starting point is the FDT [70, 71]. Let δz be the beam-averaged optical path-length displacement, which is the degree of freedom measured by the interferometer output. Let $\chi(\omega)$ be the susceptibility of that degree of freedom to a generalized force with the same spatial profile as the laser beam. The FDT relates the thermal

fluctuation spectrum of δz to the dissipative part of the response:

$$S_z(f) = \frac{4k_B T}{\omega} \left|\mathrm{Im}[\chi(\omega)]\right|, \tag{3.60}$$

where ω is the angular frequency. Equivalently, defining the admittance

$$Y(\omega) = \frac{v(\omega)}{F(\omega)} = i\omega\chi(\omega), \tag{3.61}$$

the same result can be written in terms of the real, dissipative part of $Y(\omega)$. Levin's direct method then replaces the explicit calculation of the admittance with a calculation of the power dissipated by a fictitious drive [71]:

$$\left|\mathrm{Re}[Y(\omega)]\right| = \frac{2W_{\mathrm{diss}}(\omega)}{F_0^2}. \tag{3.62}$$

Substituting this into the FDT gives

$$S_z(f) = \frac{8k_B T}{\omega^2 F_0^2} W_{\mathrm{diss}}(\omega). \tag{3.63}$$

Here F_0 is the amplitude of the generalized force and W_{diss} is the dissipated power. The force does not need to correspond to a real actuator. It is a calculation tool for the same degree of freedom seen by the laser beam. This is why the spatial shape of the fictitious force must match the spatial weighting of the readout.

For a Gaussian beam, the normalized intensity profile is

$$I(r) = \frac{2}{\pi w^2} e^{-2r^2/w^2}, \tag{3.64}$$

where w is the $1/e^2$ intensity radius. The fictitious pressure used in the Brownian thermal-noise calculation is then

$$p(r,t) = F_0 I(r) e^{i\omega t}. \tag{3.65}$$

The same weighting appears in the measured displacement:

$$\delta z(t) = \int u_z(\mathbf{r},t) I(\mathbf{r}) d^2\mathbf{r}, \tag{3.66}$$

where u_z is the surface displacement of the optic in the beam direction. This matching between the applied pressure and the readout weighting is the central point of the direct FDT method.

The low frequency approximation leads to the usual quasistatic substrate Brownian thermal-noise result,

$$S_{z,\mathrm{qs}}^{\mathrm{SBTN}}(\omega) = \frac{4k_{\mathrm{B}}T\left(1-\nu_s^2\right)\phi_s}{\sqrt{\pi}Y_s w\omega}, \tag{3.67}$$

where Y_s is the substrate Young's modulus, ν_s is the Poisson ratio, and ϕ_s is the substrate mechanical loss angle. This expression is a useful check on the high-frequency model. It is not the model that should be used in the GQuEST band.

The high-frequency model used for GQuEST replaces the quasistatic half-space picture with a finite-thickness acoustic model of the optic [68]. The first step is to treat the mirror as an infinite slab with thickness h. This keeps the important physics from the front and back surfaces of the optic without needing the full finite-radius geometry. In this picture, the mirror supports Lamb-wave acoustic motion, and the laser beam drives and reads out only the part of that motion that overlaps with the beam spot.

The slab model is then extended to the real optic geometry. The GQuEST mirrors and beam splitter are finite cylinders, so the continuous slab response becomes a discrete set of acoustic modes. This makes the detailed spectrum more complicated, but the physical picture stays the same: the optic is a small acoustic resonator, and the interferometer samples its thermally driven surface motion through a Gaussian beam profile [68].

This is what gives the BTN trace in Figs. 3.11 and 3.13 its sawtooth shape. The sharp peaks occur when the readout frequency lands near an acoustic resonance of the optic. At those frequencies, the susceptibility is large, so the FDT predicts large displacement noise. Between resonances, the acoustic response is much smaller, giving the valleys where the experiment should be operated.

The peaks are not symmetric. After a resonance, the modes that contribute tend to have shorter transverse structure across the optic face. A Gaussian beam averages over that structure, so surface motion with many positive and negative ripples across the beam spot couples poorly to the readout. This produces a sharp rise into an acoustic resonance and a slower falloff afterward. The same inverse-sawtooth structure appears in Holometer data and is identified in the high-frequency thermal-noise model as substrate and beam-splitter mechanical noise [68, 72].

For GQuEST, this structure is essential to the sensitivity. The output filter bandwidth is much narrower than the spacing between the largest acoustic features, as seen in

Fig. 3.11, so the readout frequency can be placed in a valley of the BTN spectrum rather than on a peak. In the slab approximation, the local minima scale roughly as

$$S_{z,\mathrm{min}}^{\mathrm{SBTN}}(\omega) \propto \frac{k_{\mathrm{B}} T h \phi_s}{Y_s w^2 \omega}. \tag{3.68}$$

This scaling gives the design intuition. The Brownian thermal-noise floor is reduced by using thinner optics, stiffer materials, lower mechanical loss, larger beam spots, and higher readout frequencies. The full GQuEST model uses the finite cylindrical geometry and crystalline-silicon material parameters, but Eq. 3.68 demonstrates why the design is pushed toward thin, stiff, low-loss optics and why the readout band is chosen between acoustic peaks [68].

The coating contribution is calculated differently. The coating is thin, but its mechanical loss is much larger than the substrate loss. Coating Brownian thermal noise has long been a limiting noise source in gravitational-wave interferometers [73, 74]. In the low-frequency literature, the coating contribution is usually computed by evaluating the strain energy stored in the coating and multiplying by the coating loss angle. The high-frequency model keeps the same dissipated-power logic, but uses the finite-frequency elastic response of the substrate to determine the coating strain. This is needed because the coating rides on a substrate whose acoustic response is no longer quasistatic.

The BTN trace in Fig. 3.11 should therefore be read as

$$S_L^{\mathrm{BTN}}(f) = S_L^{\mathrm{SBTN}}(f) + S_L^{\mathrm{CBTN}}(f) + S_L^{\mathrm{BSBTN}}(f), \tag{3.69}$$

where substrate Brownian thermal noise (SBTN) comes from the end-mirror substrates, coating Brownian thermal noise (CBTN) comes from the high-reflectivity coatings, and beam-splitter Brownian thermal noise (BSBTN) comes from beam-splitter substrate motion. The coating term fills in the valleys of the substrate spectrum because the coating loss is high. This is why the coating contribution can be comparable to the substrate contribution even though the coating is only a small fraction of the optic thickness because the coating has a much higher loss angle [68, 73, 74]. In the current GQuEST design, the readout band is chosen so that SBTN is locally small, making CBTN and the combined SBTN plus BSBTN comparable near the low-noise bins around 17.6 MHz [68].

The thermo-optic terms are grouped separately:

$$S_L^{\mathrm{TO}}(f) = S_L^{\mathrm{STR}}(f) + S_L^{\mathrm{STE}}(f) + S_L^{\mathrm{CTE}}(f) + S_L^{\mathrm{CTR}}(f). \tag{3.70}$$

These terms are driven by temperature fluctuations rather than elastic Brownian motion. Substrate thermoelastic noise (STE) and coating thermoelastic noise (CTE) come from thermally driven expansion. STR and coating thermorefractive noise (CTR) come from thermally driven changes in refractive index. Earlier coating thermoelastic and thermo-optic models were developed for the lower-frequency regime of gravitational-wave detectors [75, 76]. In the MHz band, several assumptions used in those models have to be revisited. The thermal diffusion length can become comparable to the coating thickness, the optical penetration depth, or even individual layer thicknesses. The high-frequency model therefore treats the coating thermo-optic terms with the relevant length scales kept explicitly [68].

The STR term is mainly a beam-splitter effect. A standing optical wave in the beam splitter would raise this term by a large factor. This is one of the main reasons the final GQuEST layout is traveling wave. Spatially separating the forward and return paths through the beam splitter removes the standing-wave contribution and keeps STR below the dominant Brownian thermal-noise background [1, 68].

The updated high-frequency model from [68] changes how the noise budget should be read. Earlier quasistatic estimates give a different balance between substrate and coating noise. The updated model finds that GQuEST is designed to operate in frequency bins where SBTN is locally low, so that CBTN and the combined SBTN plus BSBTN set the main classical background. The updated total classical noise near 17.6 MHz is about 9×10^{-22} m/$\sqrt{\text{Hz}}$ [68].

3.3.4 Photon Counting

Photon counting is the readout choice that makes GQuEST different from conventional gravitational wave detectors [2] as well as other QST interferometers [46]. In homodyne readout, a local oscillator field converts signal sidebands into a continuous power time series. The quantum fluctuations of that local oscillator appear as shot noise. In photon counting, the signal-sideband photons are measured directly. Vacuum entering the dark port does not produce signal-sideband photons, so the usual homodyne shot-noise background does not appear in the same way [10].

The single-photon-level signal-sideband readout formalism gives the direct interferometer version of this idea [10]. A stochastic displacement signal creates sideband energy, and that energy can be measured. When classical backgrounds are below the conventional homodyne shot-noise level, photon counting can collect the relevant statistical information faster than a homodyne power time series. The hard

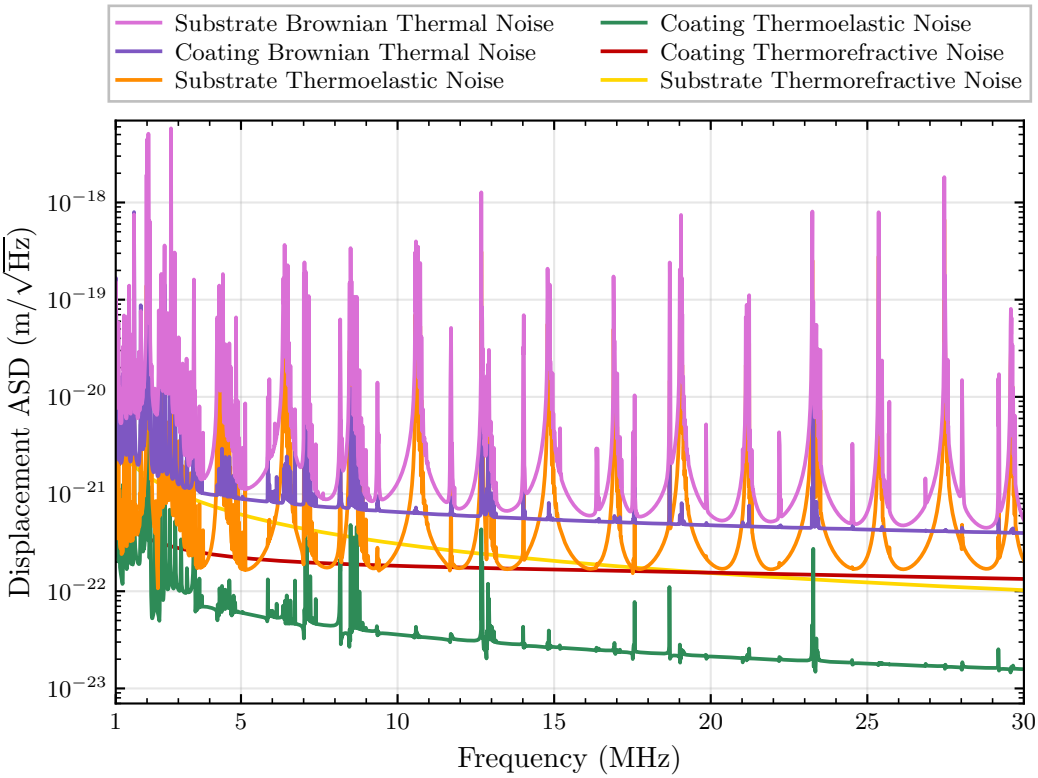


Figure 3.13: Thermal-noise mechanisms relevant for GQuEST that make up the thermal noise budget in Fig. 3.11. At MHz frequencies the quasistatic approximation fails, so the acoustic response of the optic must be included. The sawtooth structure in the BTN trace (seen more prominently at low frequencies) comes from bulk acoustic resonances of the optics and their coupling to the Gaussian beam. The readout band is chosen to lie in a local minimum between these resonances. Adapted from [68].

experimental requirement is also clear from that analysis: carrier light and technical sidebands must be rejected by many orders of magnitude before single-photon detection.

A quantum-estimation treatment of stochastic waveform measurement reaches the same basic lesson from a different direction [77]. Estimating a stochastic PSD is not the same problem as estimating the mean of a deterministic waveform. In the ideal vacuum-input case, number measurement in the right temporal mode is the natural measurement for estimating stochastic power. With loss, the optimal protocol can require non-Gaussian state preparation and non-Gaussian measurements. For the first GQuEST implementation, the practical protocol is simpler: vacuum input states, narrowband optical filters, and low-background photon detectors.

Photon counting can also be useful outside GQuEST. A proposed temporal-mode

photon-counting readout for BNS postmerger signals in future gravitational-wave detectors targets a different science case [78]. The common point is the readout. In high-frequency bands where quantum noise dominates, photon counting can be applied to selected temporal modes while ordinary homodyne readout is retained elsewhere. This makes photon counting relevant to future LIGO-like detectors and Cosmic Explorer (CE), especially above 1 kHz where postmerger information lives and shot noise is a limiting background.

For the GQuEST sensitivity estimate, the conventional homodyne shot-noise displacement PSD is

$$\overline{S}_L^{\mathrm{q}} = \frac{\hbar c}{2kP_{\mathrm{BS}}} \tag{3.71}$$

where k is the optical wave number defined as

$$k = \frac{2\pi}{\lambda} \tag{3.72}$$

and P_{BS} is the power at the beam splitter. This is the reference level for the quantum noise in Fig. 3.11. The photon-counting readout is not limited by this curve. The signal-sideband photon-flux spectral density is related to the displacement spectrum by

$$S_{\dot{N}}^{\phi}(\epsilon) = \frac{S_L^{\phi}(|\epsilon|)}{4\overline{S}_L^{\mathrm{q}}}, \tag{3.73}$$

where ϵ is the optical sideband offset from the carrier. The factor of four follows from the relation between optical gain, sideband counting, and the homodyne shot-noise reference level.

The output filter cavities transmit only a narrow band around ϵ_r. The filtered signal photon flux is

$$\dot{N}_{\mathrm{pass}}^{\phi} = \int_{-\nu}^{\infty} S_{\dot{N}}^{\phi}(\epsilon) F(\epsilon - \epsilon_r) d\epsilon \simeq \frac{\Delta\epsilon\, S_L^{\phi}(\epsilon_r)}{4\overline{S}_L^{\mathrm{q}}}. \tag{3.74}$$

The same conversion applies to classical displacement noise as

$$\dot{N}_{\mathrm{pass}}^{\mathrm{c}} = \int_{-\infty}^{\infty} S_{\dot{N}}^{\mathrm{c}}(\epsilon) F(\epsilon - \epsilon_r) d\epsilon \simeq \frac{\Delta\epsilon\, S_L^{\mathrm{c}}(\epsilon_r)}{4\overline{S}_L^{\mathrm{q}}}. \tag{3.75}$$

The values in Table 3.2 are obtained by evaluating these expressions near the current target readout band.

For one photon-counting interferometer, the signal accumulates as photon counts over time. If the signal count rate, classical-noise count rate, and detector dark count rate are stationary, then

$$\mathrm{SNR}^2_{\mathrm{counts}} = \int_0^T \frac{\left(\dot{N}^{\phi}_{\mathrm{pass}}\,dt\right)^2}{\left(\dot{N}^{\phi}_{\mathrm{pass}} + \dot{N}^{\mathrm{c}}_{\mathrm{pass}} + \dot{N}^{\mathrm{d}}\right) dt} = T\frac{\left(\dot{N}^{\phi}_{\mathrm{pass}}\right)^2}{\dot{N}^{\phi}_{\mathrm{pass}} + \dot{N}^{\mathrm{c}}_{\mathrm{pass}} + \dot{N}^{\mathrm{d}}}. \tag{3.76}$$

In the ideal limit where classical noise and detector dark counts are negligible,

$$\mathrm{SNR}^2_{\mathrm{counts}} \simeq \frac{T\Delta\epsilon}{4}\frac{S^{\phi}_L(\epsilon_r)}{\overline{S}^{\mathrm{q}}_L}. \tag{3.77}$$

For the realistic design, the thermal-noise photon flux is expected to be larger than the signal photon flux. In that limit,

$$\mathrm{SNR}^2_{\mathrm{counts}} \simeq \frac{T\Delta\epsilon}{4}\frac{\left[S^{\phi}_L(\epsilon_r)\right]^2}{\overline{S}^{\mathrm{q}}_L S^{\mathrm{c}}_L(\epsilon_r)}. \tag{3.78}$$

This is the useful design expression. The optical power sets the rate at which signal information is converted into photons, while the classical-noise photon flux sets the statistical background that must be averaged down. The numerical sensitivity should be evaluated with the current values in Tables 3.1 and 3.2, not with the older 5 m reference values.

For comparison, a conventional homodyne stochastic search has the approximate scaling

$$\mathrm{SNR}^2_{\mathrm{homodyne}} \simeq T\Delta f\left(\frac{\overline{S}^{\phi}_L}{\overline{S}^{\mathrm{q}}_L}\right)^2. \tag{3.79}$$

This comparison is the reason photon counting is attractive for GQuEST. In the classical-noise-limited photon-counting case, one power of $\overline{S}^{\mathrm{q}}_L$ appears in the denominator. In the homodyne case, two powers appear. The cost is that the current photon-counting implementation measures only the narrow bandwidth passed by the output filter cavities.

The current design is therefore a narrowband search for a broadband signal. Most of the geontropic signal power is outside the measured band. Future upgrades could use signal recycling, parallel filter cavities, or optical quantum-memory methods to measure more temporal modes at the same time [1, 10, 77]. Those upgrades are not needed for the first GQuEST implementation, but they are possible changes if the photon-counting readout works as expected.

3.3.5 Twin-Interferometer Readout

A single photon-counting interferometer can measure excess counts, but it must know the background count rate. That is difficult because the thermal-noise count rate is expected to be larger than the geontropic signal count rate. The twin-interferometer version of GQuEST addresses this problem by using the spatial coherence of the signal.

For two nearby interferometers, the geontropic signal is expected to be correlated, while the main Brownian thermal noises are not. The outputs of the two interferometers can be combined on another beam splitter before photon detection. The relative optical phase is chosen so that the correlated signal interferes constructively at one output port and destructively at the other. These become the signal and null channels. Uncorrelated noise is split equally on average. The difference between the two count rates then estimates the correlated photon flux [1, 10].

Figure 3.14 depicts this readout in the context of the full traveling-wave GQuEST optical layout. The input laser is first passed through the LFCs so that broadband laser sidebands near the photon-counting band are suppressed before the light reaches the interferometers. The two traveling-wave interferometers then produce separate dark-port fields that carry the signal-sideband information. These fields are recombined with a controllable relative phase so that the correlated signal can be directed toward either the photon-counting output or a null channel. The interferometers' combined output is then filtered by a chain of OFCs before the narrow signal passband reaches the SNSPDs. Each of the OFCs is controlled by an auxiliary laser, at half the carrier's wavelength, to reject carrier light and transmit only the selected signal sideband. The four auxiliary lasers are split off from the main laser and frequency shifted to the appropriate OFC resonance. The auxiliary lasers are not used for readout, but they are needed to control the OFCs and keep them resonant with the signal sideband, which is discussed further in Chapter 8.

The coherent-counting statistic can be written as

$$\mathrm{SNR}_{\mathrm{coh}}^2 = \int_0^T \frac{\left[\left(\dot{N}_{\mathrm{signal}} - \dot{N}_{\mathrm{null}}\right) dt\right]^2}{\left(\dot{N}_{\mathrm{signal}} + \dot{N}_{\mathrm{null}}\right) dt}. \tag{3.80}$$

If both ports are measured with equal efficiency, and if $\overline{C}^{\phi}$ is the signal coherence between the two interferometers near the readout band, this is approximately

$$\mathrm{SNR}_{\mathrm{coh}}^2 \simeq 2\left(\overline{C}^{\phi}\right)^2 \mathrm{SNR}_{\mathrm{counts}}^2. \tag{3.81}$$

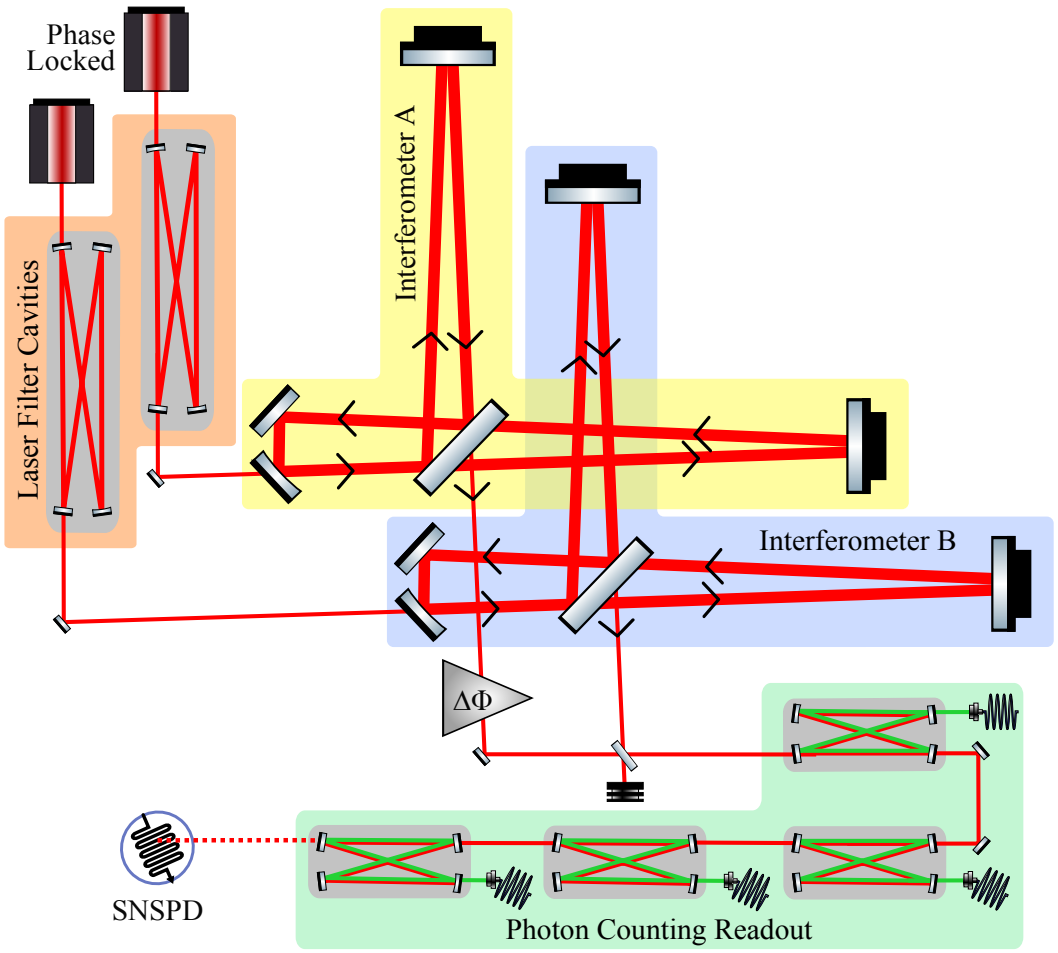


Figure 3.14: Full traveling-wave twin-IFO layout for GQuEST. The laser field is filtered by the LFCs before entering the two interferometers. The two dark-port fields are recombined with a controllable relative phase so that a correlated geontropic signal is directed to one photon-counting output while the orthogonal output acts as the null channel. The photon-counting readout contains the four OFCs, which are controlled with four auxiliary lasers and reject carrier light and transmit the selected signal sideband before detection on an SNSPD.

For the reference coherence calculation with $L_{\mathrm{sep}}/L = 0.3$, the expected coherence near the signal peak is about $\overline{C}^{\phi} = 0.88$ [1]. The main gain is not just the factor in Eq. 3.81. The null channel gives an in situ measurement of the uncorrelated background. That is what makes the twin-interferometer stage the clean test of a correlated geontropic signal.

3.4 Noise Sources

Noise in the LIGO readout can arise from test-mass motion or from optical sensing fluctuations that appear as an equivalent displacement. These contributions are expressed in strain units for comparison with the measured detector noise.

For uncorrelated sources, the expected strain PSD is

$$S_h^{\mathrm{budget}}(f) = \sum_i |C_i(f)|^2 S_i(f), \tag{3.82}$$

where $S_i(f)$ is the one-sided PSD of a noise and $C_i(f)$ is its coupling to the calibrated strain channel. The corresponding strain amplitude spectral densities (ASDs) add in quadrature when assumed to be independent.

Figure 3.15 shows the noise budgets for the Hanford and Livingston detectors during O4. The solid red curve is the measured strain noise, and the solid black curve is the sum of the known contributions. Dashed curves are estimates based on detector parameters or measurements of individual components. Dotted curves are projections obtained from auxiliary channels and their measured couplings to strain. A gap between the measured noise and the sum indicates noise that is not fully accounted for, or a change in the noise or its coupling between measurements [79].

The broad shape of the budget shows which improvements are useful in each frequency band. Below a few tens of hertz, control noise makes a large contribution to the measured spectrum. Near 100 Hz, coating thermal noise is one of the main limitations, while shot noise dominates at high frequencies [79]. The thermal-noise mechanism is described in Section 3.3.3. For LIGO, the low-frequency limit of that description applies across most of the observing band. The remaining noise sources needed to interpret this budget are discussed below.

3.4.1 Seismic Noise

Ground motion reaches an interferometer mirror through its supports. Without isolation, this motion would be much larger than the displacement caused by a gravitational wave. Suspending a mirror as a pendulum reduces the motion transmitted

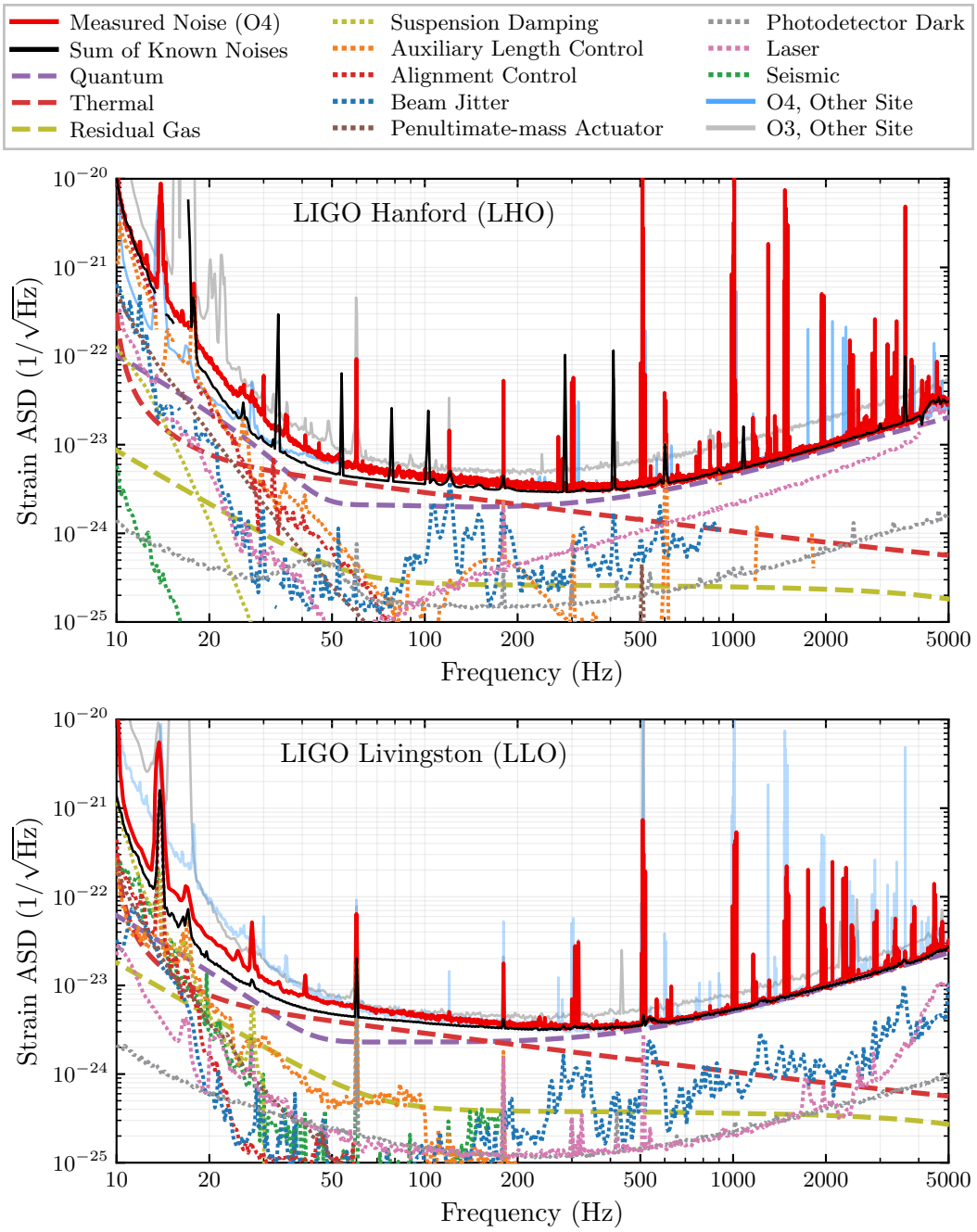


Figure 3.15: Measured noise budgets for the LIGO detectors during the fourth observing run, O4. The individual traces are physical or technical noise sources projected into the calibrated strain readout. Motion and control couplings are largest at low frequencies, thermal noise limits the middle of the band, and quantum shot noise dominates at high frequencies. Adapted from [79].

from its support above the pendulum resonance. In this simple model,

$$S_x^{\mathrm{seis}}(f) = |T_{\mathrm{sus}}(f)|^2 \, S_x^{\mathrm{support}}(f), \tag{3.83}$$

where T_{sus} is the transfer function from support displacement to mirror displacement. Well above the pendulum resonance frequency f_0,

$$|T_{\mathrm{sus}}(f)| \sim \left(\frac{f_0}{f}\right)^2 \quad (f \gg f_0), \tag{3.84}$$

so each pendulum stage provides increasing isolation with frequency. Below resonance, the mirror largely follows its support. This is the support-motion response illustrated by the pendulum example in Section 4.4.

The LIGO test masses hang from four-stage suspensions supported by actively isolated platforms. The platforms reduce low-frequency motion, where the pendulums alone provide little isolation, and the suspension stages attenuate motion in the observing band [80, 81]. This combination of isolation elements makes the direct seismic contribution fall rapidly with frequency, as shown by the green dotted trace in Fig. 3.15.

Residual ground motion can also disturb the angular alignment of the arm cavities. In the DHARD-yaw model of Chapter 7, seismic disturbances contribute to the environmental noise acting on the suspended optics. The alignment controller supplements the isolation by suppressing low-frequency angular motion, but it also feeds sensing noise back through the actuators, as described in Section 6.2. Its design therefore balances the residual angular RMS needed to maintain alignment against the strain noise introduced by alignment actuation. This control contribution is budgeted separately from direct seismic motion.

Ground motion also changes the local gravitational field through the redistribution of nearby mass. The resulting force on a test mass is called Newtonian noise. It bypasses the suspension and cannot be removed by the mechanical isolation used to limit seismic motion. Although it is not shown separately in this O4 budget, it becomes an important consideration when improving the low-frequency sensitivity beyond the present limits [82].

3.4.2 Quantum Noise

Quantum noise enters through both the optical measurement and the radiation-pressure force exerted on the mirrors. For a displacement measurement with uncor-

related imprecision and back-action, the total is

$$S_x^{\mathrm{q}}(f) = S_x^{\mathrm{imp}}(f) + |\chi_m(2\pi f)|^2 S_F^{\mathrm{ba}}(f), \tag{3.85}$$

where S_x^{imp} is the readout-imprecision PSD, S_F^{ba} is the radiation-pressure force PSD, and χ_m is the mechanical response from force to displacement. These two contributions have different dependences on optical power and frequency [32, 83].

For coherent light, the number of photons detected during a fixed interval has Poisson fluctuations: a mean count N has a standard deviation $\sqrt{N}$. These fluctuations produce shot noise in the photocurrent and limit the precision of the inferred optical phase. With the optical configuration fixed, increasing the interferometer power P reduces the shot-noise-limited strain ASD as $P^{-1/2}$ [32, 84].

The equivalent strain noise also depends on the interferometer response. At high frequencies, the finite storage time of light in the cavities reduces the response to a gravitational wave. The same readout fluctuation then corresponds to a larger strain. This produces the rising high-frequency part of the quantum-noise trace in Fig. 3.15 [85].

Fluctuations in the circulating photon number also produce a fluctuating force on the suspended mirrors. Above the suspension resonances, a freely responding test mass has $|\chi_m(2\pi f)| \simeq 1/[m(2\pi f)^2]$, so radiation-pressure displacement noise is largest at low frequencies. Its amplitude grows as $\sqrt{P}$: increasing the power to improve shot noise also increases the force noise. For an ideal measurement of a single free mass, balancing uncorrelated imprecision and back-action gives the one-sided displacement PSD at the SQL,

$$S_x^{\mathrm{SQL}}(f) = 2\hbar\,|\chi_m(2\pi f)| \simeq \frac{2\hbar}{m(2\pi f)^2}. \tag{3.86}$$

Correlations between imprecision and back-action can allow sensitivity below this SQL and are included, together with the optical response, in the full LIGO quantum-noise calculation [83, 85].

Injecting squeezed vacuum into the antisymmetric port reduces fluctuations in one optical quadrature at the expense of the conjugate quadrature. Phase squeezing reduces shot noise, but the accompanying amplitude fluctuations can increase radiation-pressure noise. In O4, a filter cavity rotates the squeezed quadrature with frequency so that the appropriate quadrature is squeezed in each band. This reduces quantum noise over a broader band than scheme with a fixed squeezing angle

allows [62]. The purple dashed curves in Fig. 3.15 include this frequency-dependent squeezing and the losses that limit its benefit [79].

These squeezing techniques improve a measurement of an optical quadrature. The photon-counting scheme in Section 3.3.4 instead estimates stochastic signal power from the energy in the filtered sidebands. Since optical vacuum alone produces no detector counts, the additive homodyne shot-noise background is avoided [10]. The signal and background counts still fluctuate, as expressed by Eq. 3.76, and changing the readout does not remove radiation-pressure-driven mirror motion. The homodyne shot-noise curve in Fig. 3.11 is therefore a reference for comparison, not a noise floor for GQuEST's photon-counting readout.

3.4.3 Technical Noise

Technical noise depends on the implementation and operating state of the detector. Its contribution is often measured by deliberately exciting a suspected source and observing the response in differential-arm length (DARM). The measured coupling is then applied to the source spectrum during normal operation. Environmental sensors provide witnesses for disturbances such as acoustic vibration or magnetic fields, allowing their projected noise to be compared with the detector sensitivity [86].

Photodetector dark noise is measured with the light blocked and referred to strain using the readout calibration. Beam jitter couples when changes in beam position or angle alter the light reaching the readout. Scattered light provides another path: a small field can leave the main beam, reflect from a moving surface, and return to the path with a varying phase. If the magnitude of the surface motion is large compared with the optical wavelength, this coupling is nonlinear and can spread low-frequency motion into the observing band [84, 86].

The residual-gas trace is an example of a noise estimated from the vacuum conditions. Molecules passing through the beam change the optical path length through fluctuations in the refractive index. Lower gas pressure reduces this contribution, which remains below the measured O4 noise in Fig. 3.15 [80]. Laser fluctuations also couple through imperfect cancellation between the arms, producing the separately budgeted laser-noise trace.

The control-noise traces account for disturbances introduced while maintaining the interferometer's operating point. Alignment control contributes alongside the suspension and auxiliary length controls. These measured noise projections provide the basis for estimating the loss of astrophysical sensitivity and how much can be

recovered through improved controller design, as developed in Section 7.1.1.

The optical response and readout determine how a signal is measured, while the noise budget identifies what limits its detection. For both LIGO and GQuEST, the sensitivity described here depends on keeping the optics aligned and the cavities at their required resonances. Feedback is needed to maintain these conditions, but the control-noise contributions show that holding an operating point is not sufficient by itself. The system must remain stable without excessive noise being introduced through the control actuation. Describing this balance requires a model of the instrument's dynamical response and how it is changed by feedback, which leads to the control-theory framework in Chapter 4.

Chapter 4

CONTROL THEORY

"Now, here, you see, it takes all the running you can do, to keep in the same place"

— Lewis Carroll, *Through the Looking-Glass*

4.1 Introduction

While most people consider control theory to be a sub-field of engineering, its roots lie in the mathematical study of dynamical systems. The earliest and most notable work on the mathematical analysis of control systems can be linked to the Scottish physicist and mathematician James Clerk Maxwell's 1868 paper on the governors [87], which are mechanical devices that regulate the speed at which engines operate. At the time they were often used in steam engines to regulate the speed of the engine. The device consists of a vertical central spinning shaft with two hinged arms attached to rotating weights. As the rotational speed of the central shaft increases, the weights move outward due to centrifugal force, which causes the arms to rise. This motion is then mechanically linked to either a throttle valve or a brake, which reduces the engine speed. If the engine slowed, the opposite would occur, causing the engine to speed up again. These types of mechanical feedback systems were widely in use during the Industrial revolution but were often designed and calibrated empirically by experienced engineers and mechanics who understood their practical operation but lacked a formal theoretical framework. Maxwell's work was the first to provide a rigorous mathematical analysis of the underlying system's stability and response characteristics. It was the beginning of control theory as a formal discipline.

In the 1868 paper, Maxwell asked the question of whether the governor would always bring the engine speed to a steady value after a disturbance, or if it could lead to oscillations or instability. This was a crucial question for the safe and efficient operation of steam engines. If the engine encountered a sudden increase in load, could the governor respond appropriately to maintain the speed, or would it over

adjust and start to oscillate wildly? Maxwell approached this problem by modeling the governor and engine system using differential equations, which he then linearized around an equilibrium point to analyze its stability giving a solution of the form

$$x(t) = A_1 e^{n_1 t} + A_2 e^{n_2 t} + A_3 e^{n_3 t}, \tag{4.1}$$

where the exponents n_i determine the stability of the system.[1] From the linearized equations, Maxwell derived what is now called a characteristic polynomial, whose roots are the values of n_1, n_2, and n_3. For Maxwell's governor-engine system the polynomial takes the form

$$MBn^3 + (MY + FB)n^2 + FYn - FG = 0, \tag{4.2}$$

where M is the moment of inertia of the governor's weights, B is the moment of inertia of the engine and load, F is the friction that damps the system, Y is the stability coefficient of the governor related to its restoring force, G is the gain of the system (how sensitive the system is to the governor's position), and n is an eigenvalue or a closed-loop pole. Maxwell points out that if the roots of the characteristic polynomial take the form $n = a \pm ib$, then the part of $x(t)$ in Eq. 4.1 that corresponds to those roots will take the form

$$e^{at} \cos(bt + \phi_0) \tag{4.3}$$

where ϕ_0 is a constant phase offset set by the initial conditions. The important insight is that the sign of a determines stability: if $a < 0$, the oscillations decay; if $a > 0$, they grow and the system is unstable. Maxwell used this analysis to derive conditions on the parameters of the governor and engine system that would ensure stability. Specifically, he then went on to find the values for the given constants under which all roots of the characteristic polynomial would have negative real parts, ensuring that any disturbance or perturbations to the system would decay over time. Maxwell's analysis of the governor system laid the groundwork for control theory by demonstrating how feedback mechanisms could be analyzed mathematically to determine their stability and performance and generated a flurry of further research on the topic. Edward John Routh, one of Maxwell's former classmates at Cambridge, then went on to develop what is now known as the Routh-Hurwitz stability criterion, also named for Adolf Hurwitz who developed the same result by a different method at the same time.

[1] This is a simplified version of Maxwell's solution. His original solution included a Vt term to account for a constant speed component.

Maxwell's pioneering work into unstable oscillations in feedback systems inspired other physicists to pursue more study into the topic. Notably, Lord Kelvin (William Thomson) was fascinated by the topic of stability in dynamical systems and incorporated Maxwell's findings into his own work [88]. Another topic to come out of Maxwell's work was the topic of self-oscillating systems, which are systems that can sustain oscillations without an external periodic input. This would be the equivalent to finding parameters and tuning a governor such that $a = 0$ in Eq. 4.3, leading to undamped oscillations. One of the most famous examples of a self-oscillating system is the violin, where the interaction between the bow and the string creates a feedback loop that sustains the vibration of the string. This phenomenon was studied in detail by Lord Rayleigh in his 1877 book *The Theory of Sound* [89], where he analyzed various self-oscillating systems and their underlying mechanisms. From these early studies of stable systems and self-oscillations, control theory continued to develop as a formal discipline, with physicists and mathematicians building on the foundational work of Maxwell, Routh, Hurwitz, and others to create a comprehensive framework for analyzing feedback systems that power modern physics and technology.

4.2 Control Systems

The term "system" is so used so often and defined so rarely that it is worth outlining what exactly is meant by the term before diving deeper into control theory. In this context, a system refers to a dynamical object that can be influenced by external input(s) and produces output(s) in response to those inputs. It can be used in a variety of ways depending on the specific application. For example, Maxwell's governor-engine system described in Section 4.1 is a physical system that can be modeled mathematically. A system can refer to the collection of interconnected components or to the components themselves. However complex, these systems can be analyzed and understood using the tools of control theory. Often times these complex systems can be broken down into their components and represented by a number of interconnected systems in a box diagram, that shows the flow of signals in a larger system. It is important to understand the mathematical representation of these lowest-level systems before trying to understand the behavior of the larger system.

4.2.1 Transfer Functions

Possibly the most important concept in control theory is that of the transfer function. A transfer function is a mathematical representation of the relationship between the input and output of a system in the frequency domain. It represents how one thing affects another. For example, in a car's suspension, one could use a transfer function to quantify how the bumps in the road (input) affect the motion of the driver's seat (output). Car suspensions act as "low-pass filters" because they tend to not transmit high-frequency inputs very well. Transfer functions are everywhere in physics; they are used to describe the behavior of electronic circuits, mechanical systems, and even biological systems. Understanding transfer functions is crucial for designing effective control systems.

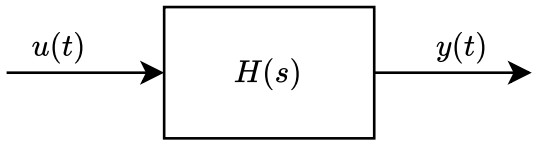


Figure 4.1: A simple input-output system. The system takes an input, $u(t)$, in the time domain and produces an output $y(t)$. The transfer function $H(s)$ describes the relationship between the input and output in the frequency domain.

Given a system as shown in Fig. 4.1, the transfer function $H(s)$ relates the input $u(t)$ to the output $y(t)$ using the Laplace transform. The Laplace transform is a mathematical operation that converts a time-domain signal into the frequency domain with the complex frequency variable

$$s = i\omega, \tag{4.4}$$

where i is the imaginary unit and ω is the angular frequency. The Laplace transform of a time-domain signal $u(t)$ is defined as

$$U(s) = \mathcal{L}\{u(t)\} = \int_0^\infty u(t)e^{-st}dt, \tag{4.5}$$

where $U(s)$ is the Laplace transform of the time-domain signal $u(t)$. From these definitions, the transfer function $H(s)$ can be expressed as the ratio of the Laplace transform of the output $Y(s)$ to the Laplace transform of the input $U(s)$:

$$H(s) = \frac{\mathcal{L}\{y(t)\}}{\mathcal{L}\{u(t)\}} = \frac{Y(s)}{U(s)}. \tag{4.6}$$

This representation can only be used for linear time-invariant systems (LTIs), which are systems that satisfy the properties of linearity and time-invariance. Linearity

means that the system's response to a sum of inputs is equal to the sum of the responses to each individual input and scaling an input by a constant scales the output by the same constant. Time-invariance means that the system's behavior does not change over time. Although many physical systems are not strictly LTIs, they can often be approximated as such over a limited range of inputs and operating conditions,[2] making transfer functions and other formula powerful tools for analyzing and designing control systems in physics.

Transfer functions are typically represented in one of two ways: either as a ratio of two polynomials in the complex variable s:

$$H(s) = \frac{N(s)}{D(s)} = \frac{b_0 + b_1 s + b_2 s^2 + \cdots + b_m s^m}{a_0 + a_1 s + a_2 s^2 + \cdots + a_n s^n}, \tag{4.7}$$

where $N(s)$ and $D(s)$ are the numerator and denominator polynomials, respectively, and b_i and a_i are the coefficients of the polynomials. Alternatively, transfer functions can be expressed in a zero-pole-gain (ZPK) representation, which specifies the zeros, poles, and gain of the system. The ZPK representation can be expressed as the factored polynomial:

$$H(s) = k \frac{(s - z_1)(s - z_2) \ldots (s - z_m)}{(s - p_1)(s - p_2) \ldots (s - p_n)}, \tag{4.8}$$

where k is the gain, z_i are the zeros, and p_i are the poles of the system. The ZPK representation is often more intuitive for understanding the behavior of a system, as it directly relates to its frequency response and stability characteristics. For transfer functions it is also helpful to define the phase and magnitude response of the system. The magnitude response $|H(s)|$ describes how the amplitude of the output signal changes with frequency, while the phase response $\angle H(s)$ describes how the phase of the output signal changes with frequency. These responses can be obtained by evaluating the transfer function at a specific frequency, f, using $s = 2\pi i f$. These responses are often plotted using Bode plots, which show the magnitude and phase response of a system as a function of frequency on stacked plots.

4.2.2 State-Space Representation

Another common way to represent LTIs is through state-space representation, which describes the system using a set of first-order differential equations. This is particularly helpful when describing a system with known equations of motion or

[2] Lots of systems can't be approximated as LTIs over any range, such as chaotic systems but these systems are outside the scope of this discussion. Physicists are often happy to linearize nonlinear systems when they perhaps should not. This is a time honored tradition.

multiple-input-multiple-output (MIMO) systems. In this representation, the system is described by a state vector $\vec{x}(t)$ of dimension n, which contains all the information about the system's current state, an input vector $\vec{u}(t)$ of dimension m, which represents the external inputs to the system, and an output vector $\vec{y}(t)$ of dimension p, which represents the system's outputs. The state-space representation can be expressed in matrix form as:

$$\dot{\vec{x}}(t) = \mathbf{A}\vec{x}(t) + \mathbf{B}\vec{u}(t) \tag{4.9}$$

$$\vec{y}(t) = \mathbf{C}\vec{x}(t) + \mathbf{D}\vec{u}(t), \tag{4.10}$$

where $\mathbf{A}$ is the $n \times n$ state matrix that describes how the system's internal state evolves over time, $\mathbf{B}$ is the $n \times m$ input matrix that describes how the inputs affect the system's internal state, $\mathbf{C}$ is the $p \times n$ output matrix that describes how the internal state affects the outputs, and $\mathbf{D}$ is the $p \times m$ feedthrough (or direct transmission) matrix that describes how the inputs directly affect the outputs without passing through the system's internal state.

The state-space representation can be used to model a wide range of systems including MIMO systems as well as single-input-single-output (SISO) systems. Whereas transfer functions are typically used for SISO systems, state-space representation is more general and its matrices allow for easy manipulation. For example, the transfer function of a SISO state-space system can be expressed as

$$H(s) = \mathbf{C}(s\mathbf{I}_n - \mathbf{A})^{-1}\mathbf{B} + \mathbf{D}, \tag{4.11}$$

where $\mathbf{I}_n$ is the identity matrix of dimension n. This relationship allows for easy conversion between state-space and transfer function representations, depending on the needs of the analysis or design. While theoretically any system can go between state-space and transfer function representations with no loss of information, in practice numerical issues can arise when performing these conversions, especially for high-order systems, systems with poorly conditioned matrices, or systems whose inputs and outputs cover many orders of magnitude. Therefore, when working with complex systems it is very important to choose the representation that is most suitable for the specific analysis and not convert unnecessarily between the two.

4.2.3 Controllability

A state space system is said to be controllable if it is possible to drive the system from any initial state to any desired final state using an appropriate input signal $\vec{u}(t)$ within a finite time. This property is crucial for designing effective control systems,

as it ensures that the system can be manipulated to achieve desired performance objectives. Mathematically, a state-space system is controllable iff

$$\text{rank}[\mathbf{B} \quad \mathbf{AB} \quad \mathbf{A}^2\mathbf{B} \quad \dots \quad \mathbf{A}^{n-1}\mathbf{B}] = n \tag{4.12}$$

where n is the dimension of the state vector $\vec{x}(t)$. If this condition is satisfied, it means that the system's internal state can be fully influenced by the input signals.

4.2.4 Observability

A state-space system is said to be observable if it is possible to determine the system's internal state $\vec{x}(t)$ from its output signals $\vec{y}(t)$ over a finite time interval. Mathematically, a state-space system is observable iff

$$\text{rank}\begin{bmatrix} \mathbf{C} \\ \mathbf{CA} \\ \mathbf{CA}^2 \\ \vdots \\ \mathbf{CA}^{n-1} \end{bmatrix} = n \tag{4.13}$$

where n is the dimension of the state vector $\vec{x}(t)$. If this condition is satisfied, it means that the system's internal state can be fully reconstructed from the output signals.

4.3 Feedback and Interconnected Systems

Once the basic building blocks of control theory are understood, the next step is to connect these systems and examine how feedback changes their behavior. By connecting multiple systems together, it is possible to model more complex behavior than any single component exhibits. Feedback allows the output of a system to be returned to its input, creating a closed-loop system. This can stabilize an unstable system or reduce the effects of noise. In many precision experiments, noise reduction is the main reason to use feedback.

4.3.1 Interconnected Systems

Interconnected systems are systems that are composed of multiple subsystems that can be connected in a variety of ways to create a larger system. The hope is that this larger system can create a more representative model of a physical system or achieve a more complex behavior than any of the individual subsystems could on their own. For example, in a car's suspension system, the shock absorbers, springs,

and tires can be modeled as interconnected systems that work together to provide a smooth ride. Related to control systems, interconnected systems are often used to model complex physical systems and noise.

Consider a simple example of a double pendulum depicted in Fig. 4.2, which consists of two pendulums connected end to end and only moving small amounts. Each pendulum can be modeled as a simple harmonic oscillator, which is a basic control system. However, when these two pendulums are connected together, they form a more complex system that exhibits chaotic behavior [90]. By modeling each pendulum as a simple harmonic oscillator and restricting the motion to small angles, it is possible to use interconnected systems to analyze the behavior of the double pendulum and understand how the two pendulums interact with each other. For example, in Fig. 4.2 the two pendulums are modeled as two interconnected systems, where the output of one system feeds into the input of the other system.

The double pendulum is a simple example of an interconnected system. The two pendula are connected together and the motion of the top pendulum is the only input to the system. The motion of the bottom pendulum is the output of the system. Here the noise, d_{N}, can be thought of as the motion of the attachment point due to vibrations. By modeling each pendulum as a simple harmonic oscillator, it is possible to analyze the behavior of the double pendulum and understand how the two pendulums interact with each other. This is just one example of how interconnected systems can be used to model complex physical systems and noise. This is a good way to model the movement of the lower pendulum as a function of the movement of the upper pendulum and the noise. It is important to note that the double pendulum is non-linear and this model is only valid for small angles. This depiction of the system is also a large simplification; the two systems are also coupled together, meaning that the output of one system feeds into the input of the other system. This is a common feature of interconnected systems and it can lead to complex behavior that is not present in the individual subsystems.

4.3.2 Feedback Systems

Feedback is the process of taking the output of a system and feeding it back into the input of the system, as depicted in Fig. 4.4. This can be used to stabilize an unstable system or reduce the effects of noise. When a controller is used to shape this feedback, it is called feedback control. Feedback control is widely used in experimental physics.

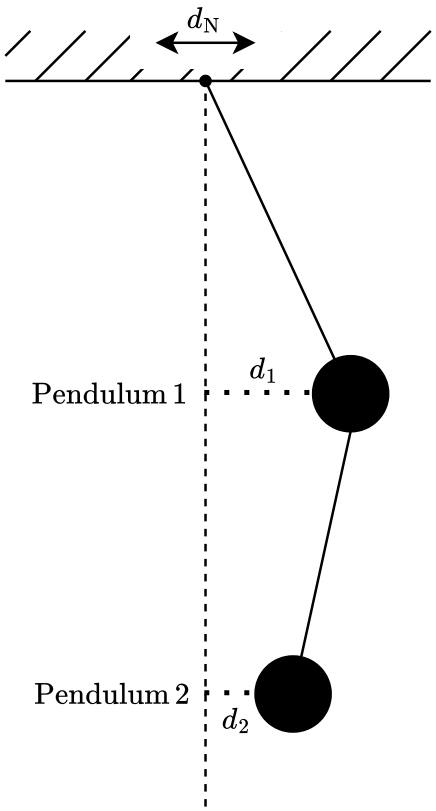


Figure 4.2: A double pendulum, which consists of two pendulums where the lower one is connected to the upper one. The displacements of the pendula are shown as d_1 and d_2. They are connected to a point, which can be thought of as the input to the system. The displacement of the first pendulum, d_1, is the output of the first system and the input to the second system, which produces the output d_2. Noise, d_N, can be introduced into this system as movement of the attachment point due to vibrations.

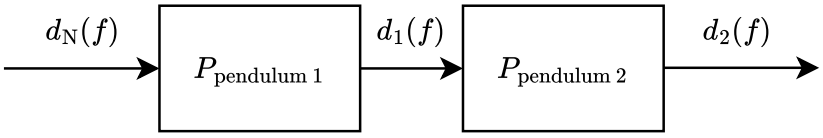


Figure 4.3: A simple interconnected system for a double pendulum, a depiction of which can be found in Fig. 4.2. The input to each pendulum is the displacement of its anchor point. Noise, $d_N(f)$, can be introduced into this system as motion of the attachment point due to vibrations. For the second pendulum, the input is the motion of the first pendulum, $d_1(f)$, which is also the output of the second pendulum. Each pendulum can be modeled as a simple harmonic oscillator, which is a basic control system. When connected together they form a more complex system but still can be analyzed using the tools of control theory.

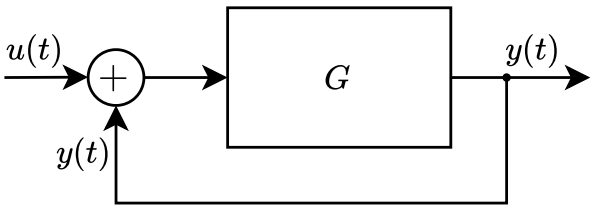


Figure 4.4: A basic feedback system. The output of the system, $y(t)$, is fed back and combined with the input into the input of the system, $u(t)$, through a feedback loop. The feedback loop can be used to stabilize an unstable system, improve the performance of a system, or achieve specific behavior. The system is now called the plant, P, and is the physical system of interest.

Consider the simple example given in Fig. 4.4. The system, P, which is now called the plant as it is the physical system of interest, takes an input, $u(t)$, in the time domain and produces an output $y(t)$. The output is then fed back and combined with $u(t)$ to create a new input to the plant. This changes the transfer function of the system and can be used to achieve desired performance. In order to achieve this desired performance, a controller, K, can be added to the feedback loop as shown in Fig. 4.5. The controller takes the output of the plant, $y(t)$, and processes it before feeding it back and creating a new input to the plant. This can be used to achieve desired performance, such as reducing the effects of noise on the output of the plant.

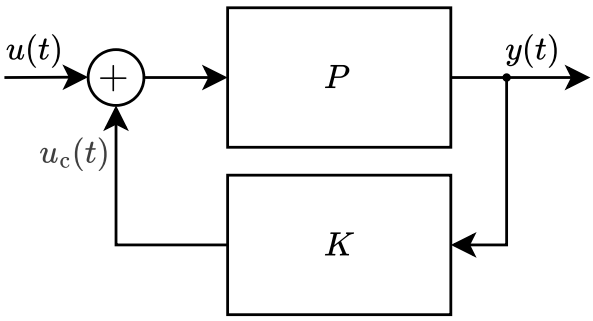


Figure 4.5: This is similar to the feedback system as in Fig. 4.4 but with a controller, K, added to the feedback loop. The controller takes the output of the system and processes it before feeding it back into the input of the system. This can be used to achieve desired performance, such as reducing the effects of noise on the system.

The effects of the feedback loop can be seen by analyzing the transfer function of the system. Without the feedback the transfer function from the input, u, to the output, y, is given by P, which is the transfer function of the plant. This means that if noise were to enter the input of the plant, it would pass through the plant getting shaped by the plant's transfer function and then appear at the output. Now with a feedback

loop with a controller in the loop as shown in Fig. 4.5, the transfer function from the input, u, to the output, y, is given by

$$H(f) = \frac{P(f)}{1 - P(f)K(f)}, \tag{4.14}$$

where $K(f)$ is the transfer function of the controller.[3] This means that if noise were to enter the input of the plant, it would pass through the plant getting shaped by the plant's transfer function and then get reduced by the feedback loop before appearing at the output. Eq. 4.14 is called the closed-loop transfer function and is the effective transfer function of the system with feedback. The transfer function of the system without feedback is given by

$$G(f) = P(f)K(f), \tag{4.15}$$

This is called the open-loop transfer function of the system. Note that the open-loop transfer function is the transfer function of all of the parts in the loop when the feedback loop is broken and not just the plant. The open-loop transfer function is important for understanding the behavior of the system with feedback, as it can be used to analyze the stability and performance of the closed-loop system.

The controller, K, can be designed in a variety of ways to achieve desired performance. The designing of controllers by hand is a skill that takes a lot of practice to hone. It is very difficult to design a controller by hand because it requires the designer to have deep understanding of the system being controlled as well as the ability to balance many different performance metrics like noise reduction, stability, and response time.

4.4 Bode Plots

Bode plots are a graphical representation of the frequency response of a system. They consist of two plots: the magnitude plot, which shows the magnitude of the transfer function as a function of frequency, and the phase plot, which shows the phase of the transfer function as a function of frequency. Since the data from a transfer function like that in Eqs. 4.7 and 4.8 are complex, it is helpful to consider their magnitude and phase separately.

[3] The controller transfer function $K(f)$ is defined to include the sign of the feedback. As a result, the closed-loop transfer function takes the form found in Eq. 4.14 rather than $H(f) = P(f)/(1 + P(f)K(f))$. Under this convention, instability occurs when $G(f) = 1$, whereas in the more common convention (where the negative feedback sign is external to the loop transfer function) the instability condition is $G(f) = -1$.

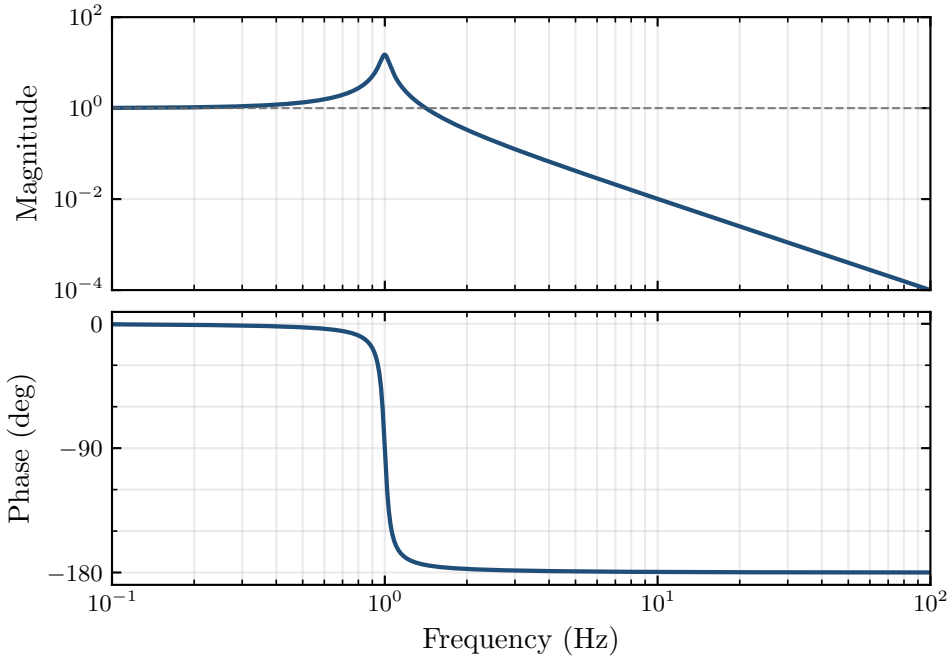


Figure 4.6: Bode plot for a simple harmonic oscillator with resonance frequency $f_0 = 1$ Hz and quality factor $Q = 15$. The magnitude plot shows the logarithmic scale of the transfer function's magnitude, while the phase plot displays the phase shift as a function of frequency.

The Bode plot shown in Fig. 4.6 is for a simple harmonic oscillator with a resonance frequency of $f_0 = 1$ Hz and a quality factor of $Q = 15$. It is a useful example because of how many systems in the world can be approximated as simple harmonic oscillators. This Bode plot can be understood as the behavior of a single pendulum, which is a simple harmonic oscillator.

If one were to hold the end of a pendulum and shake their hand back and forth at different frequencies, the pendulum would respond differently depending on the frequency of the shaking. At low frequencies, the pendulum would follow the shaking of the hand closely, resulting in the pendulum moving about as much as the hand. The magnitude response would be close to 1, meaning that the output (the motion of the pendulum) is about the same as the input (the shaking of the hand). The phase plot at lower frequencies would show that the output is in phase with the input (phase of 0°), meaning that the pendulum moves in sync with the shaking of the hand.

As the frequency of the shaking increases, the shaking frequency would approach the natural frequency of the pendulum. At $f = f_0$ the pendulum would resonate, meaning that it would oscillate with a much larger amplitude than the shaking of

the hand. This is because at resonance, the energy from the shaking is efficiently transferred to the pendulum, causing it to oscillate with a large amplitude. The magnitude response at resonance would show a peak, indicating that the output is much larger than the input. The phase plot at resonance would show a significant phase shift of $-90°$. This will be seen in the physical example as the pendulum lagging a quarter of a cycle behind the shaking of the hand.

As the frequency of the shaking continues to increase beyond the resonance frequency, the pendulum would not be able to keep up with the shaking. The energy from the shaking would not be efficiently transferred to the pendulum, and it would oscillate with a smaller amplitude than the shaking of the hand. This is because at high frequencies, the inertia of the pendulum prevents it from responding quickly enough to the shaking. The magnitude response would show a decrease, indicating that the output is smaller than the input. The phase plot at high frequencies would show that the output lags behind the input by more than $-90°$, meaning that the pendulum moves out of sync with the shaking of the hand.

LIGO uses this very same principle to design the suspension system for the test masses. The test masses are suspended by a series of pendulums, which act as low-pass filters to reduce the effect of seismic noise on the test masses. As seismic noise shakes the suspension point of the test masses, the pendulums filter out the high-frequency components, reducing the impact of the noise.

While modern control theory has moved beyond the use of Bode plots for most aspects of control design work, LIGO still uses Bode plots for a large part of its design and analysis work. Bode plots are a powerful tool for understanding the frequency response of a system and can provide valuable insights into the behavior of a system, especially when it comes to noise reduction and stability analysis, but there are many other more advanced tools that can be used for these purposes as well.

4.5 Stability

Stability is a crucial property of control systems, as it determines whether the system will return to its equilibrium state after a disturbance. A system is considered stable if, after a disturbance, it returns to its equilibrium state or remains bounded within a certain range. If a system is unstable, it will diverge from its equilibrium state and can lead to unbounded behavior. The stability of a system can be analyzed using various methods, such as the Routh-Hurwitz criterion, Nyquist criterion, or Bode

plots. Ensuring the stability of a control system is essential for its safe and effective operation, especially in applications where instability could lead to catastrophic consequences.

4.5.1 Routh-Hurwitz Criterion

The Routh-Hurwitz criterion is a mathematical test that can be used to determine the stability of a system by analyzing the coefficients of its characteristic polynomial. To determine the stability of a system using the Routh-Hurwitz criterion, one constructs the Routh array, $\mathbf{R}$, from the coefficients of the characteristic polynomial. The Routh array is constructed by

$$\mathbf{R} = \begin{bmatrix} a_n & a_{n-2} & a_{n-4} & \cdots \\ a_{n-1} & a_{n-3} & a_{n-5} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \tag{4.16}$$

where a_i are the coefficients of the characteristic polynomial $D(s)$, which is the denominator of Eq. 4.7 where the system is the full closed loop system. The stability of the system can then be determined by analyzing the first column of the Routh array. If all the elements in the first column are positive, then the system is stable. If there are any sign changes in the first column, then the system is unstable. The Routh-Hurwitz criterion provides a systematic way to determine stability without having to compute the roots of the characteristic polynomial, which can be difficult for high-order systems. In state space systems the Routh-Hurwitz criterion can be applied to the characteristic polynomial of the system, which is given by

$$D(s) = \det(s\mathbf{I}_n - \mathbf{A}), \tag{4.17}$$

where $\mathbf{I}_n$ is the identity matrix of dimension n and $\mathbf{A}$ is the state matrix of the system. The Routh-Hurwitz criterion is checking that all the roots of the characteristic polynomial have negative real parts, which ensures that any disturbance or perturbations to the system will decay over time. Examples of these systems are given in Fig. 4.7, where the stable system returns to its equilibrium state after a disturbance because the real part of all roots is negative ($\Re(p) < 0$). The unstable system diverges from its equilibrium state because at least one root has a positive real part ($\Re(p) > 0$). The marginally stable system oscillates indefinitely without returning to equilibrium or diverging because at least one root has a zero real part ($\Re(p) = 0$) and no roots have positive real parts.

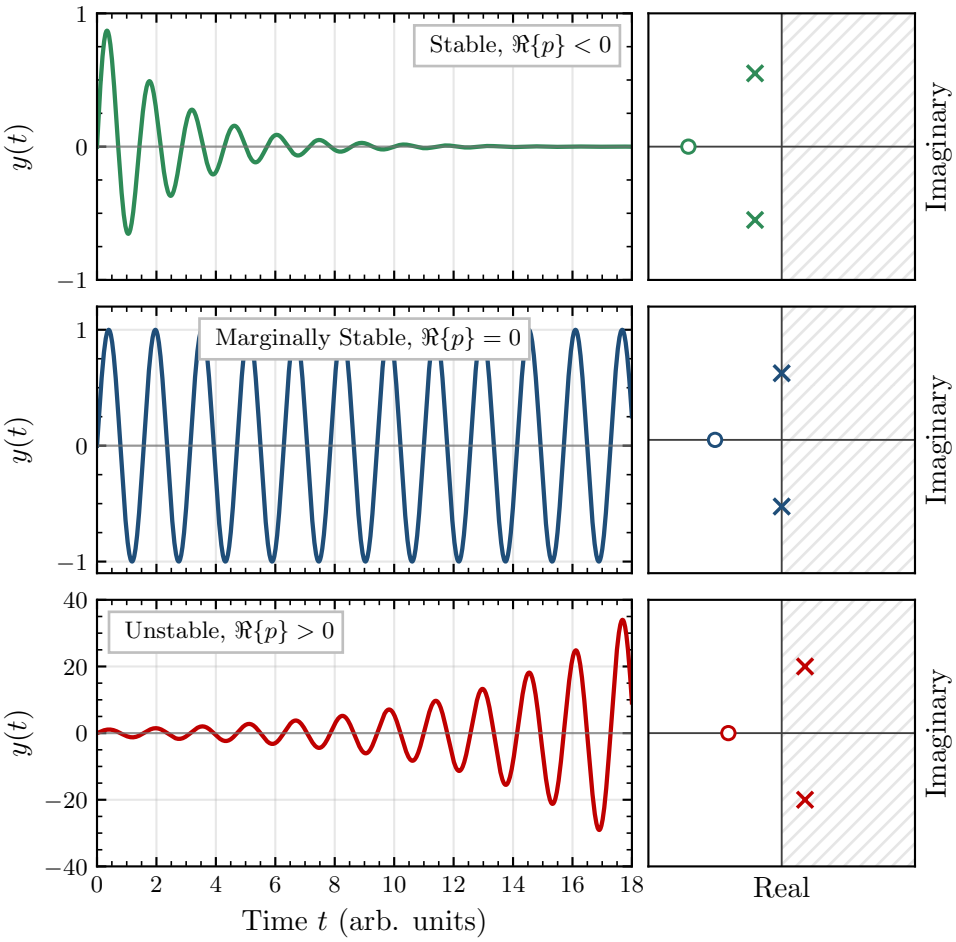


Figure 4.7: Response of stable, marginally stable, and unstable systems to a disturbance. The stable system returns to its equilibrium state after a disturbance because the real part of all roots is negative ($\Re(p) < 0$). The unstable system diverges from its equilibrium state because at least one root has a positive real part ($\Re(p) > 0$). The marginally stable system oscillates indefinitely without returning to equilibrium or diverging because at least one root has a zero real part ($\Re(p) = 0$) and no roots have positive real parts. The poles and zeros of the system are shown in the complex plane on the right of each figure, where the poles are represented by "x" and the zeros are represented by "o". The location of the poles in the complex plane determines the stability of the system. Poles with negative real parts correspond to stable systems, poles with positive real parts, which have poles in the shaded region, correspond to unstable systems, and poles with zero real parts correspond to marginally stable systems.

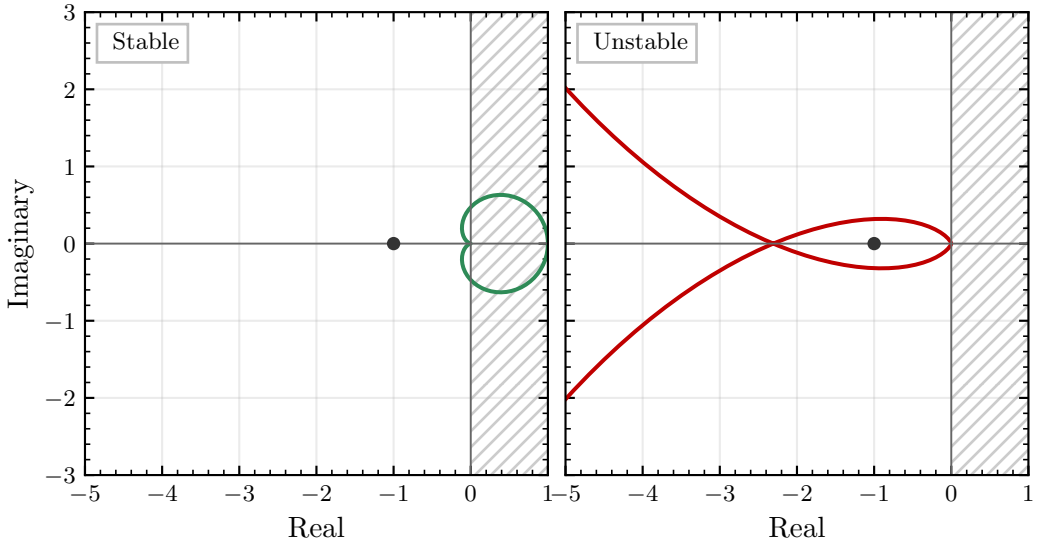


Figure 4.8: Nyquist plots for stable and unstable systems similar to those whose responses are shown in Fig. 4.7. The Nyquist plot is a plot of the complex values of the open-loop transfer function as a function of frequency. The stability of the closed-loop system can be determined by counting the number of encirclements of the critical point (-1,0) in the complex plane, marked as a dot on each plot. For the stable system, there are no encirclements of the critical point as seen on the left, indicating that the closed-loop system is stable. For the unstable system, the critical point is encircled once as seen on the right, indicating that the closed-loop system is unstable.

4.5.2 Nyquist Criterion

The Nyquist criterion is a graphical method that can be used to determine the stability of a system by analyzing the frequency response of its open-loop transfer function. By plotting the Nyquist plot, which is a plot of the complex values of the open-loop transfer function as a function of frequency, one can determine the stability of the closed-loop system by counting the number of encirclements of the critical point (-1,0) in the complex plane. The Nyquist criterion provides a visual and intuitive way to assess stability and is particularly useful for systems with complex dynamics.

The Nyquist criterion is a graphical method that can be used to determine the stability of a system by analyzing the frequency response of its open-loop transfer function. Bode plots can also be used to analyze the stability of a system by examining the gain and phase margins. By using these methods, one can ensure the stability of their systems while achieving desired performance objectives. Nyquist plots are particularly useful tools for those designing controllers by hand as they convey a lot of information about the system's stability and performance in a single plot. For

those learning controls they can be a bit overwhelming at first causing many to avoid them, but with practice they become an invaluable tool for understanding and designing control systems.

4.5.3 Stability Margins

Stability margins are measures of how close a system is to instability. Systems are either classified as stable or unstable, but in practice it is often useful to quantify how close a system is to instability. If a system is perfectly static, in that it never changes, then a marginally stable system is perfectly fine. However, in practice all systems are subject to some amount of change. This change can be due to changes in the system itself, such as changes in the parameters of the system, or it can be due to changes in the environment, such as changes in the temperature or humidity. If a system is marginally stable, then any small change in the system or the environment can cause the system to become unstable. Therefore, it is important to have some measure of how close a system is to instability, which is what stability margins provide.

Consider the double pendulum interconnected system described in Section 4.3.1 and shown in Fig. 4.3. If a controller is added to this system, it can limit how much the lower pendulum's displacement, d_2, is affected by the displacement of the upper pendulum's attachment point, d_N. For the purposes of this analysis the controller is a simple proportional controller that just adds gain to the feedback system. Thus the open loop transfer function of the system, shown in Fig. 4.9, would be given by $G(f) = P_1(f)P_2(f)K(f)$, where P_1 and P_2 are the transfer functions of the two pendula and $K(f)$ is the transfer function of the controller. When adding a controller it is possible to make a previously stable system unstable, and thus it is important to ensure that the system remains stable. One way to do this is to look at the stability margins.

Phase margin, ϕ_m, is a measure of how much the phase of the open-loop transfer function can change before the system becomes unstable. It is defined as the phase of the open-loop transfer function, $G(f)$ at the unity gain frequency (UGF).[4] The phase margin can be expressed as

$$\phi_m = \angle G(f_\mathrm{u}), \tag{4.18}$$

[4] As a reminder, the phase margin is measured from 0° as a result of the negative feedback convention explained in footnote 3.

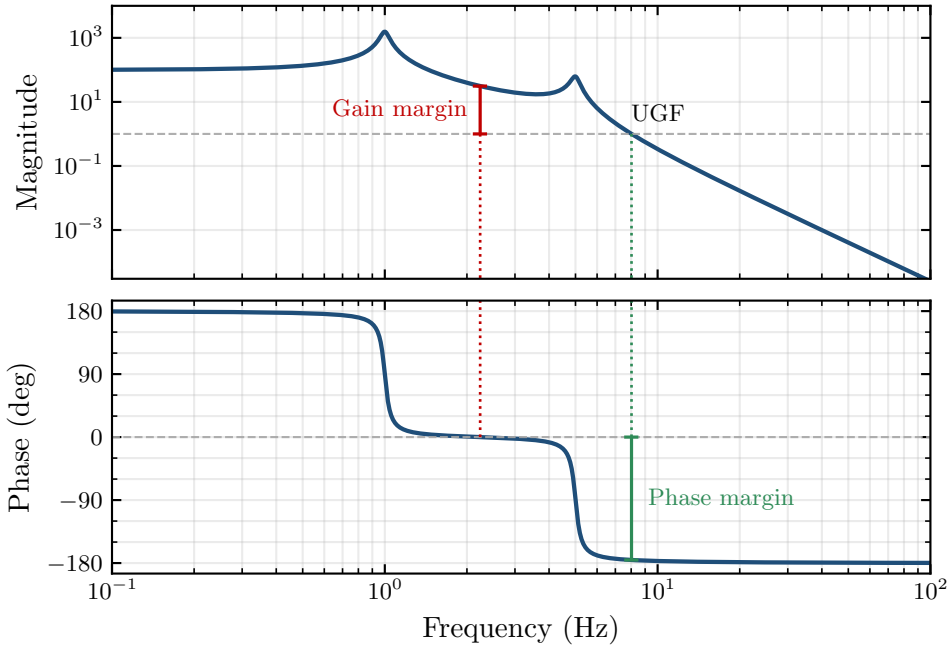


Figure 4.9: This Bode plot shows the open-loop transfer function of a double pendulum system described in Section 4.3.1 with an added controller. The open-loop transfer function is $G(f) = P_1(f)P_2(f)K(f)$, where P_1 and P_2 are the transfer functions of the two pendula and $K(f)$ is the transfer function of the controller. The unity gain frequency (UGF), f_u, is the frequency at which the magnitude of the open-loop transfer function is 1. The phase margin, ϕ_m, is the phase of the open-loop transfer function at the unity gain frequency (UGF). In this example, the phase margin is $\phi_m = 176°$, which indicates that the system is stable and can tolerate a phase change of up to 176° before becoming unstable. The controller that has been added is flat with a gain of 100.

where f_u is the UGF. As $G(f)$ approaches the UGF, the open loop gain approaches 1 as

$$\lim_{f \to f_u} |G(f)| = 1. \tag{4.19}$$

The closed loop-transfer function of the system whose open loop Bode plot is shown in Fig. 4.9, is given by

$$H(f) = \frac{P_1(f)P_2(f)}{1 - P_1(f)P_2(f)K(f)} = \frac{P_1(f)P_2(f)}{1 - G(f)}. \tag{4.20}$$

At the UGF, $|G| = 1$. If the phase of G at the UGF is 0°, then the denominator of the closed-loop transfer function approaches zero, and the closed-loop system becomes unstable. Therefore, intuitively the phase margin can be thought of as how far away from 1 the complex open loop gain is at the UGF.

Gain margin, g_m, is a measure of how much the gain of the open-loop transfer function can change before the system becomes unstable. It is defined as the inverse of the magnitude of the open-loop transfer function at the frequency where the phase of the open-loop transfer function is 0°. The gain margin can be expressed as

$$g_m = \frac{1}{G(f_c)}, \tag{4.21}$$

where f_c is the phase-crossover frequency, defined here as the smallest frequency at which $\angle G(f_c) = 0°$.

At the frequency f_c, the open-loop transfer function is real and positive. If the gain of the system is increased such that $G(f_c) \to 1$, then the denominator of the closed-loop transfer function approaches zero, and the closed-loop system becomes unstable. Therefore, intuitively the gain margin can be thought of as how far the magnitude of the open-loop transfer function is from unity at the frequency where its phase is 0°.

Another way to interpret the gain margin is to consider scaling the open-loop gain by a multiplicative factor k. Instability occurs when

$$kG(f_c) = 1. \tag{4.22}$$

The loop would maintain stability as long as $k < g_m$ and thus Eq. 4.21 can be derived by solving for k when the instability condition is met. The system can be scaled by a factor from 0 to g_m without causing instability. Note that while it is the preference of many physicists to treat gain linearly and not express it in dB, the gain margin should be expressed in dB as $20 \log_{10}(g_m)$, which corresponds to the allowable increase in gain before instability occurs.

4.6 PID Controllers

The most common type of controller is the proportional-integral-derivative (PID) controller. A PID controller applies a control signal based on the error signal,

$$e(t) = r(t) - y(t), \tag{4.23}$$

which is the difference between the desired output, $r(t)$, and the measured output, $y(t)$. In the time domain, the control input $u_c(t)$ is given by

$$u_c(t) = K_P e(t) + K_I \int_0^t e(\tau)\, d\tau + K_D \frac{de(t)}{dt}. \tag{4.24}$$

Taking the Laplace transform (assuming zero initial conditions), the transfer function of the controller is

$$K_{\mathrm{PID}}(s) = K_{\mathrm{P}} + \frac{K_{\mathrm{I}}}{s} + K_{\mathrm{D}}s, \tag{4.25}$$

where K_{P} is the proportional gain, K_{I} is the integral gain, and K_{D} is the derivative gain.

The proportional term provides a control action proportional to the instantaneous error. The integral term accumulates the error over time and eliminates steady-state offsets. The derivative term acts on the rate of change of the error, providing predictive damping and improving transient response. By appropriately tuning these gains, one can shape the closed-loop response to achieve desired performance objectives such as stability, disturbance rejection, and bandwidth.

Part II

Optimal Control

Chapter 5

QUANTIZATION NOISE AND FILTER ARCHITECTURE

"Round and round. Everything repeats. Am I going in circles,
or dare I hope I am on a spiral?"

— C. S. Lewis, *A Grief Observed*

The control theory background in Chapter 4 introduced the standard building blocks of feedback systems. Practical digital implementations add another constraint: every stored number has finite precision. Quantization noise denotes the error introduced when a value from a dense or continuous set is mapped onto a smaller set of representable values. In digital signal processing (DSP), repeated arithmetic operations can accumulate such errors and can raise the noise floor of a digital filter.

Quantization appears whenever a value space with many possible values is represented by a smaller value space. Mapping real numbers to integers provides a simple example: $3/4$ cannot be represented exactly as an integer, so a rule such as rounding maps $3/4$ to 1. Digital computers face the same problem because each numeric data type contains only a finite number of bits.[1] The discussion below focuses on quantization noise produced by floating-point arithmetic.

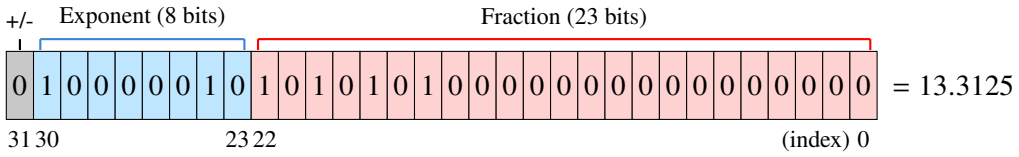


Figure 5.1: An example of 13.3125 expressed as a 32-bit floating-point number according to the IEEE 754 standard [92]. The sign bit is shown in gray, the exponent bits are shown in blue, and the fraction bits are shown in red. The stored exponent is $10000010_2 = 130$. With the single-precision bias of 127, the unbiased exponent is $130 - 127 = 3$. The fraction field represents the bits after the implicit leading 1, so the fraction is $0.1010101_2 = 0.328125$. The represented value is $(-1)^0 \times (1 + 0.328125) \times 2^3 = 13.3125$.

[1] Software methods can extend numeric range or precision, although such methods carry performance and implementation costs [91].

The exact result of a floating-point operation may require more bits than the destination format can store.A rounding rule must therefore map the exact result back onto the nearest representable format. Two common rules are:

- Truncation: The base-β expansion of x is cut after the $(p-1)^{\text{th}}$ stored digit.
- Round-to-nearest: The result is set to the nearest representable floating-point number. In a tie, the representable number with an even least significant stored digit is chosen.

IEEE 754 specifies arithmetic as though an exact result were computed first and then rounded according to the active rounding mode [92]. Consider the real-valued sum

$$0.06 + 54340.0 = 54340.06 \tag{5.1}$$

carried out with 32-bit floating-point values. The decimal number 0.06 is not exactly representable in binary floating point; the nearest single-precision value is approximately 0.0599999986589. The relevant single-precision bit patterns are

$$0.06 \mapsto \texttt{00111101011101011100001010001111}, \tag{5.2}$$

$$54340.0 \mapsto \texttt{01000111010101000100010000000000}, \tag{5.3}$$

$$\mathrm{fl}_{32}(0.06 + 54340.0) \mapsto \texttt{01000111010101000100010000001111}. \tag{5.4}$$

The decimal value of the rounded-sum bit pattern is 54340.05859375. Two errors are present. First, representation error appears before addition because 0.06 is stored with an error of approximately -1.3411×10^{-9}. Second, rounding after addition maps an exact intermediate value near 54340.0599999986589 to the nearest representable single-precision number, 54340.05859375. Relative to the real-valued sum 54340.06, the arithmetic result contains an error of approximately -1.40625×10^{-3}. In the present discussion, the latter error is treated as the quantization noise of the operation.

5.1 Quantization Noise in Filters

Digital signal processing applies many arithmetic operations to a sampled signal. As in the addition example above, each finite-precision operation can add a small error. In a block diagram, each error source can be represented as an additive noise block.

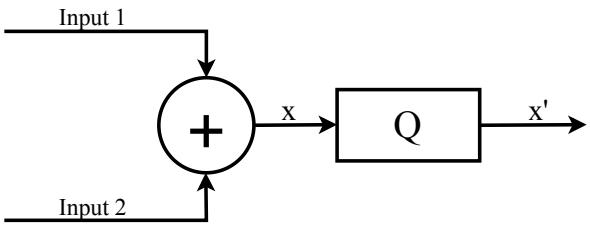


Figure 5.2: A block diagram of a quantizer.

For a rounding quantizer with fixed spacing, the output can be written as

$$x' = q \,\mathrm{round}\left(\frac{x}{q}\right), \tag{5.5}$$

where x is the quantizer input and $q > 0$ is the quantization interval, equal to one least significant bit (LSB) for a uniform fixed-point quantizer. Tie handling has no effect on the root-mean-square estimate below because tie points have zero measure. The input-output relation, shown in Fig. 5.3, is a staircase.

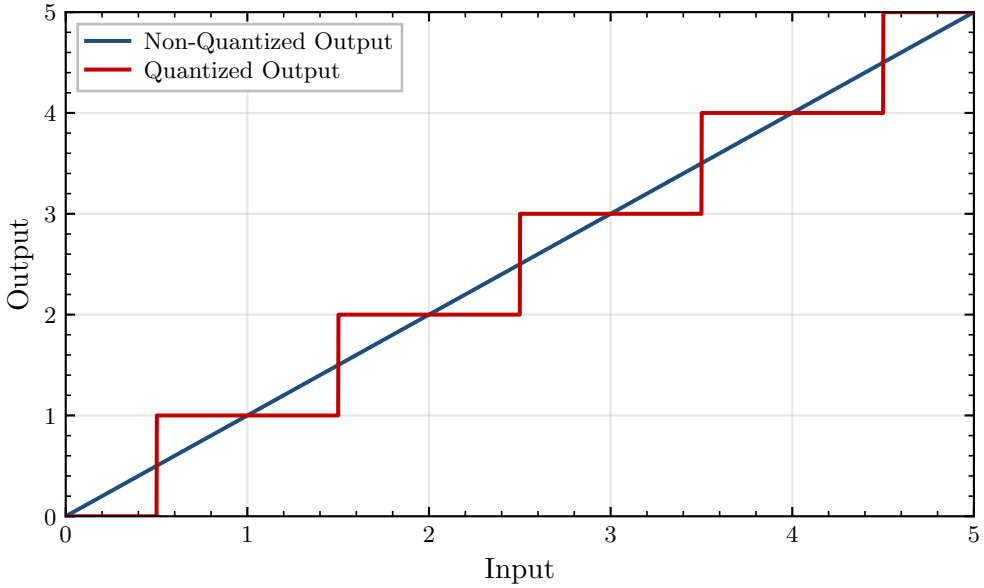


Figure 5.3: The input-output relation of a rounding quantizer with step size $q = 1$. The output x' is the nearest integer to x and changes only at half-integer thresholds.

For $q = 1$, the squared error in Fig. 5.4 lies between 0 and 0.25 and repeats with period q. A useful scalar measure for a quantizer with spacing q is the root-mean-square error,

$$e_{\mathrm{rms}} = \sqrt{\frac{1}{x_2 - x_1}\int_{x_1}^{x_2} e(x)^2 dx} \tag{5.6}$$

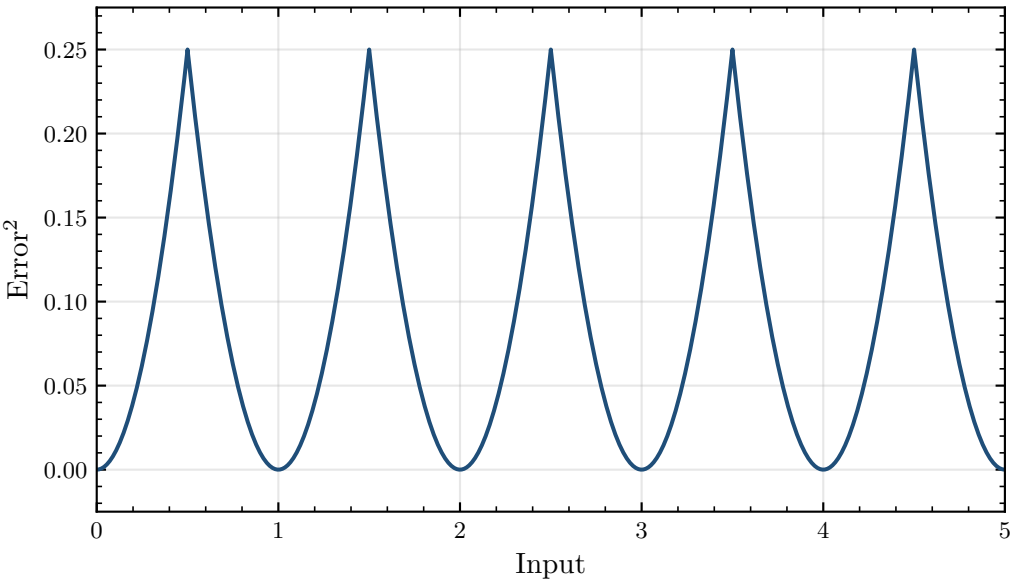


Figure 5.4: The squared error, e^2, between the input and output of a rounding quantizer with $q = 1$. The error is periodic, with a peaked pattern that repeats every quantization step. The maximum squared error is $(q/2)^2 = 0.25$, which occurs when the input is exactly halfway between two quantization levels.

where $e(x) = x' - x$ is the quantization error. Because $e(x)^2$ is periodic, integration over one quantization interval is sufficient. On an interval centered on a quantization level, the error runs from $-q/2$ to $q/2$, so

$$e_{\text{rms}}^2 = \frac{1}{q}\int_{-q/2}^{q/2} \epsilon^2 d\epsilon = \frac{q^2}{12}, \tag{5.7}$$

or

$$e_{\text{rms}} = \frac{q}{\sqrt{12}}. \tag{5.8}$$

The corresponding variance is $e_{\text{rms}}^2 = q^2/12$. Treating quantization error as additive white noise then allows system-level noise estimates. Quantization error is not always white or independent of the signal, but dithering or sufficiently busy input signals often make the white-noise model a good approximation.

For normalized binary floating-point numbers within a single binade, the exact spacing between adjacent representable values is fixed. A format with p significand bits, including the implicit leading bit, has spacing

$$\Delta = 2^{e-(p-1)} \tag{5.9}$$

for values in the binade $[2^e, 2^{e+1})$. For architecture estimates it is useful to write the same local spacing in the form

$$q = A2^{-B}, \tag{5.10}$$

where A is the local arithmetic scale and B is the effective spacing exponent used in the estimate. For exact single-precision spacing in the binade $[8, 16)$, $A = 8$ and $B = 23$, giving $q = 8\,2^{-23} \approx 9.536 \times 10^{-7}$. The double-precision estimates below use $B = 53$ with A chosen from the observed internal signal scale, which approximates the local floating-point spacing at the level needed for the source-count estimates.

The same result gives the standard ideal analog-to-digital converter (ADC) relation, SNR $= 6.02N + 1.76$ dB. For an N-bit bipolar ADC with step size q, the largest sinusoidal peak amplitude without clipping is $q2^{N-1}$. The corresponding RMS signal amplitude is $q2^{N-1}/\sqrt{2}$. With quantization-noise RMS value $q/\sqrt{12}$, the ideal full-scale SNR is

$$\mathrm{SNR}_{\mathrm{dB}} = 20\log_{10}\left(\frac{q2^{N-1}/\sqrt{2}}{q/\sqrt{12}}\right) = 6.02N + 1.76. \tag{5.11}$$

Under the white-noise approximation, the quantization noise is spread uniformly over the two-sided Nyquist band from $-f_s/2$ to $f_s/2$ [93]. The two-sided PSD is therefore

$$N_0 = \frac{q^2}{12f_s} \tag{5.12}$$

and the corresponding amplitude spectral density is

$$n_0 = \frac{q}{\sqrt{12f_s}} \tag{5.13}$$

This chapter uses the two-sided convention for analytic estimates. A one-sided ASD is larger by a factor of $\sqrt{2}$.

5.2 Direct Form II

Direct form II (DFII) is a common filter architecture for implementing a second-order section with fewer delay elements than Direct Form I. For the structure shown in Fig. 5.5, the output is calculated from the difference equations

$$y(n) = b_0w(n) + b_1w(n-1) + b_2w(n-2), \tag{5.14}$$

with

$$w(n) = x(n) - a_1 w(n-1) - a_2 w(n-2), \tag{5.15}$$

where $x(n)$ is the input sample, a_i and b_i are filter coefficients, and n is the sample index.

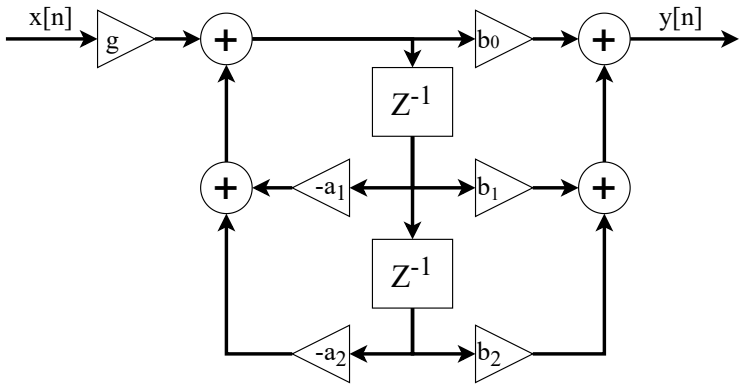


Figure 5.5: A block diagram of a Direct Form II (DFII) second-order section.

In the block diagram shown in Fig. 5.5, finite-precision arithmetic can be modeled by inserting quantizers after additions and multiplications.

5.2.1 Measuring Noise in a DFII Filter

Floating-point arithmetic adds quantization noise at summation nodes and multiplications. The DFII structure in Fig. 5.5 therefore contains ten possible modeled quantization-noise sources when every summation and nontrivial multiplication is rounded. Noise generated in the feedback path enters the recursive state and is filtered by the section transfer function. Noise generated after the feed-forward multiplications enters the output directly. Under the additive white-noise approximation, an output-noise estimate for a DFII section is

$$N_{\text{out}}(f) = N_b \left(\sum_{\text{inputs}} A_i^2 |H(f)|^2 + \sum_{\text{outputs}} A_j^2 \right), \tag{5.16}$$

where $N_b = 2^{-2B}/(12 f_s)$ follows from Eqs. 5.10 and 5.12, A_i and A_j are local scale factors for the corresponding arithmetic operations, and $H(f)$ is the section transfer function from an input-referred noise source to the output.

For the filter shown in Fig. 5.5 with $H(f) = 1$ and a common scale A for all nonzero operations,

$$n_{\text{out}}(f) = \sqrt{N_{\text{out}}(f)} = \sqrt{\frac{2^{-2B}}{12 f_s} 9A^2} = 4.2 \times 10^{-17} \frac{1}{\sqrt{\text{Hz}}}, \tag{5.17}$$

for $f_s = 16384$ Hz, $B = 53$, and $A \approx 56$. The estimate assumes $b_0 = 1$, so the b_0 multiplication acts as a pass-through and does not add a multiplication roundoff source. This leaves nine active sources for the null-filter estimate.

Quantization noise depends strongly on input amplitude, internal state amplitudes, and filter coefficients. A structure with a lower nominal noise count can still perform worse for particular coefficients or inputs. Estimates such as Eq. 5.17 are useful for scaling arguments, but direct measurement remains necessary for a real implementation.

5.3 Low-Noise Biquad Form I

Low-Noise Biquad Form I is used in the LIGO control system and was suggested by Matt Evans [94]. The structure is a second-order section realization related to DFII, but with a different internal state choice. The difference equations are

$$y(n) = x(n) + c_1 w(n-1) + c_2 u(n-1), \tag{5.18}$$

where

$$w(n) = x(n) + a_{11} w(n-1) + a_{12} u(n-1), \tag{5.19}$$

$$u(n) = u(n-1) + w(n-1), \tag{5.20}$$

and the coefficients a_{11}, a_{12}, c_1, and c_2 can be calculated from the DFII filter coefficients using

$$\begin{aligned} a_{11} &= -a_1 - 1 & \qquad c_1 &= b_1 - a_1 \\ a_{12} &= -a_2 - a_1 - 1 & \qquad c_2 &= b_2 - a_2 + b_1 - a_1 \end{aligned} \tag{5.21}$$

This form assumes that the section gain has been factored out so that $b_0 = 1$. Substitution of these coefficients gives

$$\frac{Y(z)}{X(z)} = \frac{1 + b_1 z^{-1} + b_2 z^{-2}}{1 + a_1 z^{-1} + a_2 z^{-2}}, \tag{5.22}$$

so the realization has the same transfer function as the normalized DFII section.

The low-noise biquad reorganizes the internal states so a null filter can pass the input without adding quantization noise. The reorganization adds an internal summation but can reduce the effect of coefficient roundoff and internal quantization for narrow filters.

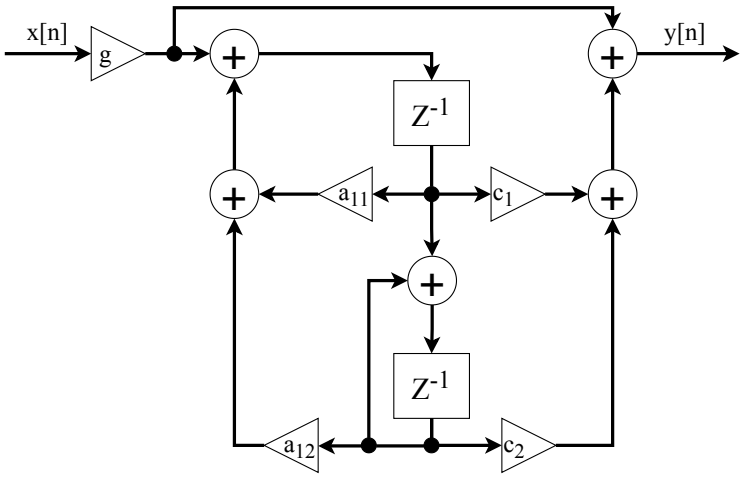


Figure 5.6: A block diagram of the Low-Noise Biquad Form I filter.

A general state-space filter can have a substantially higher computational cost than a cascade of second-order sections. An Nth-order state-space implementation can require $(N+1)^2$ multiplications per output sample, compared with $2(N+1)$ multiplications for common direct-form realizations [95].

5.3.1 Estimating Noise in a Low-Noise Biquad Form I

Quantization-noise estimates for a low-noise biquad are less direct than estimates for a DFII section. A complete analysis follows from a state-space representation and the propagation of each noise source through the corresponding transfer function. A full treatment is outside the scope of the present chapter; [95] gives the standard formulation.

For an order-of-magnitude estimate, the filter with added quantizers in Fig. 5.7 can be simplified.

In the additive-noise model, a quantizer contributes an added noise term. Since $(A+B)+Q = Q+(A+B)$ for arbitrary signals A and B, quantization-noise sources can be moved across adjacent additions without changing the algebraic sum. Moving the noise sources after c_1 and c_2 to the output and the noise sources after a_{11} and a_{12} toward the input gives the simplified diagram in Fig. 5.8.

Further simplification combines the $4Q$ blocks near the output and input. As shown in Fig. 5.9, splitting a path and then adding equal quantization-noise sources to both branches is equivalent, for a noise-power estimate, to adding the same independent sources before the split. The branch amplitudes are not exactly equal in the full filter, so the step is an approximation valid only when the relevant gains are near unity.

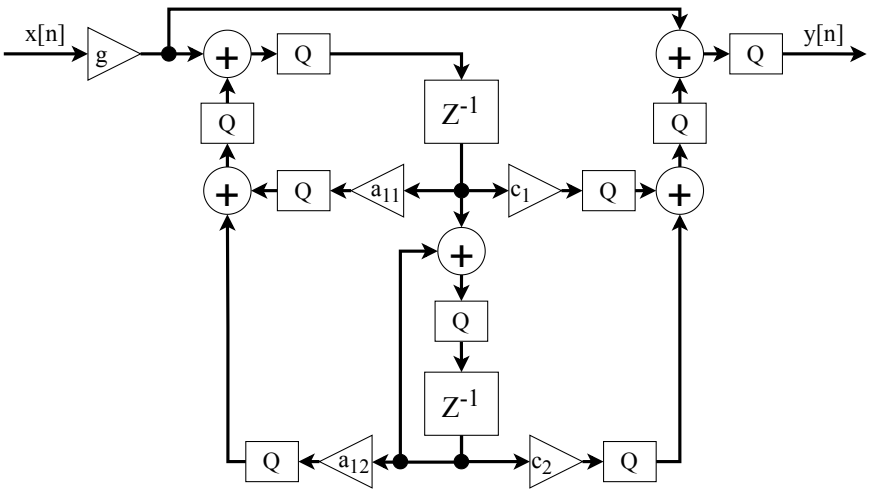


Figure 5.7: The Low-Noise Biquad Form I filter with quantization noise sources added after each arithmetic operation.

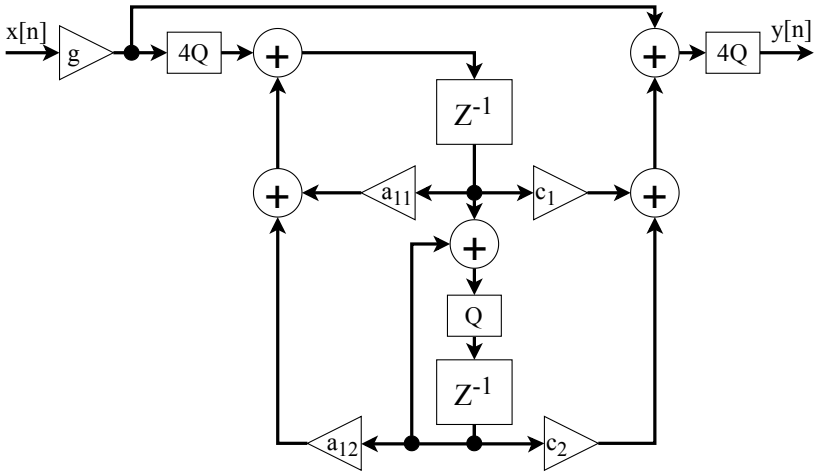
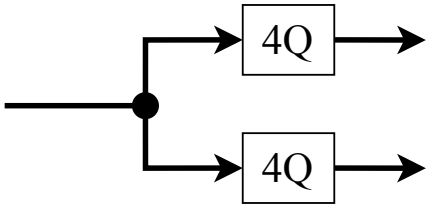
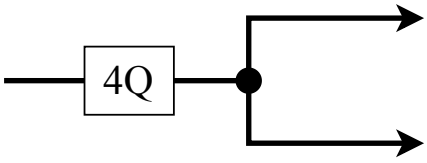


Figure 5.8: A simplified model of the Low-Noise Biquad Form I filter with quantization noise sources.



(a) A diagram showing the junction with one input and two outputs, where four quantization noise blocks are attached to the two outputs.



(b) A diagram showing the junction with one input and two outputs, where four quantization noise blocks are attached to the two inputs.

Figure 5.9: An illustration of the equivalence of placing quantization noise sources before or after a signal split.

After combination of the $4Q$ blocks, one quantizer remains inside the loop. Forward propagation would split the source at the junction and feed part of the source back into the same loop. Backward propagation would require the reverse of the split approximation above. For the present estimate, the remaining quantization-noise source is propagated through the delay block and split at the junction. The approximation is coarse but gives the correct scale for comparison between architectures.

As in the DFII estimate, take $f_s = 16384$ Hz, $B = 53$, and $A \approx 56$.[2] The output-noise estimate is

$$N_{\text{out}}(f) \approx 5N_b A^2 |H(f)|^2. \tag{5.23}$$

For a null filter with $H(f) = 1$, the amplitude spectral density becomes

$$n_{\text{out}}(f) = A\sqrt{5N_b}|H(f)| = 3.1 \times 10^{-17} \frac{1}{\sqrt{\text{Hz}}}. \tag{5.24}$$

Because the estimates above depend on simplified source propagation and approximate internal amplitudes, the measurement method in Section 5.6 provides a more reliable quantization-noise value for a particular filter and input.

5.4 Low-Noise Biquad Form II

Low-Noise Biquad Form II differs from the realization in Section 5.3 by removing the internal summation loop around the second delay block, the block before a_{12} and c_2 in Fig. 5.6 [96]. Removing the summation reduces the number of arithmetic operations and eliminates one quantization-noise source.

The coefficients can be calculated from the difference equations

$$y(n) = x(n) + c_1 w(n-1) + c_2 w(n-2) \tag{5.25}$$

and

$$w(n) = x(n) - a_1 w(n-1) - a_2 w(n-2). \tag{5.26}$$

Equations 5.25 and 5.26 can be put into the same form as the DFII equations in Eqs. 5.14 and 5.15. Beginning with

$$0 = w(n) - x(n) + a_1 w(n-1) + a_2 w(n-2). \tag{5.27}$$

[2] A pass-through null filter with all nontrivial coefficients equal to zero generates no quantization noise in the corresponding zero multiplications or additions.

Adding the right-hand side of Eq. 5.27 to Eq. 5.25 gives

$$\begin{aligned} y(n) &= x(n) + c_1 w(n-1) + c_2 w(n-2) + w(n) - x(n) + a_1 w(n-1) + a_2 w(n-2) \\ y(n) &= w(n) + c_1 w(n-1) + c_2 w(n-2) + a_1 w(n-1) + a_2 w(n-2) \\ y(n) &= w(n) + (c_1 + a_1) w(n-1) + (c_2 + a_2) w(n-2). \end{aligned} \tag{5.28}$$

The resulting expression is

$$y(n) = w(n) + (c_1 + a_1) w(n-1) + (c_2 + a_2) w(n-2). \tag{5.29}$$

Since Eqs. 5.15 and 5.26 are identical and Eqs. 5.14 and 5.29 have the same form, coefficient matching gives

$$c_1 = b_1 - a_1 \qquad\qquad c_2 = b_2 - a_2. \tag{5.30}$$

As with Form I, this coefficient mapping assumes a normalized section with $b_0 = 1$ or a separate section gain.

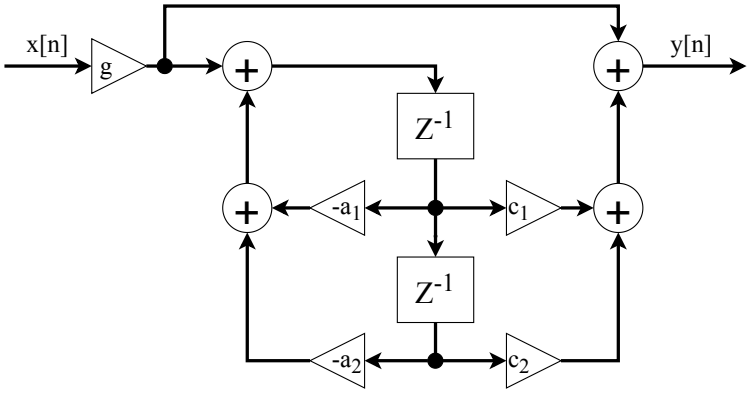


Figure 5.10: A block diagram of the Low-Noise Biquad Form II filter.

5.4.1 Estimating Noise in Low-Noise Biquad Form II

The same source-propagation technique gives an estimate for the second low-noise biquad. Each summation and gain is modeled with an added quantization-noise block as in Fig. 5.2. Noise blocks after the c_1 and c_2 gains are propagated to the filter output, while noise blocks after the a_1 and a_2 gains are propagated to the point before the first split. The resulting intermediate diagram is shown in Fig. 5.11.

As in the first low-noise biquad estimate, the four quantizers near the input cannot be moved past the split unless the corresponding branch sources are matched. Combining the four right-side sources with the input-side sources moves the modeled sources to the input, as shown in Fig. 5.12.

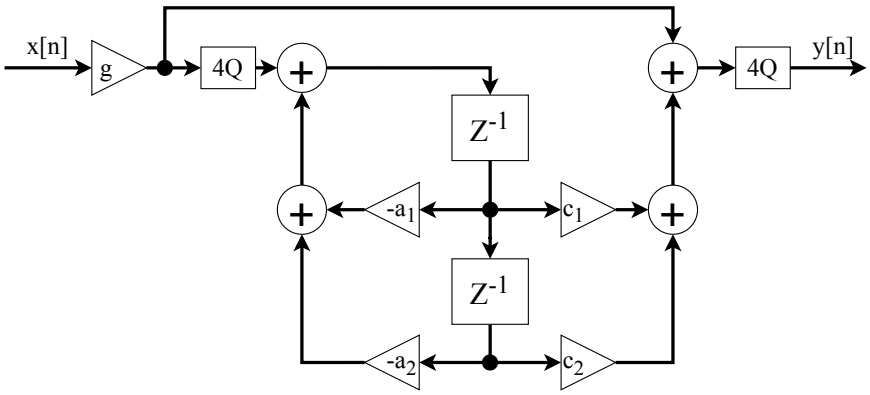


Figure 5.11: An intermediate representation of the Low-Noise Biquad Form II filter with quantization noise sources.

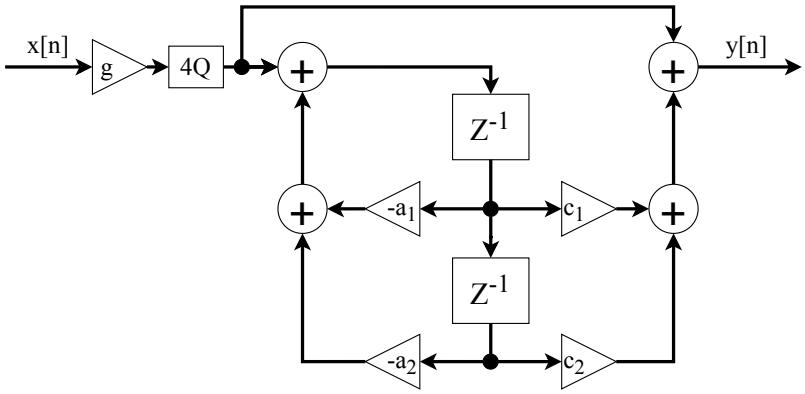


Figure 5.12: The final simplified model of the Low-Noise Biquad Form II filter with all quantization noise sources referred to the input.

With all modeled sources referred to the input, the quantization noise is filtered by the section transfer function:

$$N_{\text{out}}(f) = 4N_b A^2 |H(f)|^2. \tag{5.31}$$

For a null filter with $H(f) = 1$, $A \approx 56$, $B = 53$, $f_s = 16384$ Hz, and $N_b = 2^{-2B}/(12f_s)$, the output amplitude spectral density is

$$n_{\text{out}}(f) = \sqrt{4N_b} A |H(f)| = 2.8 \times 10^{-17} \frac{1}{\sqrt{\text{Hz}}}. \tag{5.32}$$

5.5 State-Space Realizations

State-space representations were introduced for continuous-time LTI systems in Eqs. 4.9 and 4.10 and used to assemble control models in Section 6.3.1. For the digital filters considered here, the corresponding discrete-time SISO form is

similarly

$$\mathbf{x}[n+1] = \mathbf{A}\mathbf{x}[n] + \mathbf{B}u[n], \tag{5.33}$$

$$y[n] = \mathbf{C}\mathbf{x}[n] + \mathbf{D}u[n], \tag{5.34}$$

with scalar input and output and an internal state vector $\mathbf{x}[n]$. This realization has the same transfer function as a direct-form or biquad implementation when the matrices are constructed from the same poles, zeros, and gain. However, the internal coordinates are different. Since floating-point quantization depends on the magnitude of the intermediate values being added and multiplied, two algebraically equivalent realizations can produce different quantization noises.

For one normalized second-order section with DFII coefficients $[g, a_1, a_2, b_1, b_2]$, one convenient controllable realization is

$$\mathbf{A} = \begin{bmatrix} -a_1 & -a_2 \\ 1 & 0 \end{bmatrix}, \qquad \mathbf{B} = \begin{bmatrix} g \\ 0 \end{bmatrix}, \tag{5.35}$$

$$\mathbf{C} = \begin{bmatrix} b_1 - a_1 & b_2 - a_2 \end{bmatrix}, \qquad \mathbf{D} = \begin{bmatrix} g \end{bmatrix}. \tag{5.36}$$

The transfer function of Eq. 5.36 is

$$H(z) = g\frac{1 + b_1 z^{-1} + b_2 z^{-2}}{1 + a_1 z^{-1} + a_2 z^{-2}}, \tag{5.37}$$

which is the same normalized section used above. Cascaded filters may be formed by connecting the state-space systems in series, although it is possible to use a state space system with higher-order matrices to represent a complex filter with a single state-space representation. Thus, the state-space implementation changes the numerical realization, not the intended filter response.

The particular state coordinates are not unique. If $\mathbf{T}$ is nonsingular, the similarity transformation

$$\mathbf{A}' = \mathbf{T}^{-1}\mathbf{A}\mathbf{T}, \qquad \mathbf{B}' = \mathbf{T}^{-1}\mathbf{B}, \qquad \mathbf{C}' = \mathbf{C}\mathbf{T}, \qquad \mathbf{D}' = \mathbf{D} \tag{5.38}$$

preserves the transfer function but changes the state basis. The quantization noise can therefore change under a similarity transformation even though the exact arithmetic input-output response does not. The tool described below tests several such coordinate choices. The Schur form is the same matrix form used by the Riccati and deflating-subspace solvers discussed in Section 6.3.3. In the same way, here it is used as an orthogonal realization transformation that can reduce the dynamic

range of the state update without changing the poles. The `AB09ND` option calls the SLICOT singular-perturbation approximation routine introduced in Sections 6.3.4 and 6.3.5 for state-space conditioning and reduction [97, 98]. In this application the requested order is the original order, so the routine is used mainly as a numerically conditioned realization transformation; when redundant states are detected, it can also return a lower-order minimal realization.

The cost of a full state-space realization is higher than a cascade of second-order sections, because an N-state SISO update involves matrix-vector products rather than a small fixed number of section operations. The motivation for testing it is therefore not that it is always preferable, but that poorly scaled filters may benefit from a different internal basis. The measurement procedure, discussed in Section 5.6, quantifies that tradeoff for each filter individually.

5.6 Measurement

As discussed in Sections 5.2.1 and 5.3.1, quantization noise depends strongly on filter coefficients and input signal. A single architecture-level number cannot capture the noise behavior of every filter realization. A more useful quantity is the measured quantization noise of one specified filter driven by a representative input.

Since quantization noise arises from finite binary precision, increasing numeric precision decreases quantization noise. Equations 5.16 and 5.23 show $N_{\text{out}} \propto 2^{-2B}$, so each additional significand bit reduces the noise power by a factor of four and the amplitude spectral density by a factor of two.

In practical measurements, quantization noise can be approximated by the difference between a calculation using the target precision and a calculation using much higher precision. Because higher precision greatly suppresses quantization noise, the higher-precision output can serve as a reference. Subtracting the target-precision output from the reference output gives an estimate of the quantization noise.

5.6.1 Example: Fourth-Order Notch Filter

Consider a representative two-tone input signal for a digital filter.[3] The first tone is at $f_1 = 1$ Hz, and the second tone is at $f_2 = 4096$ Hz, one quarter of the sampling rate $f_s = 16384$ Hz. With the chosen amplitudes, the sampled input is

$$x[n] = \sin\left(2\pi f_1 \frac{n}{f_s}\right) + 10^{-9} \sin\left(2\pi f_2 \frac{n}{f_s}\right), \tag{5.39}$$

[3] A version of the example appears in [94].

where n is the sample index.

The signal is passed through a fourth-order notch filter located at 1 Hz with a zero quality factor of 1 and a pole quality factor of 10^6.

A 300-second input signal is passed through the digital filter using 64-bit floating-point arithmetic and NumPy's extended-precision `longdouble` type.[4] The scaled input and output are shown in Fig. 5.13. The two-tone input is useful because the notch suppresses the lower frequency while the higher frequency lies at half of the Nyquist frequency and is sampled four times per period.

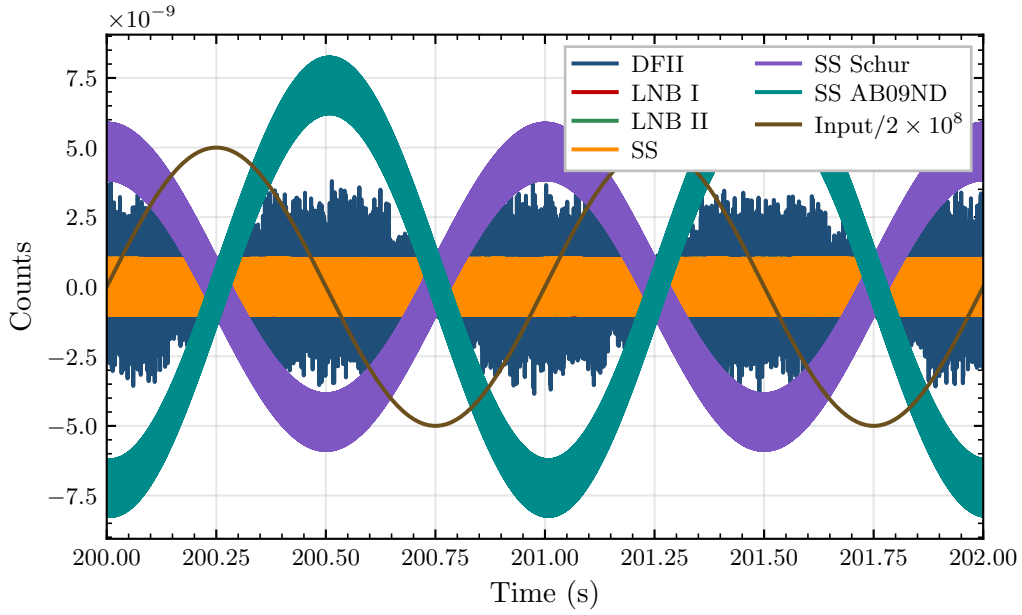


Figure 5.13: The time-domain response of a fourth-order notch filter to the two-tone input signal for DFII, two low-noise biquad realizations, and three state-space realizations. The input contains a unit-amplitude 1 Hz sinusoid and a 10^{-9}-amplitude 4096 Hz sinusoid.

The output ASD in Fig. 5.14 shows the filtered input frequencies at 1 Hz and 4 kHz clearly for the lower-noise realizations. The same filtered tones are hidden by the DFII noise floor. An ideal noiseless filter would produce output only at the two input frequencies, with amplitudes set by the filter response. The broadband nonzero amplitude is therefore attributed to quantization noise. Taking the DFII quantization noise near the high-frequency tone to be approximately 10^{-11} counts/$\sqrt{\text{Hz}}$ and the

[4] On the analysis platform, `longdouble` occupied 128 bits and had a machine epsilon of 1.08×10^{-19}, compared with 2.22×10^{-16} for `float64`. It therefore provided a higher-precision reference, although it was not the IEEE binary128 format. The calculations were performed using the LIGO Lab's Jupyter Hub.

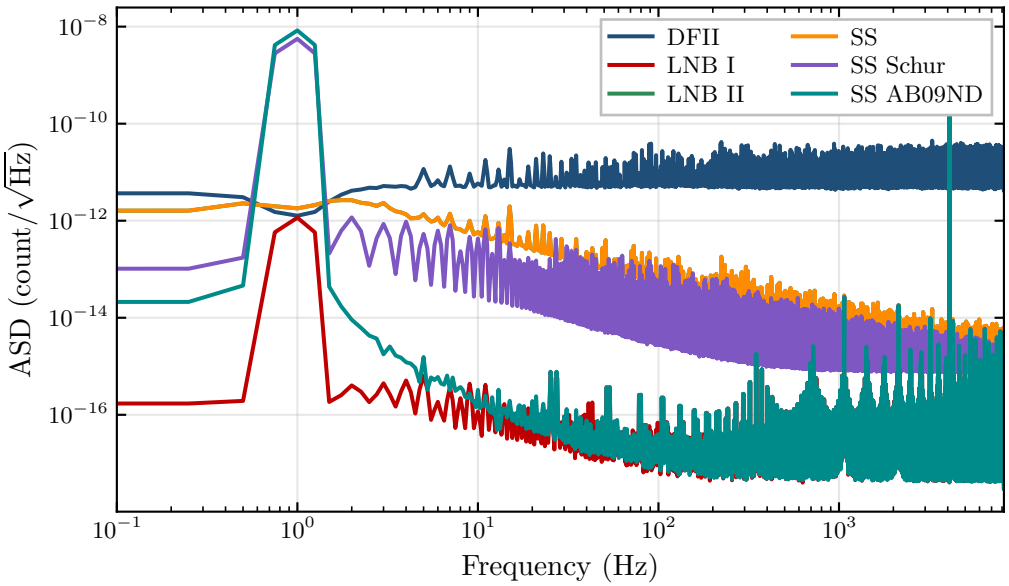


Figure 5.14: The output amplitude spectral density of the fourth-order notch filter for the tested implementations. The realization determines the broadband numerical floor and whether the weak high-frequency tone remains above that floor.

high-frequency tone to be approximately 10^{-9} counts/$\sqrt{\text{Hz}}$ gives $\text{SNR}_{\text{DFII}} \approx 100$. For Low-Noise Biquad Form I and the `AB09ND`-conditioned state-space realization, a nearby noise level of approximately 10^{-17} counts/$\sqrt{\text{Hz}}$ gives SNR $\approx 10^8$ for the same tone; the other tested realizations fall between these limits. For these data, the filter realization determines whether the 10^{-9} tone remains visible.

After filtering with both 64-bit and extended-precision arithmetic, the quantization-noise estimate is the difference between the two outputs. The `longdouble` calculation serves as a practical reference because its quantization noise is far below the 64-bit contribution. The results in Fig. 5.15 show the quantization-noise spectra for the tested realizations, confirm that the noise floor in Fig. 5.14 is quantization noise, and provide a check on the estimated SNR. The subtraction method gives a quantitative noise estimate for a specified filter and input signal.

5.7 Procedure for a Generalized Measurement

Since quantization noise depends strongly on filter input and coefficients, both must be included in any quantitative statement about a filter. The following procedure gives a practical measurement method [84].

First, choose a representative input signal. For an operating experiment, record the filter input while the experiment is in a normal operating state. The input

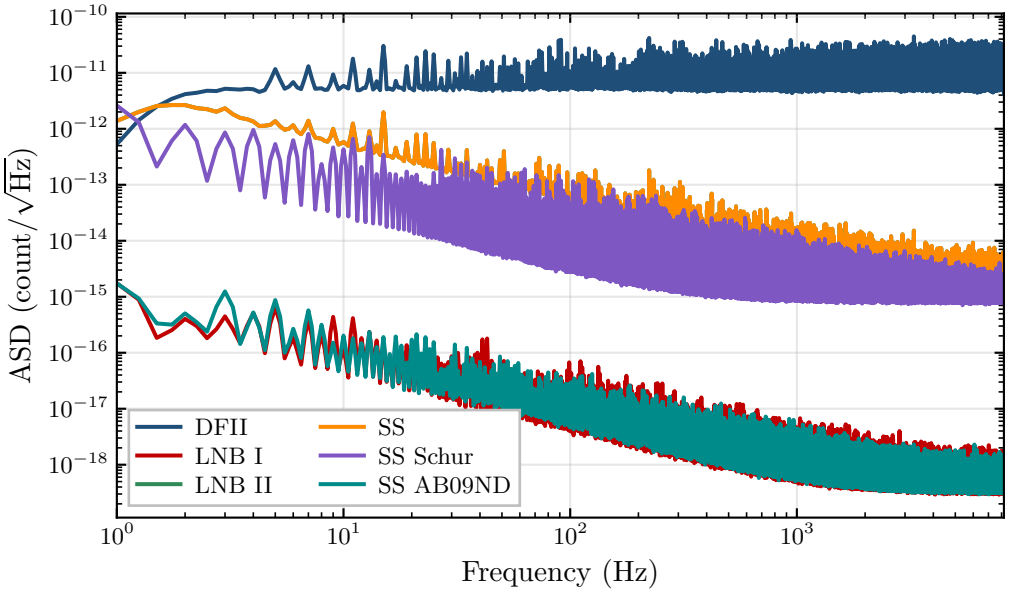


Figure 5.15: The quantization-noise spectrum for the fourth-order notch filter implementations. The spectrum is obtained by subtracting the 64-bit output from the extended-precision `longdouble` (128-bit) reference output and then taking the ASD of the difference.

should include the amplitudes and spectral content that the filter sees during normal use, since both affect the internal state amplitudes and therefore the floating-point spacing. A quiet input can underestimate the quantization noise of a filter that usually runs close to a large carrier or control signal. A record length of at least 25 seconds is usually adequate, although longer records may be needed for low-frequency content, high sample rates, or nonstationary signals.

Next, pass the recorded input through two digital filters: one using the target system precision and one using higher precision. The sample rate, filter structure, gain normalization, and section ordering should match the target implementation. The filter states should also be initialized in the same way for both calculations. If an artificial initial condition is used, the startup transient should be removed before the spectrum is estimated.

The coefficient choice depends on the quantity being measured. If the goal is arithmetic quantization noise alone, the target-precision coefficients should be used in both calculations. Both filters then share the same coefficient roundoff error, and the subtraction isolates the error generated by finite-precision arithmetic during filtering. If the goal is the total numerical error of an implementation, including coefficient quantization, then the reference filter should use higher-precision coefficients and

the target filter should use the coefficients stored by the real system.

Let $y_{\text{target}}[n]$ be the output of the target-precision filter and $y_{\text{ref}}[n]$ be the output of the higher-precision reference filter. The estimated quantization-noise time series is

$$e_q[n] = y_{\text{target}}[n] - y_{\text{ref}}[n]. \tag{5.40}$$

The sign of the subtraction has no effect on the PSD or ASD. The reference precision should be high enough that reference quantization noise is well below the target contribution. Reference precision can be checked by repeating the calculation with still higher precision when available, or by verifying that the result scales with the expected 2^{-B} dependence when the target precision is changed.

After both filter outputs are computed, subtract the outputs and plot the PSD or ASD of $e_q[n]$. The same spectral-estimation settings should be used for all filters being compared: window, segment length, overlap, detrending, and one-sided or two-sided convention. The resulting spectrum estimates the quantization noise produced by the target-precision implementation for the chosen input.

For a control filter, the result is most useful when it is compared with the filtered signal level or with a requirement on output noise. A frequency-dependent numerical SNR can be defined as

$$\text{SNR}_q(f) = \frac{n_y(f)}{n_q(f)}, \tag{5.41}$$

where $n_y(f)$ is the output ASD and $n_q(f)$ is the ASD of $e_q[n]$. The ratio should be interpreted only in frequency bands where the output signal or requirement is meaningful. A large ratio indicates that numerical noise is far below the filtered signal. A small ratio identifies filters where coefficient scaling, section ordering, or filter realization should be checked.

In the LIGO Control and Data acquisition System (CDS) system, as in many digital control systems, the process can be automated by pulling recorded input data and coefficients for each filter. Automated analysis can flag filters with low SNR caused by quantization noise. A flagged filter can then be tested with alternate realizations to determine whether another architecture lowers the quantization noise.

5.8 Quantization-Aware Filter Realization Selection

The generalized measurement in Section 5.7 can be automated to compare filter architectures across a large control system. The scan begins with a Foton filter

file and extracts each populated filter section. Every candidate realization is then given the same input time series, sample rate, initial conditions, and target transfer function. The direct-form and low-noise-biquad candidates use the same second-order-section coefficients, while the state-space candidates are constructed from the equivalent transfer function and, where applicable, transformed using Schur form or `AB09ND`.

Each realization is evaluated once at the precision used by the control system and again at a higher reference precision. If $y_{k,\mathrm{target}}[n]$ and $y_{k,\mathrm{ref}}[n]$ are the two outputs for architecture k, its estimated quantization-noise time series is

$$e_{q,k}[n] = y_{k,\mathrm{target}}[n] - y_{k,\mathrm{ref}}[n]. \tag{5.42}$$

The same spectral-estimation settings are used for every candidate. The resulting noise PSD, $S_{q,k}(f)$, is reduced to the relative noise-power score

$$P_{\mathrm{rel},k} = \frac{\int_0^{f_s/2} S_{q,k}(f)\, df}{\int_0^{f_s/2} S_{\mathrm{in}}(f)\, df}, \tag{5.43}$$

where $S_{\mathrm{in}}(f)$ is the input PSD. Normalizing by the input power allows filters with different signal levels to be compared using the same metric.

Before ranking the candidates, the scan checks that the realization is finite and that the state-space transformations preserve the intended transfer function to numerical precision. Any candidate that produces non-finite samples or fails its transformation checks is excluded. The architecture with the smallest $P_{\mathrm{rel},k}$ among the remaining candidates is recorded as the numerical winner for that filter. This definition makes the ranking reproducible, but it is a screening criterion rather than an automatic deployment decision. If several candidates have similar scores, the quantization-noise spectra must be compared in the frequency band relevant to the control loop. Computational cost and compatibility with the real-time system can then break a near tie. For some filters it is advantageous to add a frequency weighting to the integral in Eq. 5.43 to emphasize the frequency band of interest. The weighting should be chosen to reflect the control system's sensitivity to noise at different frequencies. A similar approach is used in Section 7.2.2 to define a frequency-weighted figure of merit for control filters. This, along with accurately shaped input signals, would allow the scan to identify the best architecture for each filter in a way that is relevant to the control system's performance. Both are left for future work. The current scan uses a uniform weighting and a standard test input to demonstrate the method.

To demonstrate the method on filters from an operating interferometer, the automated scan was run on 94 populated sections from the Caltech 40m prototype interferometer's ETMX suspension filters [99]. Because every filter was driven by the same standard test input rather than an input representative of normal operation, this scan is a proof of concept. A deployment-quality test should replay a typical recorded input for each filter, since its amplitude and spectral content determine the internal states and resulting quantization noise [84, 94]. The target calculations used 64-bit arithmetic, with NumPy's extended-precision `longdouble` type as the reference. The resulting winner counts are given in Table 5.1.

Low-Noise Biquad Form I, the realization used in the LIGO control system [94], is the most frequent individual winner, giving the lowest relative noise power for 44 of the 94 sections. It is therefore a strong general-purpose architecture, especially because it is much less computationally expensive than any of the state-space variants, but it is not uniformly optimal. The state-space variants collectively win 48 sections, led by `AB09ND` followed by Schur form with 35; Direct Form I wins the remaining two. These alternatives can produce less integrated numerical noise when their internal coordinates are better scaled for a particular set of coefficients. The preferred architecture can also depend on frequency: a realization without the smallest full-band $P_{\mathrm{rel},k}$ may still be quietest within the experimentally relevant control band.

Architecture	Filters won
Low Noise Biquad Form 1 (LIGO)	44
State Space + AB09ND + Schur	35
State Space + Schur	5
State Space + AB09ND	4
State Space + Schur + AB09ND	4
Direct Form I	2
Total scanned filter sections	94

Table 5.1: Lowest-noise architectures in a proof-of-concept scan of 94 populated filter sections from the Caltech 40m ETMX suspension Foton file. The rankings compare 64-bit outputs with extended-precision `longdouble` references using Eq. 5.43 and a common standard test input. The LIGO Low-Noise Biquad Form I is the most frequent individual winner, although conditioned state-space realizations are quieter for many filters.

The scan rejects invalid implementations and provides a consistent first ranking. Representative operating inputs and frequency-dependent requirements should de-

termine the final architecture.

5.9 Summary

Finite arithmetic introduces a noise source whose spectrum depends on both numerical precision and the path taken through a filter's internal calculations. Two realizations of the same transfer function are therefore not physically equivalent once they are implemented on a digital system. The analytic noise estimates and the comparison with a higher-precision reference provide complementary ways to evaluate this effect: one identifies how roundoff propagates through the realization, while the other measures the numerical noise produced by a particular filter and input.

A digital-filter design is consequently not complete when its poles and zeros have been chosen. Its realization must also be checked against the expected signal and the noise requirements of the system in which it will operate. The notch-filter measurement in Section 5.6.1 demonstrates why this distinction can matter for weak signals, and the Caltech 40m scan demonstrates that the same test can be applied across an existing control system. Low-Noise Biquad Form I remains a strong practical default, but representative operating data and frequency-dependent requirements are needed to identify the cases in which another realization provides a meaningful advantage. Treating quantization noise as a measurable part of commissioning prevents finite arithmetic from becoming an unintended limit on the physical measurement.

Chapter 6

OPTIMAL CONTROLS

"Between stimulus and response, there is a space. In that space is our power to choose our response. In our response lies our growth and our freedom"

— Unknown, Found in an unidentified library book[1]

Relationship to Published Work

The work presented in this chapter and Chapter 7 was conducted as part of this thesis and originally published in "Robust bilinear-noise-optimal control for gravitational-wave detectors: A mixed LQG and $\mathcal{H}_\infty$ approach" [100]. The paper was organized as a bridge from classical to modern control: it first introduced the LIGO controls-noise problem, formulated the astrophysical objectives, and presented the resulting controller families before giving the state-space and Riccati-equation derivation. Supporting derivations and alternate formulations were placed in appendices.

The thesis separates that material by purpose and reverses the order so that the general synthesis method precedes its interferometer application. This chapter collects the generalized state-space construction, the $\mathcal{H}_2$/linear quadratic Gaussian (LQG) and $\mathcal{H}_\infty$ Riccati methods, and the mixed-sensitivity implementation. The paper's appendices on translating the algebraic notation and on alternate noise arrangements are promoted into the main discussion in Sections 6.6 and 6.7. A detailed derivation connecting state covariance, output variance, and the Lyapunov equation has also been added in Section 6.4.

Chapter 7 then applies this framework to the LIGO Hanford DHARD-yaw loop. It contains the interferometer background, bilinear-noise model, figures of merit, Pareto fronts, and robustness-constrained controller results. The derivation of the BNS range weighting is expanded in Section 7.2.2. The comparison with AI-

[1] Stephen R. Covey recalled finding this passage in a library book, but he did not record the author or title and was unable to locate the source when he later returned. Although the words are now frequently attributed to Viktor Frankl, no evidence has been found that Frankl wrote them. A similar passage in a 1963 article by Rollo May is the most likely inspiration for Covey's recollection.

informed and convex-optimization controller synthesis in Section 7.6 is new to the thesis. This organization allows the reusable control method developed here to lead directly into its experiment-specific demonstration.

6.1 Introduction

The control-theory concepts reviewed in Chapter 4 describe how feedback changes a system's response and how the resulting loop can be evaluated for stability. They do not, by themselves, determine which stabilizing controller should be used. In a precision interferometer, and really all experiments, that choice must balance residual motion, sensing noise, actuation noise, control effort, and robustness over a wide frequency range. The first step in optimal control is therefore to translate the physical purpose of the loop into a precise mathematical objective.

Optimal-control methods make this translation explicit. The plant dynamics, disturbances, sensors, actuators, and quantities to be minimized are assembled into a single system model. Weighting filters assign physical meaning and relative priority to the performance outputs, and the controller is chosen to minimize the resulting cost while respecting the system dynamics and stability requirements. An LQG controller is used to minimize integrated noise, while an $\mathcal{H}_\infty$ constraint limits the largest closed-loop response. Their combination produces a mixed-sensitivity controller that balances average noise performance against robustness.

6.2 Noise in Feedback Systems

Feedback redistributes noise rather than eliminating it. A disturbance acting on the plant, noise added by the sensor, and noise introduced through the actuator enter the loop at different locations and therefore propagate through different closed-loop transfer functions. Increasing the loop gain can suppress one contribution while increasing another or requiring more control effort. A controller cannot therefore be optimized from the nominal plant response alone.

The optimization model must identify each independent noise source, where it enters, which signals are available to the controller, and which outputs represent experimental performance. Colored disturbances can be represented by shaping filters driven by white-noise inputs, while additional output filters convert physical quantities into the desired figures of merit (FOMs). The $\mathcal{H}_2$ norm then measures the integrated mean-square response of those outputs, and the $\mathcal{H}_\infty$ norm bounds the largest response over frequency. The chosen inputs, outputs, and weighting filters define what "optimal" means for the physical system.

6.3 Riccati Equation Approach to Mixed-Sensitivity Controls

The mixed-sensitivity method used here is solved through algebraic Riccati equations. Classical loop quantities remain useful for interpreting the resulting controller, but the numerical synthesis is performed on a generalized state-space model. The $\mathcal{H}_2$ portion supplies the noise-minimizing LQG solution, while the $\mathcal{H}_\infty$ condition restricts the worst-case gain of a designated input-output path.

The development begins by assembling the interconnected optical, mechanical, sensing, and actuation models into a monolithic state-space realization. It then relates the $\mathcal{H}_2$ norm to output noise through the Lyapunov equation and constructs the corresponding LQG controller from solutions to the Riccati equations. The mathematical assumptions required by the solvers are stated explicitly, along with system augmentations that enforce them without changing the relevant input-output response. The $\mathcal{H}_\infty$ bound is introduced last and combined with the LQG objective to form the mixed-sensitivity problem. This ordering connects each algebraic step to the physical model and makes the synthesis procedure reproducible.

6.3.1 State-Space System Description

To define an LQG problem, the "System Model" component of Fig. 7.1 must be assembled into a monolithic state-space. All of the component subsystems are individually modeled or fit from data and converted to state-space form. They are then connected according to the diagram using composition rules, and the entire system, excluding the controller, is put into the form:

$$\dot{\vec{x}} = \mathbf{A}\vec{x} + \mathbf{B}_1\vec{u}_i + \mathbf{B}_2\vec{u}_\mathrm{c} \tag{6.1}$$

$$\vec{y}_\mathrm{FOM} = \mathbf{C}_1\vec{x} + \mathbf{D}_{12}\vec{u}_\mathrm{c} \tag{6.2}$$

$$\vec{y}_\mathrm{meas} = \mathbf{C}_2\vec{x} + \mathbf{D}_{21}\vec{u}_i + \mathbf{D}_{22}\vec{u}_\mathrm{c} \tag{6.3}$$

where $\mathbf{A}$ is the $n \times n$ state matrix, $\mathbf{B}_j$ is the $n \times m_j$ input matrix, $\mathbf{C}_i$ is the $p_i \times n$ sensing matrix, $\mathbf{D}_{ij}$ is the $p_i \times m_j$ feed-through matrix, $\vec{x}$ is the state vector, $\vec{u}_\mathrm{c}$ is the control input, $\vec{u}_i$ is the vector of inputs for simulating noise in the system, $\vec{y}_\mathrm{FOM}$ is the vector of FOM outputs, and $\vec{y}_\mathrm{meas}$ is the measured output. Using $s = 2\pi j f$,

the system transfer function matrix, $\mathbf{T}$, is given by

$$\mathbf{T}(f) = \left[\begin{array}{c|cc} \mathbf{A} & \mathbf{B}_1 & \mathbf{B}_2 \\ \hline \mathbf{C}_1 & 0 & \mathbf{D}_{12} \\ \mathbf{C}_2 & \mathbf{D}_{21} & \mathbf{D}_{22} \end{array}\right] \tag{6.4}$$

$$\equiv \begin{bmatrix} \mathbf{C}_1 \\ \mathbf{C}_2 \end{bmatrix} (s - \mathbf{A})^{-1} \begin{bmatrix} \mathbf{B}_1 & \mathbf{B}_2 \end{bmatrix} + \begin{bmatrix} 0 & \mathbf{D}_{12} \\ \mathbf{D}_{21} & \mathbf{D}_{22} \end{bmatrix}. \tag{6.5}$$

It is important to note that while the system described by $\mathbf{T}$ is often called the "plant" in controls literature, in this chapter it will continue to be called the system model so as to differentiate it from the plant in Fig. 7.1. Given a $\mathbf{T}$ in the proper form, it is possible to calculate a controller, K, that minimizes the $\mathcal{H}_2$ norm of the combined system shown in Fig. 6.1 [101, 102].

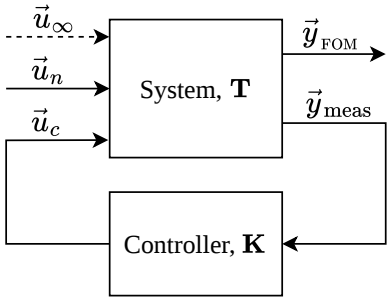


Figure 6.1: A block diagram depicting the connection of the controller to the plant and abstracting away the internal structure of other block diagrams, such as Fig. 7.1. The transfer function from $\vec{u}_n$ to $\vec{y}_{\text{FOM}}$ is $\mathbf{T}_2$, discussed in Section 6.3.1. The transfer function from $\vec{u}_\infty$ to $\vec{y}_{\text{FOM}}$ is referred to as $\mathbf{T}_\infty$. This abstraction is depicted to clarify that transformations which preserve the T_2 or T_∞ response do not affect the optimal closed-loop response, but may influence the optimal $\mathbf{K}$, requiring it to be mapped through inverse transformations. Adapted from [100].

In the following, the inputs and outputs of the $\mathbf{T}_2$ and $\mathbf{T}_\infty$ components of this system will need to be distinguished. Those can be distinguished by segmenting the noise inputs into the white noise and infinity contributions, using

$$\mathbf{B}_1 = \begin{bmatrix} \mathbf{B}_{1n} & \mathbf{B}_{1\infty} \end{bmatrix}, \tag{6.6}$$

$$\mathbf{D}_{21} = \begin{bmatrix} \mathbf{D}_{21w} & \mathbf{D}_{21\infty} \end{bmatrix}, \tag{6.7}$$

and

$$\vec{u}_i = \begin{bmatrix} \vec{u}_n \\ u_\infty \end{bmatrix}. \tag{6.8}$$

6.3.2 Calculating the $\mathcal{H}_2$ Norm

With the state-space representation of the system, the total RMS noise for any controller can be calculated. Along with evaluating the transfer function equations of Eq. 6.5 to construct the frequency-resolved noise budget, matrix mathematics from numerical Lyapunov solvers can be used to efficiently calculate the total RMS noise without numerically evaluating integrals[2]. This technique is used to construct the plots of Figs. 7.4 and 7.7. The same point of view is also used to construct application-specific FOMs: once a quantity such as the BNS range loss has been rewritten as an integrated output spectrum, as in Section 7.2.2, it can be evaluated by the same $\mathcal{H}_2$ and Lyapunov approach.

The $\mathcal{H}_2$ norm of the system is a measure of the RMS at $\vec{y}_{\mathrm{FOM}}$ output from the white noise input $\vec{u}_n$. For a continuous system with a transfer function, $\mathbf{T}_2(s)$, the $\mathcal{H}_2$ norm is given by

$$\|\mathbf{T}_2\|_{\mathcal{H}_2} = \sqrt{\frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{trace}\left[\mathbf{T}_2^{\dagger}(j\omega)\,\mathbf{T}_2(j\omega)\right]\,d\omega}. \tag{6.9}$$

The $\mathcal{H}_2$ norm is calculated by solving the Lyapunov equation given by

$$\mathbf{A}^T L_o + L_o \mathbf{A} + \mathbf{C}^T\mathbf{C} = 0, \tag{6.10}$$

$$\mathbf{A} L_{\mathrm{c}} + L_{\mathrm{c}}\mathbf{A}^T + \mathbf{B}\mathbf{B}^T = 0, \tag{6.11}$$

where L_o is the observability Gramian of ($\mathbf{C}$, $\mathbf{A}$), and L_{c} is the controllability Gramian of ($\mathbf{A}$, $\mathbf{B}$). Both are determined by numerically solving the Lyapunov equations. In this expression, the matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ are the matrices of the $\mathbf{T}_2$ system whose $\mathcal{H}_2$ norm is being measured (usually with a controller solved and attached). The $\mathcal{H}_2$ norm can then be calculated by

$$\|\mathbf{T}_2\|_{\mathcal{H}_2} = \sqrt{\mathrm{trace}\left[\mathbf{B}^T L_o \mathbf{B}\right]} = \sqrt{\mathrm{trace}\left[\mathbf{C} L_{\mathrm{c}} \mathbf{C}^T\right]}. \tag{6.12}$$

Note that these Lyapunov equation methods apply only when $\mathbf{A}$ is stable, but that condition is guaranteed for the closed-loop systems returned by LQG-based solvers. Allowing $\mathbf{A}$ to be unstable changes the interpretation to an $\mathcal{L}_2$ norm calculation, and can be solved using the SLICOT function `AB13BD` [103]. It has been observed that standard Lyapunov equation solvers sometimes fail to produce accurate noise estimates due to the dynamical range of the transfer functions involved. More

[2] This is well known in the control systems community, but less well known in the GW interferometer control commissioning community.

advanced Lyapunov solvers that compute the "square root" of the noise covariance matrices using Hammarling's method and return Cholesky decompositions are more numerically reliable [104, 105].

The use of Lyapunov equations to compute the $\mathcal{H}_2$ norm is foundational to LQG theory and can be used to derive the following Riccati equations for optimal control [106]. While this is widely known in the controls community and textbooks, it is included here for a physics audience.

6.3.3 Optimal LQG Controllers

Some of the major points of the LQG numerical solution process are outlined here, as they inform the mixed-sensitivity method to follow.

The optimal LQG controller, $\mathbf{K}_{\mathrm{LQG}}$, is found by solving two continuous algebraic Riccati equations,

$$\mathbf{A}^T\overline{X} + \overline{X}\mathbf{A} - \left(\overline{X}\mathbf{B}_2 + \mathbf{C}_1^T\mathbf{D}_{12}\right)\left(\mathbf{D}_{12}^T\mathbf{D}_{12}\right)^{-1}\left(\mathbf{B}_2^T\overline{X} + \mathbf{D}_{12}^T\mathbf{C}_1\right) + \mathbf{C}_1^T\mathbf{C}_1 = 0, \quad (6.13)$$

and

$$\mathbf{A}\overline{Y} + \overline{Y}\mathbf{A}^T - \left(\overline{Y}\mathbf{C}_2^T + \mathbf{B}_1\mathbf{D}_{21}^T\right)\left(\mathbf{D}_{21}\mathbf{D}_{21}^T\right)^{-1}\left(\mathbf{C}_2\overline{Y} + \mathbf{D}_{21}\mathbf{B}_1^T\right) + \mathbf{B}_1\mathbf{B}_1^T = 0. \quad (6.14)$$

Each of these Riccati equations corresponds to the Lyapunov equations of the prior section. The $\overline{Y}$ is the observability Gramian for the optimally-controlled full-state feedback system, where the estimator/controller is unknown but assumed to be a Kalman filter. It is also a positive definite symmetric matrix. As the sensing noise becomes smaller, the Kalman gain increases and $\overline{Y}$ becomes increasingly positive semidefinite, eventually unable to be resolved numerically. Analogously, $\overline{X}$ is a positive (semi)definite matrix and the controllability Gramian of the closed-loop system. The stabilizing solutions, defined below, $\overline{X} \geq 0$ and $\overline{Y} \geq 0$, are then used to calculate the Kalman gain, K_s, and the regulator feedback, F_s, using:

$$F_s \equiv -\left(\mathbf{D}_{12}^T\mathbf{D}_{12}\right)^{-1}\left(\mathbf{B}_2^T\overline{X} + \mathbf{D}_{12}^T\mathbf{C}_1\right) \quad (6.15)$$

and

$$K_s \equiv -\left(\mathbf{D}_{21}\mathbf{D}_{21}^T\right)^{-1}\left(\mathbf{C}_2\overline{Y} + \mathbf{D}_{21}\mathbf{B}_1^T\right). \quad (6.16)$$

The strictly proper controller that solves the $\mathcal{H}_2$-control problem is given by

$$\overline{\mathbf{K}} = \left[\begin{array}{c|c} \mathbf{A} + \mathbf{B_2}F_s + K_s^T\mathbf{C_2} & K_s^T \\ \hline -F_s & 0 \end{array}\right]. \quad (6.17)$$

Note that the above equations do not use $\mathbf{D}_{22}$, and assume it to be zero. One can remove or add the D_{22} using a feedforward term on $\mathbf{L}$, which can be reinterpreted (and numerically implemented) as a feedback term on $\overline{\mathbf{K}}$. This amounts to a linear canonical transformation to find the final optimal LQG controller $\mathbf{K}_{\mathrm{LQG}}$ of the as-given system:

$$\mathbf{K}_{\mathrm{LQG}} = \overline{\mathbf{K}}(\mathbf{I}_{p_2} + \mathbf{D}_{22}\overline{\mathbf{K}})^{-1}, \tag{6.18}$$

where $\mathbf{I}$ is the identity matrix of dimension p_2.

The existence and numerical solvability of these equations depend on many results in control theory and mathematical analysis. Two that are relevant to the mixed-sensitivity approach are emphasized here.

The first is the separability of the solutions. The controller $\overline{\mathbf{K}}$ is computed from two independent sub-problems. The two independent problems are the linear-quadratic observer (a Kalman filter) and the linear-quadratic regulator. Each problem has stability guarantees for the limiting case of state feedback or state estimation, yet these combine to form a single control system that is itself stable and that stabilizes the system under feedback. This separability is a major reason that the solution is non-iterative and can be proven to be globally optimal.

The second key result emphasized here is the development of deflating subspace methods to solve algebraic Riccati equations [107, 108]. In short, these methods transform either of the above equations into a single matrix representing a particular system of equations. The matrix is transformed into a Schur form, upper triangular with its eigenvalues along the diagonal. During the conversion to Schur form, a modified QR or QZ algorithm reorders the stable eigenvalues first along the diagonal. Once in this form, the system of equations can be solved using a column of the orthogonal transformation matrix, the deflating subspace. In the original formulation of Laub, the matrix is "Hamiltonian", a form that ensures that half of the eigenvalues are stable. That form can be generalized [109] to solve directly for the feedback control as well. The choice of eigenvalues represents the poles of the closed-loop system, ensuring stability under optimal control. Under the conditions of the following section, the Gramians $\overline{X}$ and $\overline{Y}$ are ensured to exist and to be positive definite, which in turn ensures that the controller, as an independent system, is itself stable.

These two are mentioned because they are essential for the solution of the mixed-sensitivity problem and for further improving numerical robustness for high dynamic

range gravitational wave system models. For example, the `ordqz` algorithm that is essential to the Scipy implementation of the deflating subspace method, on which these implementations are based, greatly benefits from matrix rescaling [110, 111]. This has motivated custom (re-)implementations of these solvers and testing against quad-precision linear algebra implementations, where the physical dynamic range is a substantially smaller fraction of the floating-point epsilon. This is left for future work, and several libraries exist in the Julia language that may make such a future effort possible. In testing, iterative refinement has not been found to improve upon the reliability of a deflating subspace solver [112].

6.3.4 Mathematical Assumptions and Their Physical Interpretation

In order to calculate the optimal controllers for the system given in Eqs. 6.1 and 6.5, the following assumptions must be met to ensure that the problem is well-posed. These are presented mathematically here, followed by the physical requirements and manipulations of the system under consideration used to generically enforce them in the list that follows, as well as Section 6.3.5.

1. The inputs, $\vec{u}_n$, must be unit white noise.
2. $(\mathbf{A}, \mathbf{B}_2)$ is stabilizable and $(\mathbf{C}_2, \mathbf{A})$ is detectable. This assumption is necessary for an internally stable closed-loop system. More restrictive assumptions are given in [101].
3. $\mathbf{D}_{12}$ and $\mathbf{D}_{21}$ both have full column rank.
4. $\mathbf{D}_{12}^T\mathbf{D}_{12}$ is invertible and numerically non-singular.
5. $\mathbf{D}_{21}\mathbf{D}_{21}^T$ is invertible and numerically non-singular.
6. $\mathbf{D}_{11} = 0$. Note that $\mathbf{D}_{22}$ is often also assumed to be 0 for separable solutions, but does not have to be. If $\mathbf{D}_{22} \neq 0$, the optimal controller can be computed as though it is zero, and then transformed as shown in Eq. 6.18.
7. $\text{rank}\begin{bmatrix} \mathbf{A} - j\omega\mathbf{I}_n & \mathbf{B}_2 \\ \mathbf{C}_1 & \mathbf{D}_{12} \end{bmatrix} = n + m_2, \quad \forall\omega \in \mathbb{R}$, where n is the order of $\mathbf{A}$, m_2 is the dimension of $\vec{u}_c$, and $\mathbf{I}$ is the identity matrix also of order n.
8. $\text{rank}\begin{bmatrix} \mathbf{A} - j\omega\mathbf{I}_n & \mathbf{B}_1 \\ \mathbf{C}_2 & \mathbf{D}_{21} \end{bmatrix} = n + p_2, \quad \forall\omega \in \mathbb{R}$, where p_2 is the dimension of $\vec{y}_{\text{meas}}$.

These formal assumptions restrict the types of systems that can be solved for using this method. The reasoning behind each assumption, as well as a translation into more useful language, is presented as follows:

item 1 This assumption requires that the noise entering the system is Gaussian white noise. This requirement is ensured for systems that require non-white noise by using noise-shaping filters, E and M, on the two noise inputs, which unwhiten the noise to admit a physical source. Note that at least one of the two noise shapes must have a non-singular D matrix, indicating that the noise becomes asymptotically white at sufficiently high frequency. This is typically the measurement noise, M. This is required formally to enforce item 3, but is physically necessary to make a well-posed optimization tradeoff between low-frequency system noise and measurement noise.

item 2 This requirement is required for the control and estimation sub-problems of the separation principle to be individually numerically well-posed. Physically, an observable but unstabilizable state contributes to RMS but cannot be controlled to suppress its contribution. Thus, it is either a fundamental problem from its real noise contribution or should simply be ignored. If ignored, model reduction should be used. Because of well-known issues in choosing tolerances, structural algorithms to produce minimal realizations [113] are not preferred. Instead, Grammian-based truncation algorithms both reduce systems and improve other numerical properties. Singular perturbation approximation-based model reduction is employed [98] using `AB09ND` from SLICOT [97].

items 3 and 5 These requirements are necessary to make a well-posed minimization problem where the control system couples white noise into its states through $\mathbf{D}_{21}$ and then that noise is coupled to the output through $\mathbf{D}_{12}$. Without those requirements, there is excessive freedom to choose how the noise is rolled off. Along with requirements on the E and M shaping filters, this also requires that the "plant" component has a non-singular D matrix, and thus cannot be strictly proper. Additionally, at least one of the FOM filters must have a non-singular D matrix. Accommodating more general strictly proper plant systems is non-trivial, but is generically possible with the augmentation methods presented in the following section.

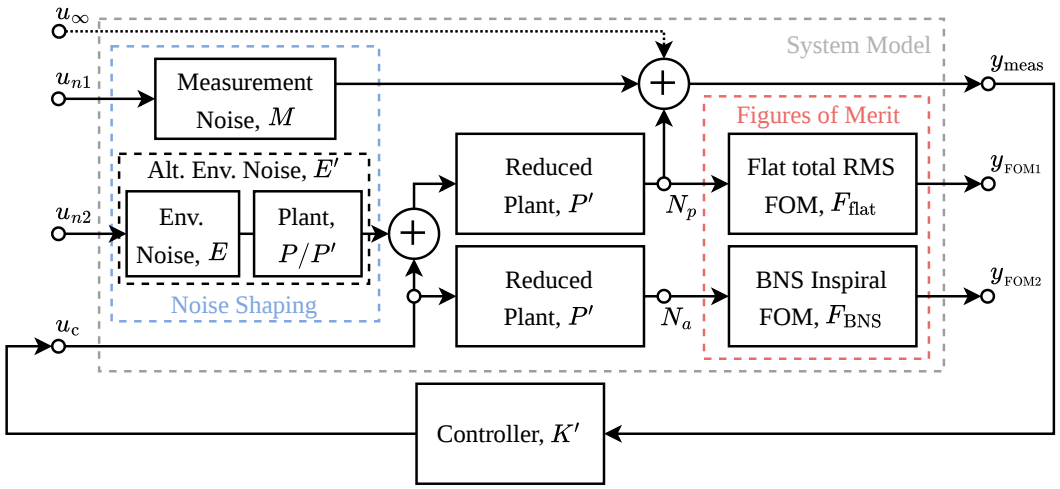


Figure 6.2: The layout of the augmented and modified plant system with a controller attached. This alteration of Fig. 7.1 shows the modifications that preserve **T** while better ensuring the reliability conditions of Section 6.3.4. Adapted from [100].

item 6 The $\mathbf{D}_{11}$ condition is necessary to prevent an infinite noise power from contaminating the problem. The $\mathbf{D}_{22}$ conditions are a mathematical convenience in the problem statement but can be compensated for by advanced problem restructuring [114].

items 7 and 8 If either of these rank conditions is insufficiently met, then the open-loop gain may have a zero on the imaginary line, and the existence of solutions is not guaranteed, or at least substantially more nuanced. It can often be solved anyway with the numerical reliability of generalized eigenvalue and descriptor-system techniques. See remark 14.1 of [115].

6.3.5 System Augmentation

In order to accommodate the assumptions of Section 6.3.4, some augmentations and manipulations are made to make the solver more applicable to general interferometer systems. One example of such augmentations, which has already been discussed, is the inclusion of noise shaping and the FOMs in Section 7.2.2. Further manipulations change the system laid out in Fig. 7.1 into a new modified system of Fig. 6.2.

To make the new system equivalent to the original, its transfer matrix, $\mathbf{T}(f)$ from the $\vec{u}_n$ inputs to the $\vec{y}_{\text{FOM}}$ outputs must be preserved. The control inputs and outputs can also be modified during this process, but only in ways that can be stably inverted. For a SISO plant and control systems, there is a general method for meeting the assumptions necessary for the solver, as long as the noise shaping filters and FOMs

meet the physical conditions outlined at the end of Section 6.3.4.

The following manipulations ensure that item 4 is met. To do so, much of the plant must be included in the environmental noise block. Most physical plants, such as optical systems or suspensions, have some roll-off at high frequency. This makes them strictly proper systems and implies that the plant subsystem's state-space representation has $\mathbf{D} = 0$. By moving most of the plant to the environmental noise block, it can be assured that the remainder has a balanced number of poles and zeros, so that $\mathbf{D}_{12} \neq 0$. If the plant is stable, then the only part of the plant that is left in the plant block is the delay or other non-minimum-phase zeros, which can not be moved without changing the physical interpretation. Unstable poles must also remain in the plant section to preserve the problem formulation. The requirement is that the reformulated plant system, P', has $\lim_{f\to\infty} P'(f) > 0$ while containing the same non-minimum phase zeros and unstable poles as P.

Under this redefinition, P' becomes the new plant subsystem, K' is the new calculated controller, and $G' = K'P'$ is the new open-loop gain. Because the $\mathbf{T}_2$ and $\mathbf{T}_\infty$ transfer functions are preserved, the open-loop gain must be the same as the original formulation, thus $G = G'$. After computing K', the controller can be compensated using $K \equiv K'P'/P$ to find the controller of the original formulation. Note that this representation may not be a proper transfer function, and thus may not admit a state-space representation. A simple reason the aforementioned solvers fail is that their outputs are in state-space form, which may not be able to represent optimal controllers, even if the process can properly pose the optimization problem. For SISO systems, the final controller may be computed from K' using a zero-pole-gain (ZPK) form or using a descriptor state-space representation.

After transformation, Eq. 7.11 becomes

$$S_\mathrm{p} = \left|\frac{1}{1-G'}\right|^2 S'_\mathrm{env} + \left|\frac{G'}{1-G'}\right|^2 S_\mathrm{meas}, \tag{6.19}$$

where

$$S'_\mathrm{env} = |P/P'|^2\, S_\mathrm{env}, \tag{6.20}$$

which is equivalent to the original formulation.

A useful convention for simple plant systems, such as in the alignment sensing control (ASC) example considered here, is to make $|P'| = 1$ (an inner factor), so that it captures delay and stability effects, without any additional shaping. For this

convenient convention, the computed $K' = G$ when the system has no delay, $P' = 1$. This convention always has the property $|K'| = |G|$.

If the plant is unstable, all the unstable poles must remain in the plant. To ensure that item 4 is still met, the stable version of the unstable poles, i.e. the pole swapped to the left-hand plane, must be added to the plant as zeros. Then those same stabilized poles must be added to the environmental noise system as poles. Since there will be the same number of poles and zeros in P, $\mathbf{D}_{22} \neq 0$. When executed correctly, the stable poles added to E' will cancel the zeros added to P'.

These modifications are essential and relatively simple to implement for SISO models. The task of refactoring MIMO systems in an analogous manner is left for future work. It is reiterated that the composition of P' uses only the unstable poles and non-minimum phase zeros of P. Making P' an *inner coprime* factor of P will satisfy the conditions. The factorization described above was implemented using ZPK representations of SISO filters due to the unavailability of inner/outer factorization functions in Python's wrappers of SLICOT (e.g., `SB08CD`), but coprime factorization is anticipated to be effective in extending this formulation for MIMO plants.

In addition to the structural modifications, the system representation is also numerically modified. Given that many individual control systems in LIGO deal with noise over multiple orders of magnitude and models can contain over a hundred states, numerical problems can quickly become a problem for the solvers. To combat this, the system undergoes a matrix balancing stage [116, 117]. This better conditions the problems for generalized eigenvalue solvers. Before solving and after attaching the controller, the system's order was reduced using square-root singular perturbation approximation with preliminary scaling of $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ [98, 118]. These conditioning steps can be essential to solving for complex controllers of gravitational-wave interferometer noise models.

6.3.6 Calculating the $\mathcal{H}_\infty$ Norm

The $\mathcal{H}_\infty$ norm is a measure of the peak gain of a closed-loop system over all frequencies. When used as a reinterpretation of the $\mathbf{T}_2$ transfer function, it can be considered as a measure of a system's robustness against an environment (or adversary) that directs all input noise power to the worst possible frequency. For specific $\mathbf{T}_\infty$ designs, it can be used to compensate specific, known perturbations to $\mathbf{A}$ of the plant system. For gravitational-wave detectors, the stationary noise does not

actively counter the control systems, but has such a large magnitude that it prevents solutions with acceptable stability margins.

The design of the $\mathbf{T}_\infty$ transfer function, using u_∞ to the $\vec{y}_{\text{FOM}}$ outputs, specifically targets the phase margin, just like hand-tuned classical controls designs. The $\mathcal{H}_\infty$ norm represents the maximum possible gain that is achievable by the system given all possible disturbances.

For a continuous system with a transfer function, $\mathbf{T}_\infty(s)$, the $\mathcal{H}_\infty$ norm is most formally defined as

$$\|\mathcal{H}_\infty\| = \sup_\omega \sigma_{\max}\left[\mathbf{T}_\infty(j\omega)\right], \tag{6.21}$$

where sup is the supremum over ω and $\sigma_{\max}$ is the maximum singular value of the matrix $\mathbf{T}(j\omega)$.

A means of efficiently finding this norm is another highly worthwhile result from mathematical controls called the bounded real lemma. Unlike directly solving for the $\mathcal{H}_2$ norm, the $\mathcal{H}_\infty$ norm can be detected by first assuming a bound, γ, where $\|\mathbf{T}_\infty\|_{\mathcal{H}_\infty} < \gamma$. Thus, finding the optimal $\mathcal{H}_\infty$ controller requires a search for the minimum value of γ. The notion to test this comes from finding the zeros of $0 = \mathbf{T}_\infty(f)\mathbf{T}_\infty^\dagger(f) - \gamma^2\mathbf{1}$, which is the cross-spectrum matrix of the system (positive semidefinite), less the γ bound. This equation can be produced from a state-space connected to its time-reversed transpose. Finding the zeros of this equation is equivalent to finding the poles of the system $\gamma^{-2}\mathbf{T}_\infty(f)\mathbf{T}_\infty^\dagger(f)$ that has its output fed back to its input. The A matrix of this feedback system is the following Hamiltonian matrix, $\mathbf{H}$, expressed here using $\mathbf{A}, \mathbf{B}, \mathbf{C}$ of $\mathbf{T}_\infty$,

$$\mathbf{H} \equiv \begin{bmatrix} \mathbf{A} & \gamma^{-2}\mathbf{B}\mathbf{B}^T \\ -\mathbf{C}^T\mathbf{C} & -\mathbf{A}^T \end{bmatrix}. \tag{6.22}$$

It follows that, if there are no eigenvalues of this matrix along the imaginary line, then $\|\mathcal{H}_\infty\| < \gamma$ for the $\mathbf{T}_\infty$ system. Thus, it is possible to choose a value for γ, construct the Hamiltonian matrix, check the spectrum of the solution, then adjust the gamma value and repeat until the lowest gamma value is found within the desired precision. The above algorithm has been highly refined, and it can be directly computed using the SLICOT routine `AB13DD` [119].

This result is presented for a physics audience to indicate that $\mathcal{H}_\infty$ bounds can be clearly and concisely stated and can be related to a physically meaningful feedback relation. The mixed-sensitivity approach to follow was originally derived by

applying Eq. 6.22 to a noise minimization problem of Eq. 6.12 using Lagrangian constraints.

Finally, the use of singular values in the statement of the norm, Eq. 6.21, may not be intuitive. For the single input of u_∞, there can be only one singular value of this matrix, which must be the 2-norm of the vector. This leads to the simple γ-constraint of Eq. 7.50 without further concern about the complexity of solving for singular values.

6.3.7 Optimal $\mathcal{H}_\infty$ Controllers

Here, the solution method for $\mathcal{H}_\infty$-optimal control is presented to relate it to the mixed-sensitivity approach and as background for a physics audience. The primary takeaway is in how $\mathcal{H}_\infty$ utilizes the parameter γ as a bound on its norm, which is also utilized in mixed-sensitivity control.

Similar to the $\mathcal{H}_2$ controller, the $\mathcal{H}_\infty$ controller is also found at the minimum of its respective norm. The $\mathcal{H}_\infty$ norm is a measure of the worst-case scenario for the control system as the maximum gain from an input disturbance to the output over all frequencies. In order to calculate an optimal controller, $\mathbf{K}_{\infty(\text{sub})}$, for the system given in Eqs. 6.1 and 6.3, it must follow the assumptions of Section 6.3.3.[3]

Unlike solving for the optimal $\mathcal{H}_2$ controller, the $\mathcal{H}_\infty$ controller, like its norm, has one free parameter, γ, where $\|\mathcal{H}_\infty\| < \gamma$. Thus, finding the optimal $\mathcal{H}_\infty$ controller requires a search for the minimum value of γ.

A sub-optimal $\mathcal{H}_\infty$ controller is found by solving two continuous algebraic Riccati equations,

$$\mathbf{A}^T\overline{X}_\infty + \overline{X}_\infty\mathbf{A} - \overline{X}_\infty\left(\gamma^{-2}\mathbf{B}_1\mathbf{B}_1^T - \mathbf{B}_2\mathbf{B}_2^T\right)\overline{X}_\infty + \mathbf{C}_1^T\mathbf{C}_1 = 0 \tag{6.23}$$

and

$$\mathbf{A}\overline{Y}_\infty + \overline{Y}_\infty\mathbf{A}^T - \overline{Y}_\infty\left(\gamma^{-2}\mathbf{C}_1^T\mathbf{C}_1 - \mathbf{C}_2^T\mathbf{C}_2\right)\overline{Y}_\infty + \mathbf{B}_1\mathbf{B}_1^T = 0, \tag{6.24}$$

where $\overline{X}_\infty \geq 0$ and $\overline{Y}_\infty \geq 0$ must be stabilizing solutions. They are guaranteed to exist as $\gamma \to \infty$, but will fail to exist at some minimum $\gamma_{\min}$. The resulting admissible controller is

$$\mathbf{K}_{\infty(\text{sub})} = \left[\begin{array}{c|c} \mathbf{A}_{K\infty} & Z_\infty\overline{Y}_\infty\mathbf{C}_2^T \\ \hline -\mathbf{B}_2^T\overline{X}_\infty & 0 \end{array}\right], \tag{6.25}$$

[3] The solution presented here, adapted from [101], is less general than the solution to the LQG problem, given in Section 6.3.3. The Glover-Doyle solution is more general, although it goes beyond the needs of most $\mathcal{H}_\infty$ controllers [120, 121].

where

$$\mathbf{A}_{K\infty} = \mathbf{A} + (\gamma^{-2}\mathbf{B}_1\mathbf{B}_1^T - \mathbf{B}_2\mathbf{B}_2^T)\overline{X}_\infty \\ -Z_\infty\overline{Y}_\infty\mathbf{C}_2^T\mathbf{C}_2 \tag{6.26}$$

and $Z_\infty = (\mathbf{I}_n - \gamma^{-2}\overline{Y}_\infty^T\overline{X}_\infty)^{-1}$. The optimal $\mathcal{H}_\infty$ controller, $\mathbf{K}_\infty$, is the unique controller where γ is at its minimum possible value, $\gamma_{\min}$ for these equations to produce a closed-loop-stable system. The failure to produce a solution below this minimum can happen in several ways [122].

One can inspect the form of Eq. 6.24 and notice that as $\gamma \to \infty$, it recovers the equations and solution of the LQG problem. More generally, for the set of sub-optimal controllers $\gamma_{\text{opt}} < \gamma < \infty$, this method of $\mathcal{H}_\infty$ solution calculates the "central" controller that obeys the γ bound while minimizing a $\mathbf{T}_2 = \mathbf{T}_\infty$ problem. Thus, for mixed-sensitivity problems for which $\mathbf{T}_2$ should be the same as $\mathbf{T}_\infty$, a simple computation of a sub-optimal controller also solves the mixed-sensitivity problem [101].

For the phase-margin constraint and construction of $\mathbf{T}_\infty$ presented here, the sub-optimal central controllers of the associated $\mathcal{H}_\infty$-optimal control problem lack all information about the environmental and sensing noise weights. They are thus not useful for noise minimization. The following shows how to find noise-optimal solutions within the set of sub-optimal controllers.

6.4 Lyapunov Noise Derivation

To measure the RMS noise in a system, the Lyapunov equation can be used. A specific form of the Lyapunov equation is designed to yield the RMS noise at the output of a system. This derivation is important for understanding how to design controllers that minimize the RMS noise at the output.

The state in the state-space representation, given in Eqs. 4.9 and 4.10, is denoted by $\vec{x}(t)$. It is useful to begin with a scalar example using the same symbol and then extend the notation to the full vector case.

6.4.1 Relation Between Covariance and RMS

To build intuition, first consider a scalar random variable x. The covariance of x with itself is simply its variance, which is defined as

$$\text{Cov}[x] = \langle (x - \langle x \rangle)\,(x - \langle x \rangle) \rangle\,. \tag{6.27}$$

This quantity measures the mean-square fluctuation of x about its mean value.

Expanding the product inside the expectation gives

$$\mathrm{Cov}[x] = \left\langle x^2 - x\langle x\rangle - \langle x\rangle x + \langle x\rangle^2 \right\rangle . \tag{6.28}$$

Using the linearity of expectation, and noting that $\langle x\rangle$ is a constant with respect to the averaging operation, this becomes

$$\mathrm{Cov}[x] = \langle x^2\rangle - \langle x\langle x\rangle\rangle - \langle\langle x\rangle x\rangle + \left\langle \langle x\rangle^2 \right\rangle \tag{6.29}$$

$$= \langle x^2\rangle - \langle x\rangle\langle x\rangle - \langle x\rangle\langle x\rangle + \langle x\rangle^2. \tag{6.30}$$

Combining like terms yields

$$\mathrm{Cov}[x] = \langle x^2\rangle - 2\langle x\rangle^2 + \langle x\rangle^2 \tag{6.31}$$

$$= \langle x^2\rangle - \langle x\rangle^2. \tag{6.32}$$

If the expectation value of x is defined as the mean, written as

$$\langle x\rangle = \mu, \tag{6.33}$$

then the covariance may be written as

$$\mathrm{Cov}[x] = \langle x^2\rangle - \mu^2. \tag{6.34}$$

Now define the root-mean-square value through

$$\mathrm{rms}^2 = \langle x^2\rangle. \tag{6.35}$$

Substituting this into the previous result gives

$$\mathrm{Cov}[x] = \mathrm{rms}^2 - \mu^2. \tag{6.36}$$

The previous equation shows that the covariance and the squared RMS are equal only when the mean vanishes. This is a common scenario in signal processing and physics and especially control theory. Having a DC offset in the signal is often undesirable, and therefore it is common to design systems that have zero-mean outputs. Even when a DC offset is present, it can be removed by simply subtracting the mean from the signal, which is a common preprocessing step in data analysis and control design.

In the important special case of a zero-mean variable, $\mu = 0$, one finds

$$\mathrm{Cov}[x] = \langle x^2\rangle = \mathrm{rms}^2, \tag{6.37}$$

and therefore

$$\sqrt{\text{Cov}[x]} = \text{rms}. \tag{6.38}$$

This scalar result is useful because it provides intuition for the corresponding matrix expressions that arise when dealing with the vector-valued state $\vec{x}(t)$ in a state-space system.

6.4.2 Covariance of the Output of a State-Space System

Consider the state-space system, given in Eqs. 4.9 and 4.10, which is repeated here for convenience:

$$\dot{\vec{x}}(t) = \mathbf{A}\vec{x}(t) + \mathbf{B}\vec{u}(t), \tag{6.39}$$

$$\vec{y}(t) = \mathbf{C}\vec{x}(t) + \mathbf{D}\vec{u}(t), \tag{6.40}$$

where $\vec{x}(t)$ is the n-dimensional state, $\vec{u}(t)$ is the m-dimensional input, and $\vec{y}(t)$ is the p-dimensional output. The covariance of the output, $\vec{y}$, is defined by

$$\text{Cov}[\vec{y}] = \langle \vec{y}\vec{y}^T \rangle - \langle \vec{y} \rangle \langle \vec{y} \rangle^T. \tag{6.41}$$

Substituting Eq. 6.40 gives

$$\text{Cov}[\vec{y}] = \left\langle (\mathbf{C}\vec{x} + \mathbf{D}\vec{u}) (\mathbf{C}\vec{x} + \mathbf{D}\vec{u})^T \right\rangle - \langle \mathbf{C}\vec{x} + \mathbf{D}\vec{u} \rangle \langle \mathbf{C}\vec{x} + \mathbf{D}\vec{u} \rangle^T. \tag{6.42}$$

Now assume that the output is zero mean, so that $\langle \vec{y} \rangle = 0$. Under this assumption,

$$\text{Cov}[\vec{y}] = \left\langle (\mathbf{C}\vec{x} + \mathbf{D}\vec{u}) (\mathbf{C}\vec{x} + \mathbf{D}\vec{u})^T \right\rangle. \tag{6.43}$$

Expanding the product gives

$$\text{Cov}[\vec{y}] = \left\langle \mathbf{C}\vec{x}\vec{x}^T\mathbf{C}^T + \mathbf{C}\vec{x}\vec{u}^T\mathbf{D}^T + \mathbf{D}\vec{u}\vec{x}^T\mathbf{C}^T + \mathbf{D}\vec{u}\vec{u}^T\mathbf{D}^T \right\rangle. \tag{6.44}$$

If there is no direct feedthrough, so that $\mathbf{D} = 0$, this expression can be simplified. Since the goal of this derivation is to come up with an equation to calculate the RMS noise at the output, it is reasonable to assume that $\mathbf{D} = 0$. When $\mathbf{D} \neq 0$, the frequency response of the system does not tend towards zero at high frequencies, and therefore the RMS noise at the output is not well defined. In that case, the output noise power would be infinite, and the RMS noise would be infinite as well. Since the solutions to $\mathcal{H}_2$ optimal control problems are closed loop stable, their frequency response tends towards zero at high frequencies, and therefore they have

well-defined RMS noise at the output. The assumption that $\mathbf{D} = 0$ is perfectly acceptable for the subset of systems that are of interest here.

In the $\mathbf{D} = 0$ case, the output covariance reduces to

$$\mathrm{Cov}[\vec{y}] = \left\langle \mathbf{C}\vec{x}\vec{x}^T\mathbf{C}^T \right\rangle . \tag{6.45}$$

Because $\mathbf{C}$ is deterministic, it can be taken outside the expectation to give

$$\mathrm{Cov}[\vec{y}] = \mathbf{C}\langle\vec{x}\vec{x}^T\rangle\mathbf{C}^T. \tag{6.46}$$

For a zero-mean state, $\langle\vec{x}\vec{x}^T\rangle = \mathrm{Cov}[\vec{x}]$, as extracted from Eq. 6.37, and therefore

$$\mathrm{Cov}[\vec{y}] = \mathbf{C}\,\mathrm{Cov}[\vec{x}]\,\mathbf{C}^T. \tag{6.47}$$

This result shows that the output covariance is obtained by mapping the state covariance, $\mathrm{Cov}[\vec{x}]$, through the output matrix, $\mathbf{C}$. Therefore, to understand how noise propagates to the output, it is necessary to understand how the state covariance evolves in time.

6.4.3 State Vector Covariance Dynamics

The next goal is to determine how the state covariance evolves in time. Define the state covariance matrix, $\mathbf{X}(t)$, as

$$\mathbf{X}(t) \equiv \mathrm{Cov}[\vec{x}(t)] = \langle\vec{x}(t)\vec{x}^T(t)\rangle - \langle\vec{x}(t)\rangle\langle\vec{x}(t)\rangle^T. \tag{6.48}$$

Continuing with the assumption from Eq. 6.47 that the state has zero mean, such that $\langle\vec{x}(t)\rangle = 0$, this reduces to

$$\mathbf{X}(t) = \langle\vec{x}(t)\vec{x}^T(t)\rangle. \tag{6.49}$$

Taking the time derivative gives

$$\dot{\mathbf{X}}(t) = \frac{d}{dt}\langle\vec{x}\vec{x}^T\rangle \tag{6.50}$$

$$= \left\langle\dot{\vec{x}}\vec{x}^T\right\rangle + \left\langle\vec{x}\dot{\vec{x}}^T\right\rangle, \tag{6.51}$$

where the good old fashioned chain rule has been applied inside the expectation. Substituting the state equation (Eq. 6.39) gives

$$\dot{\mathbf{X}} = \left\langle(\mathbf{A}\vec{x} + \mathbf{B}\vec{u})\,\vec{x}^T\right\rangle + \left\langle\vec{x}(\mathbf{A}\vec{x} + \mathbf{B}\vec{u})^T\right\rangle. \tag{6.52}$$

Expanding each term yields

$$\dot{\mathbf{X}} = \left\langle \mathbf{A}\vec{x}\vec{x}^T \right\rangle + \left\langle \mathbf{B}\vec{u}\vec{x}^T \right\rangle + \left\langle \vec{x}\vec{x}^T\mathbf{A}^T \right\rangle + \left\langle \vec{x}\vec{u}^T\mathbf{B}^T \right\rangle. \tag{6.53}$$

Since the system matrices are deterministic, they can be taken outside the expectations, giving

$$\dot{\mathbf{X}} = \mathbf{A}\langle \vec{x}\vec{x}^T \rangle + \langle \vec{x}\vec{x}^T \rangle \mathbf{A}^T + \mathbf{B}\langle \vec{u}\vec{x}^T \rangle + \langle \vec{x}\vec{u}^T \rangle \mathbf{B}^T. \tag{6.54}$$

Using the definition of $\mathbf{X}$ from Eq. 6.48, this becomes

$$\dot{\mathbf{X}} = \mathbf{A}\mathbf{X} + \mathbf{X}\mathbf{A}^T + \mathbf{B}\langle \vec{u}\vec{x}^T \rangle + \langle \vec{x}\vec{u}^T \rangle \mathbf{B}^T. \tag{6.55}$$

At this stage, the evolution of the state covariance depends on the input-state cross-correlation terms $\langle \vec{u}\vec{x}^T \rangle$ and $\langle \vec{x}\vec{u}^T \rangle$. These must now be evaluated.

6.4.4 Evaluation of the Cross Terms

To evaluate the cross-correlation terms, the state is written as the convolution of the input with the system's impulse response:

$$\vec{x}(t) = \int_{-\infty}^{t} e^{\mathbf{A}(t-\tau)}\mathbf{B}\vec{u}(\tau)\, d\tau. \tag{6.56}$$

The use of these integration bounds assumes that the system is asymptotically stable so that the effect of any initial condition has decayed away. Using this expression, the first cross term becomes

$$\langle \vec{u}(t)\vec{x}^T(t) \rangle = \left\langle \vec{u}(t) \int_{-\infty}^{t} \vec{u}^T(\tau)\mathbf{B}^T e^{\mathbf{A}^T(t-\tau)}\, d\tau \right\rangle. \tag{6.57}$$

Interchanging the expectation and the integration gives

$$\langle \vec{u}(t)\vec{x}^T(t) \rangle = \int_{-\infty}^{t} \langle \vec{u}(t)\vec{u}^T(\tau) \rangle \mathbf{B}^T e^{\mathbf{A}^T(t-\tau)}\, d\tau. \tag{6.58}$$

Now let the input covariance be described by the matrix $\mathbf{W}$, such that

$$\langle \vec{u}(t)\vec{u}^T(\tau) \rangle = \mathbf{W}\,\delta(t-\tau), \tag{6.59}$$

where $\mathbf{W}$ specifies which input channels are white-noise channels and what their relative strengths are. If the input vector $\vec{u}(t)$ contains both stochastic disturbance inputs and non-stochastic channels, then $\mathbf{W}$ is generally not the full identity matrix of dimension m.

For example, suppose the m input channels are ordered as

$$\vec{u}(t) = \begin{bmatrix} u_{\mathrm{n1}}(t) \\ u_{\mathrm{c}}(t) \\ u_{\mathrm{n2}}(t) \\ u_{\mathrm{ref}}(t) \end{bmatrix}, \tag{6.60}$$

where $u_{\mathrm{n1}}(t)$ and $u_{\mathrm{n2}}(t)$ are independent unit white-noise inputs, while $u_{\mathrm{c}}(t)$ (the control input) and $u_{\mathrm{ref}}(t)$ (a reference input) are not white-noise inputs. In that case,

$$\mathbf{W} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \tag{6.61}$$

The ones indicate the channels that are being treated as unit white-noise inputs, while the zeros indicate channels that do not directly contribute white-noise power in this model.

In the special case where all m input channels are independent unit white-noise processes, one has

$$\mathbf{W} = \mathbf{I}_m. \tag{6.62}$$

This can be set up rather easily by simply temporarily removing any non-unit white-noise input channels from the input vector $\vec{u}(t)$ and the corresponding columns from the input matrix, $\mathbf{B}$.

Substituting Eq. 6.59 into the integral gives

$$\langle \vec{u}(t)\vec{x}^T(t) \rangle = \int_{-\infty}^{t} \mathbf{W}\,\delta(t-\tau)\mathbf{B}^T e^{\mathbf{A}^T(t-\tau)}\,d\tau. \tag{6.63}$$

Evaluating the integral using the Dirac delta function yields

$$\langle \vec{u}(t)\vec{x}^T(t) \rangle = \frac{1}{2}\mathbf{W}\mathbf{B}^T \tag{6.64}$$

and thus, by the same reasoning,

$$\langle \vec{x}(t)\vec{u}^T(t) \rangle = \frac{1}{2}\mathbf{B}\mathbf{W}. \tag{6.65}$$

6.4.5 Continuous-Time Lyapunov Equation

Substituting the results from Eqs. 6.64 and 6.65 into Eq. 6.55 gives

$$\dot{\mathbf{X}} = \mathbf{A}\mathbf{X} + \mathbf{X}\mathbf{A}^T + \mathbf{B}\left(\frac{1}{2}\mathbf{W}\mathbf{B}^T\right) + \left(\frac{1}{2}\mathbf{B}\mathbf{W}\right)\mathbf{B}^T. \tag{6.66}$$

Combining the final two terms yields

$$\dot{\mathbf{X}} = \mathbf{A}\mathbf{X} + \mathbf{X}\mathbf{A}^T + \mathbf{B}\mathbf{W}\mathbf{B}^T. \tag{6.67}$$

Define the process noise covariance matrix as

$$\mathbf{Q} \equiv \mathbf{B}\mathbf{W}\mathbf{B}^T. \tag{6.68}$$

As shown in Eq. 6.61, the matrix $\mathbf{W}$ determines which elements of the input vector are modeled as white-noise channels. Entries equal to 1 correspond to unit white-noise inputs, entries equal to 0 correspond to channels that are not modeled as direct white-noise inputs, and more general positive values can be used to represent non-unit noise strengths. If different noise channels are correlated, then $\mathbf{W}$ will also contain off-diagonal terms.

With the process noise covariance matrix $\mathbf{Q}$ defined in this way, the state covariance satisfies the continuous-time Lyapunov equation

$$\dot{\mathbf{X}} = \mathbf{A}\mathbf{X} + \mathbf{X}\mathbf{A}^T + \mathbf{Q}. \tag{6.69}$$

In steady state, the covariance no longer changes in time, so that $\dot{\mathbf{X}} = 0$. The Lyapunov equation then reduces to

$$\mathbf{A}\mathbf{X} + \mathbf{X}\mathbf{A}^T + \mathbf{Q} = 0. \tag{6.70}$$

This equation determines the steady-state covariance of the state vector. In the special case considered above, where $\mathbf{D} = 0$, this relation reduces to the simple form expressed in Eq. 6.47. However, when the feedthrough matrix $\mathbf{D}$ is nonzero, the output covariance contains additional terms and must be treated separately.

6.5 Optimal LQG Controllers With a Bounded $\mathcal{H}_\infty$ Norm

The previous section introduced controls concepts for integration into a mixed-sensitivity solver and defined the relevant state-space representation.

Mixed-sensitivity control offers the best of both LQG and $\mathcal{H}_\infty$ optimization, minimizing noise while ensuring robust performance. Despite its considerable benefits,

it is less commonly known or taught, as it is advanced and algebraically cumbersome to describe and solve.

The methodology can be incorporated into existing designs in a straightforward manner through the addition of a u_∞ input, such that end-users will only need to input the various components of Fig. 7.1, and software can do the rest, shielding the user from the problem description. This then requires numerical algorithms that can reliably solve such mixed-sensitivity problems despite the aforementioned numerical challenges, without requiring careful user tuning. This is largely, though not entirely, achieved through the following methods.

For context, mixed-sensitivity control was studied shortly after the development of a reliable $\mathcal{H}_\infty$- optimal solution using the separation principle. Interestingly, the more direct method of Lagrangian constraints was used to apply the $\mathcal{H}_\infty$ bound to the LQG noise minimization problem. This was explored in the two initial papers leading into this subject [123, 124]. These two works establish the necessary conditions as a set of coupled Riccati equations to solve and maintain a weaker form of the separation principle. Shortly thereafter, sufficient conditions were determined [125] and the two algebraic approaches were related as duals of each other [126]. A notable aspect of these alternate approaches is that they set up the problem using a similar arrangement of inputs and outputs. Namely, there are three sets of inputs and two sets of outputs, or vice versa. One of each is used for observing and control, and the remainder to define the optimization and bound norms. For the phase-margin-enforcing approach considered here, the arrangement of inputs and outputs is why the $\mathbf{T}_2$ and $\mathbf{T}_\infty$ matrices are related and lead to the constraint of Eq. 7.50.

Three inputs and outputs are needed to define a fully general problem that decouples the $\mathbf{T}_2$ and $\mathbf{T}_\infty$ transfer matrices, but this removes all remaining algebraic separability. Since the original algebraic formulations, ongoing work has re-framed the mixed-sensitivity problem using linear matrix inequalities (LMIs) on a convex domain [127–129]. Initial attempts to optimize the system under consideration used this approach through the MATLAB function `h2hinfsyn`. This function uses an LMI approach. For the systems considered here, it was only successful for $\zeta = 0$ problems, with no BNS weighting. It was also unacceptably slow. As to why, it can only be speculated that the LMI formulation, or the specific numerical implementation and internal numerical tolerances, are not currently structured or balanced to handle the large dynamic range of gravitational-wave interferometer models. Alternatively, they may be particularly unsuited when the $\mathbf{T}_2$ weighted

optimization problem is defined so differently than the central controllers of the rather dissimilar $\mathbf{T}_\infty$ problem. For the present study, the earlier algebraic approach of Berstein-Haddad (BH) is utilized [123, 124].

The convention of u_{inf} being on the input side is used, as that is required for the system layout. This follows the conventions of [101] and is dual to [123, 124]. The BH equations are otherwise used, as they present algebraic solutions for the most general problem. The following definitions from BH, in the notation adopted here, are used to express the algebraic Riccati equations that follow.

$$\mathbf{R}_1 = \mathbf{B}_{1n}\mathbf{B}_{1n}^T, \tag{6.71}$$

$$\alpha^2\hat{\mathbf{R}}_2 = \mathbf{R}_2 = \mathbf{D}_{21n}\mathbf{D}_{21n}^T, \tag{6.72}$$

$$\mathbf{R}_{12} = \mathbf{B}_{1n}\mathbf{D}_{21n}^T, \tag{6.73}$$

$$\mathbf{V}_1 = \mathbf{C}_1\mathbf{C}_1^T, \tag{6.74}$$

$$\mathbf{V}_2 = \mathbf{D}_{12}\mathbf{D}_{12}^T, \tag{6.75}$$

and

$$\mathbf{V}_{12} = \mathbf{C}_1^T\mathbf{D}_{12}. \tag{6.76}$$

These factors are related to the noise inputs and FOM outputs. More relations that map the present formulation back to the BH papers are given in Section 6.6. The expressions below require structural similarities between the D matrices of the RMS and $\mathcal{H}_\infty$ problems. Namely, they have to be proportional so that the β factor can be defined as the ratio, $\beta^2 = \mathbf{D}_{21\infty}\mathbf{D}_{21\infty}^T/\mathbf{R}_2$. This is easy to ensure for SISO models, but may be more difficult in MIMO extensions.

The factor α is nominally 1. The equations below are all expressed in terms of $\hat{\mathbf{R}}_2$ rather than $\mathbf{R}_2$ or $\mathbf{R}_{2\infty} = \mathbf{D}_{21\infty}\mathbf{D}_{21\infty}^T$, leading to the included factors of α and β. By choosing $\alpha > 1$, the noise matrices in the problem are effectively rescaled, which is useful in finding solutions.

The set of equations that must then be solved is the following three Riccati equations in the positive semidefinite matrices $\mathbf{Q}$, $\mathbf{Z}$, $\hat{\mathbf{Q}}$. Included below are the equations along with some convenience definitions to shorten the Riccati equations and highlight their quadratic form. The first of the equations, solving for $\mathbf{Q}$, is independent of the others, indicating the remaining separability of the problem. The equations in $\mathbf{Z}$, $\hat{\mathbf{Q}}$ are inter-coupled and both rely on the solution to $\mathbf{Q}$. Their simultaneous solution requires an iterative method. The equations are:

$$0 = \mathbf{A}^T\mathbf{Q} + \mathbf{Q}\mathbf{A} + \mathbf{V}_{1\infty} - (\mathbf{Q}\mathbf{B}_2 + \mathbf{V}_{12\infty})\mathbf{V}_{2\infty}^{-1}(\mathbf{B}_2^T\mathbf{Q}^T + \mathbf{V}_{12\infty}^T), \tag{6.77}$$

$$\mathbf{Q}_a \equiv \mathbf{Q}\mathbf{B}_2 + \mathbf{V}_{12\infty}, \tag{6.78}$$

$$\overline{\mathbf{A}}_Z \equiv \mathbf{A}^T - \frac{\beta^2}{\alpha^2\gamma^2}\hat{\mathbf{Q}}\mathbf{R}_1, \tag{6.79}$$

$$0 = \overline{\mathbf{A}}_Z^T\mathbf{Z}' + \mathbf{Z}'\overline{\mathbf{A}}_Z + \mathbf{R}_1 - \mathbf{Z}'\left(\frac{\mathbf{C}_2^T\hat{\mathbf{R}}_\mathbf{2}^{-1}\mathbf{C}_2}{\alpha^2} - \frac{\beta^4\hat{\mathbf{Q}}\mathbf{R}_1\hat{\mathbf{Q}}}{\alpha^4\gamma^4} - \frac{\beta^2\mathbf{Q}_a\mathbf{V}_{2\infty}^{-1}\mathbf{Q}_a^T}{\alpha^2\gamma^2}\right)\mathbf{Z}', \tag{6.80}$$

$$\mathbf{Z} \equiv \mathbf{Z}'/\alpha^2, \tag{6.81}$$

$$\overline{\mathbf{A}}_{\hat{Q}} \equiv \mathbf{A}^T - \mathbf{C}_2^T\hat{\mathbf{R}}_\mathbf{2}^{-1}\mathbf{C}_2\mathbf{Z}, \tag{6.82}$$

and

$$0 = \overline{\mathbf{A}}_{\hat{Q}}\hat{\mathbf{Q}} + \hat{\mathbf{Q}}\overline{\mathbf{A}}_{\hat{Q}}^T + \mathbf{Q}_a\mathbf{V}_{2\infty}^{-1}\mathbf{Q}_a^T + \hat{\mathbf{Q}}\frac{\beta^2}{\gamma^2}\mathbf{Z}\mathbf{C}_2^T\hat{\mathbf{R}}_\mathbf{2}^{-1}\mathbf{C}_2\mathbf{Z}\hat{\mathbf{Q}}. \tag{6.83}$$

Using the solutions ($\mathbf{Q}$, $\mathbf{Z}$, and $\hat{\mathbf{Q}}$) to the above equations, the optimal controller is then computed to be

$$\mathbf{K}_{\mathcal{H}_{2/\infty}} = \left[\begin{array}{c|c} \mathbf{A}_\mathrm{K} & \mathbf{B}_\mathrm{K} \\ \hline \mathbf{C}_\mathrm{K} & 0 \end{array}\right], \tag{6.84}$$

where

$$\mathbf{B}_\mathrm{K} = \mathbf{Q}_a\mathbf{V}_{2\infty}^{-1}, \tag{6.85}$$

$$\mathbf{C}_\mathrm{K} = -\hat{\mathbf{R}}_\mathbf{2}^{-1}\mathbf{C}_2\mathbf{Z}, \tag{6.86}$$

and

$$\mathbf{A}_\mathrm{K} = \mathbf{A}^T + \mathbf{C}_2^T\mathbf{C}_\mathrm{K} - \mathbf{B}_\mathrm{K}\mathbf{B}_2^T - \mathbf{B}_\mathrm{K}\mathbf{D}_{22}^T\mathbf{B}_\mathrm{K}. \tag{6.87}$$

The last term on $\mathbf{A}_\mathrm{K}$ accounts for the $\mathbf{D}_{22} \neq 0$, so no further adjustment is required.

The challenge in solving these equations is in finding the simultaneous solutions. Some discussion from [123] is noted here. As long as $\hat{\mathbf{Q}}$ is positive semidefinite, then the solution for $\mathbf{Z}$ will be noise-optimal and stable, but may not obey the given $\mathcal{H}_\infty$ constraint. Likewise, for any positive semidefinite $\mathbf{Z}$, a solution for $\hat{\mathbf{Q}}$ indicates that the control is stable and the $\mathcal{H}_\infty$ bound is met, although it may not be noise optimal. These equations are solved iteratively using a holonomy method, where parameters are scanned from a parameter space where solutions are reliably found

to one where they are more challenging to find. The initial condition of the iteration is for $\hat{\mathbf{Q}} = 0$.

Instead of performing holonomy on solutions where γ is transitioned from ∞ to its final value, α is instead transitioned from ∞ to the value 1 while the coupled equations are solved alternately. This method is physically motivated. It is equivalent to scaling the sensing noise from being large and dominant, where loops do not need aggressive shaping, to the final form where sensing noise is many orders of magnitude below the environmental noise, while maintaining γ at its nominal desired value.

6.5.1 Controller Stability and Implementation

While the closed-loop system is guaranteed to be stable with either the $\mathcal{H}_2$ or the $\mathcal{H}_\infty$ bounded controller, the controllers alone are not guaranteed to be stable. It has been observed that the outputs of the mixed-sensitivity solver produce unstable $\mathbf{K}_{\mathcal{H}_{2/\infty}}$.

In theory, there is nothing concerning about a controller alone being unstable since the overall stability of the system is determined in the closed-loop case. However, controllers that are themselves unstable may still be undesirable in some applications. Especially when gaining lock in one of the interferometer's degrees of freedom (DOFs), a stable controller is beneficial. A controller that is independently stable and whose closed loop is stable is called a "strongly stable" controller. The equations for strongly stable controllers have been computed in [130, 131]. However, implementation is beyond the goals of the present chapter.

It has been found that in situations where the solved controller is unstable, simply flipping its unstable poles to the stable side still generally results in a closed-loop stable system. This is performed during a reduction stage to lower the order of the controller for implementation in digital control. It has been verified that the final system model using an order-reduced controller is still stable and that the phase margin of the order-reduced controller is still acceptable after the reduction step.

6.6 Translation of the BH Equations

The notation table of [123], as well as the notation table and Fig. 1 mapping in [124] can be used to translate the notation used here into the algebraic equations, through

the following mappings:

$$A \to \mathbf{A}^T, \qquad B \to \mathbf{C}_2^T, \qquad C \to \mathbf{B}_2^T, \tag{6.88}$$

$$D \to \mathbf{D}_{22}^T, \qquad E_1 \to \mathbf{B}_{1n}^T, \qquad E_{1\infty} \to \mathbf{B}_{1\infty}^T, \tag{6.89}$$

$$E_\infty \to \mathbf{D}_{11\infty}^T, \qquad E_2 \to \mathbf{D}_{21n}^T, \qquad E_{2\infty} \to \mathbf{D}_{21\infty}^T, \tag{6.90}$$

$$D_1 \to \mathbf{C}_1^T, \qquad \text{and} \qquad D_2 \to \mathbf{D}_{12}^T. \tag{6.91}$$

Because E_∞ is structurally zero in the problem setup, the following BH parameters are simplified as:

$$\mathbf{M} = \mathbf{1}, \qquad \mathbf{N} = \mathbf{1}, \tag{6.92}$$

$$\mathbf{V}_{1\infty} = \mathbf{V}_1, \qquad \mathbf{V}_{2\infty} = \mathbf{V}_2, \qquad \mathbf{V}_{12\infty} = \mathbf{V}_{12}, \tag{6.93}$$

$$\mathbf{R}_{01\infty} = \mathbf{0}, \qquad \mathbf{R}_{02\infty} = \mathbf{0}. \tag{6.94}$$

This simplification leads to the reduced expressions

$$\mathbf{R}_{1\infty} = \mathbf{B}_{1\infty}\mathbf{B}_{1\infty}^T, \tag{6.95}$$

$$\mathbf{R}_{2\infty} = \mathbf{D}_{21\infty}\mathbf{D}_{21\infty}^T, \tag{6.96}$$

$$\mathbf{R}_{12\infty} = \mathbf{B}_{1\infty}\mathbf{D}_{21\infty}^T, \tag{6.97}$$

and

$$\hat{\mathbf{R}}_2 \equiv \mathbf{R}_2/\alpha^2 = \mathbf{R}_{2\infty}/\beta^2, \tag{6.98}$$

where

$$\alpha \equiv 1. \tag{6.99}$$

It is also assumed that there is no additional shaping filter on the u_∞ input, causing the $E_{1\infty}$ term from BH to be zero. This leads to the additional simplifications:

$$\mathbf{R}_{1\infty} = \mathbf{0} \quad \text{and} \quad \mathbf{R}_{12\infty} = \mathbf{0} \tag{6.100}$$

Note that all of these are then applied to the full form equations of [124] and then simplified into the **Z**-form description following Section V of [123].

6.7 Alternate Noise Arrangement

Along with the model presented in the main text, there are two major alternate approaches to where to place the FOMs to model bilinear-type noise couplings, and a third approach is also suggested. Each has its own reasons for implementation.

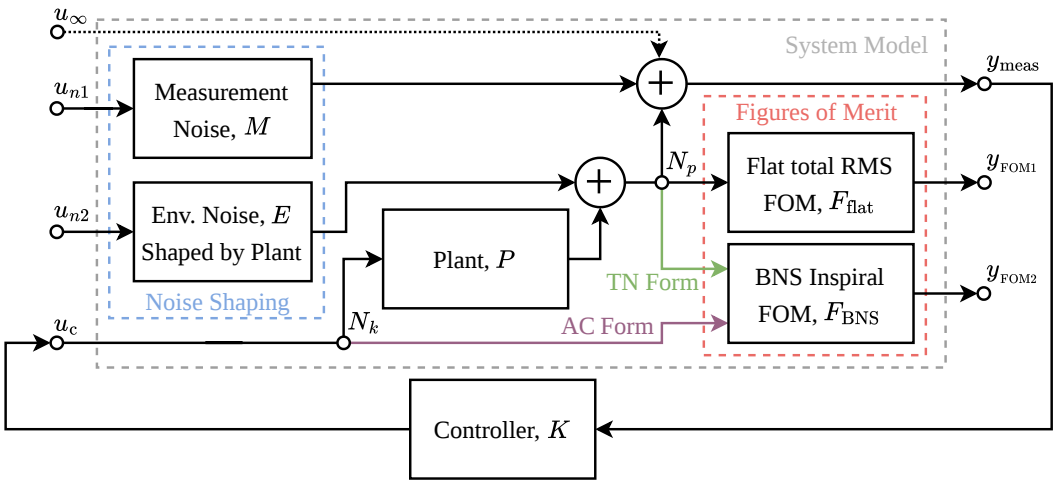


Figure 6.3: Two alternate system layouts to Fig. 7.1, the noise-augmented plant system with a feedback controller attached. In this configuration, the environmental noise block is the equivalent of $|EP|$ rather than just E as in all previous system layouts. This change does not alter the system but allows an open-loop spectrum at point y_{meas} to be fit, as is shown in Fig. 7.2. Here, the FOMs are arranged showing two alternate bilinear noise optimizations. The first, the actuator-centered (AC) form often used by LIGO, is where the BNS FOM is connected to the control signal, u_{c}, and treated as a noise, N_{K}, with the purple connection. The second is the total-noise (TN) form that can be used for non-coupled DOFs and connects the BNS FOM to the plant's state noise N_{p} with a green connection. Only one of these two alternate connection forms can be used at once. Adapted from [100].

Another alteration present in this system is the apparent bypass of the plant by the environmental noise. In this configuration, the environmental noise-shaping filter now also includes the shape of the plant. This means the new shape of the alternate E' filter can be written as $|EP|$. Letting this input noise-shaping filter represent the environmental noise passed through the plant allows it to be connected after the plant, since it already has picked up the shape from the plant. This alteration allows the very useful option of fitting the filter to an open-loop spectrum at y_{meas} as is shown in Fig. 7.2.

6.7.1 Actuator-Centered form

For the ASC bilinear noise model often assumed by LIGO controls commissioners, the controls noise arises solely from the u_{c} feedback contribution to the noise. This type of bilinear noise injection can be modeled by the following coupling.

$$h_{signal}(t) = h_{\text{gw}}(t) + N_{\text{det}}(t) + \chi N_{\text{p}}(t) u_{\text{c}}(t). \tag{6.101}$$

This coupling results when the coupling function is computed from

$$|C|^2 = \frac{S_{h,\,\mathrm{exc}} - S_h}{S_{\mathrm{K},\,\mathrm{exc}} - S_{\mathrm{K}}}, \tag{6.102}$$

where the excited channel in the denominator is the output of the control filter (for alignment, it is related to units of torque on the mirror).

Fig. 6.3 shows how to set up the augmented plant system to implement this alternate form. Note that the injection of the environmental noise is now after the plant, which puts it into the same units as the measurement noise. This makes the injection into DARM appear as:

$$S_{\mathrm{K}} = \left|\frac{G(f)}{1 - G(f)}\right|^2 (S_{\mathrm{env}} + S_{\mathrm{meas}}), \tag{6.103}$$

for this arrangement of noises in the model.

This arrangement is primarily a conceptual change from the model used in the main text. In it, the coupling function takes on different units since it is modeled before entering the plant, and so it is in units of DARM length/torque for the alignment control example considered here.

6.7.2 Total-Noise Optimization Form

The form and connection point of the FOMs that is perhaps the most appropriate for quadratically-nonlinear (rather than bi-linear) systems requires using the total noise S_{p} as discussed in Section 7.2.2. This setup is detailed in Fig. 6.3 with the connection from N_{p} to the BNS FOM. The connection of both FOMs to N_{p} ensures that both the actuation point controls noise, as well as environmental noise, are minimized together. This is a more typical arrangement of inputs/outputs and weight filters for LQG problems and does not require two copies of the plant. It corresponds to quadratically nonlinear noise couplings of the form

$$h_{signal}(t) = h_{\mathrm{gw}}(t) + N_{\mathrm{det}}(t) + \chi N_{\mathrm{p}}^2(t). \tag{6.104}$$

In terms of numerical implementation, there are several reasons that the LIGO preferred forms in Section 6.7.1 and the main text are less preferable to the total-noise form. One of the major reasons is that in LIGO's form, most of the low-frequency environmental noise shows up in S_{c} when $|G|$ is large. The dynamic range of the BNS FOM covers nearly all of the range of a double float. Thus, in computing the noise contribution in this model, there are numerical difficulties multiplying

the environmental noise filter with the BNS-weighting filter inside a state-space. The second reason is that bilinear noise is well motivated to be driven by N_p, or at least some mixture of N_p and u_c. In some instances, LIGO has found that further accelerating the roll-off of G does not reduce the noise in DARM. The primary noise coupling chosen here provides a reason for that result: the bilinear noise can also be driven by residual environmental noise in the plant that is reflected in N_p rather than in N_a or $u_\mathrm{c} = N_K$.

6.7.3 Measurement-Only Optimization Form

A final alternate approach would be to use a BNS FOM channel with the noise coupling $S_{\mathrm{c}'} = |G/(1-G)|^2 S_\mathrm{meas}$ only, without the environmental noise contribution. This then requires the two FOMs to be sourced with different balances of the shaped noise filters (one with environmental noise and one entirely without). A means to implement this through the block diagram without two copies of the unknown controller, K, is not known; however, it should be possible to develop a new set of Riccati equations to solve. This might be done by determining the equivalent noise model using a sum of Lyapunov equations, and then solving for the constrained noise minimization problems using the techniques of [106, 123, 124, 132]. Doing so is likely to further reduce the separability structure expected of LQG, and may result in three coupled Riccati equations, rather than the two coupled equations of mixed-sensitivity control. This alternate approach could be primarily useful for numerical implementation reasons, and perhaps would be useful to investigate using a convex optimization approach that can more easily represent and solve constrained optimization problems.

Chapter 7

OPTIMAL CONTROL OF LIGO'S ASC DHARDY LOOP

"The proof of the pudding is in the eating"

— Miguel de Cervantes, *Don Quixote*[a]

[a]The quoted wording comes from Peter Anthony Motteux's English translation of Cervantes's Spanish proverb "al freír de los huevos lo verá" ("you will see when the eggs are fried").

LIGO is a ground-based gravitational wave interferometer that first detected gravitational waves in 2015 [2]. Since its first detection, the LIGO Scientific Collaboration (LSC) has been working to reduce noise in the instrument to increase its sensitivity and, therefore, detect even fainter gravitational wave signals [79, 133]. These significant improvements in LIGO's sensitivity, as well as the inclusion of additional detectors, have led to an increased detection rate that expands astrophysical study [79, 133–137].

There is considerable interest in reducing low-frequency noise in LIGO (<30 Hz). There are many low-frequency waveforms, like high-mass BBH inspirals, whose detection would benefit from an increased sensitivity in the low-frequency band [138]. Other waveforms that extend to higher frequencies will benefit from higher signal-to-noise detections, improved parameter estimation, and potentially early warning of ongoing events during the low-frequency portion of long-lived waveforms. Currently, at these low frequencies, both LIGO detectors are limited by noise introduced by the feedback control loops that stabilize auxiliary degrees of freedom needed to maintain interferometer operation. Due to the use of resonant cavities, even small variations in the auxiliary degrees can result in highly nonlinear responses that can produce noise. The lowest meaningful order of nonlinearity results in two channels, one a control signal and one a drifting parameter, which multiply to produce bilinear noise that is indistinguishable from an astrophysical strain signal. Improving the controllers used in LIGO's feedback loops and generally improving the processes for creating these controllers could significantly reduce the low-frequency noise in the detectors, as well as allowing for resources to be spent on other areas of

improvement.

Motivation

The LIGO instrumentation community widely uses classical control design methods. While an entire system has many degrees of freedom, they are almost all designed and engineered as individual SISO signal paths using transfer functions, Bode plots, and frequency-resolved budgets of noise contributions. These subsystems and control loops are designed depending on how they interconnect to affect the principal measurement channels. The software interfaces of LIGO are optimized for this approach, and it is largely successful; however, it has several weaknesses. First, the loops are all hand-designed, which can get the system operating fast, but compounds negatively for global optimization. In particular, without a lower bound on performance, a halting criterion for optimizing is established either by experience or by exhaustion. Furthermore, the design of coupled MIMO systems must often be iterative, as loop shapes and noise consequences are interdependent. More modern state-space controls approaches are an attractive research topic, promising guaranteed existence, stability, and optimality properties even for complicated, intercoupled MIMO systems.

So far, modern controls approaches have not been widely adopted due to the unique requirements of gravitational-wave detectors. First, these approaches require very specific cost functions to form well-posed optimization problems. Second, these approaches usually assume white noise inputs; thus, system models must be augmented with noise-shaping filters. The fitting of such filters is nontrivial but now reliable with modern algorithms. The third issue is that interferometers rarely have actuator cost functions (unlike, e.g., aerospace); thus, typical examples of regulator cost do not immediately apply. The fourth issue is that interferometer loops are limited by transmission delays, which create non-minimum-phase zeros that must be incorporated into the control loop design. The final and highly significant hurdle is dynamic range. The range of shaped noise in gravitational-wave detectors spans 8–12 orders of magnitude, from low-frequency seismic motion to interferometric sensors. The combination of these factors causes optimal solutions to sit on the edge of stability regions, where even the best-known numerical algorithms struggle.

Approach

An intermediate and transitional approach is adopted in an effort to supplant hand-tuning while managing these challenges. Modern methods are used to pose and

solve a SISO controls problem. Future work can extend these examples to more intrinsically MIMO problems. A major focus of the present study is the problem definition, through which the modern controls problem is posed using FOMs that are relevant to optimizing gravitational-wave detectors. Furthermore, the mixed-sensitivity approach developed here is necessary to find robust solvers that have acceptable gain and phase margins. This approach is mathematically complex to describe, but has largely been solved in previous controls literature. By finding solutions with acceptable phase margins, much like hand-designed controllers, the numerical solutions are generally better behaved and do not live "near the edge" of admissible solution regions, where small perturbations, numerically or in the instrument plants, push them outside. A mixed-sensitivity solver rapidly arrives at controllers qualitatively similar to hand designs but guaranteed optimal for the defined FOMs that are tied to the astrophysics objectives.

The general state-space and mixed-sensitivity synthesis framework was developed in Chapter 6. This chapter applies that framework to the LIGO Hanford DHARD-yaw loop. It first defines the controls-noise and astrophysical performance objectives, constructs the shaped-noise and FOM model, and derives the bilinear-noise trade-off. The pure LQG Pareto front is then compared with the robustness-constrained controller families, followed by comparisons with AI-informed and convex-optimization approaches. Keeping the application and results in this chapter allows the measured plant, physical noise couplings, and controller performance to be discussed together after the general synthesis method has been established.

7.1 Background

LIGO's sensitivity is limited by the sum of a number of individual noise sources in the interferometer. The extent to which these noise sources obscure astrophysical signals is expressed using the SNR. The SNR of a signal can be calculated from the Fourier transform of the event waveform, $h_{\mathrm{gw}}(f)$, and the PSD of the interferometer's noise, $S_h(f)$, using

$$\mathrm{SNR}^2 = 4\int_0^\infty \frac{\left|h_{\mathrm{gw}}(f)\right|^2}{S_h(f)}df. \tag{7.1}$$

The detection range,

$$d_{\mathrm{range}} = \frac{d_{\mathrm{gw}}}{2.26}\frac{\mathrm{SNR}}{\rho_0}, \tag{7.2}$$

is the average effective distance to which events louder than a threshold SNR called ρ_0 can be detected. The range can be calculated using the SNR formula and the

distance to the fiducial source. For an $h_{\rm gw}(f)$ waveform modeled from a source at distance $d_{\rm gw}$, the detection range averaged over the sky is included above. The threshold SNR is usually taken to be $\rho_0 = 8$, and the factor of 2.26 accounts for the average of the interferometer antenna pattern over the sky [139].

The astronomical detection range is critical to the interferometer's sensitivity because it determines the detection rate, $\mathcal{R}_{\rm det}$. The astronomical detection range, detection rate, and SNR are related through

$$\mathcal{R}_{\rm det} \propto d^3_{\rm range} \propto {\rm SNR}^3. \tag{7.3}$$

Even minor decreases in the interferometer's noise can have a significant impact on the detection rate. One of LIGO's long-standing goals is increasing the detection rate for BNS inspirals. Binary neutron star waveforms are taken to be the standard fiducial for computing SNR as a FOM for the interferometer performance considered here.

7.1.1 Noise from Control Systems

At the low frequency end of its detection band, LIGO's sensitivity is limited by noise injected by control systems. At frequencies below LIGO's detection band, the control systems stabilize the operating points of otherwise highly nonlinear optical cavity resonators and interferometers. Noise coming from the control actuation of cavity alignment and auxiliary length are two of the largest of all known noise sources in the detector below 10 Hz [79, 133]. Interferometric controls noise is any noise that arises from the addition of a stabilizing or noise-suppressing controller to a DOF of the interferometer system. When a controller is added in a feedback loop to a DOF, some of the noise inherent to that DOF, along with new measurement noise, passes through the controller and is (re-)injected into the system at the actuator. While a controller can reduce the overall noise of a DOF, it also introduces a new noise source from the controller's sensor. This occurs when the loop bandwidth is higher than the frequency where (input-referred) plant noise becomes lower than the sensing noise. For complex systems with many DOFs and many control systems, the controls noise can become a significant noise source. From optics controllers to seismic isolation, each LIGO detector has more than 300 controllers that all maintain its operating point but add noise to the principal readout: the DARM. This is the measurement of arm-length differences caused by gravitational waves, and thus the output of interest for astrophysics.

A model is constructed here for how a control system for a single DOF affects the signal and introduces noise. Bilinear noise is later expressed in terms of this model. The plant noise of some degree of freedom under feedback control, N_c, affects the signal seen at the output of the interferometer through

$$h_{\text{signal}}(f) = h_{\text{gw}}(f) + N_{\text{det}}(f) + C(f)N_c(f), \tag{7.4}$$

where $h_{\text{gw}}(f)$ is a potential signal, N_{det} is all the detector noise from sources other than the DOF being modeled, and $C(f)$ is a frequency-dependent coupling of the DOF into DARM. This coupling is often indirect and sourced from bilinear effects. For simple estimates, it is often possible to assume $C(f)$ is constant across all frequencies, but in the general case, measurement data necessitates that it be frequency-dependent. An example measurement of $C(f)$ and physical motivation are described in Section 7.2.2. The PSD of the total noise in the detector is thus

$$S_h(f) = S_{\text{det}}(f) + |C(f)|^2 S_c(f), \tag{7.5}$$

where $S_{\text{det}}(f)$ is the PSD of all the detector noise from sources other than the DOF being modeled and $S_c(f)$ is the PSD of the noise of interest, in this case a definition of controls noise. By linearizing Eq. 7.1, and including Eq. 7.5's expression of $S_h(f)$, the controls noise dependent SNR becomes

$$\text{SNR}^2 = \text{SNR}^2_{\text{det}} - \text{SNR}^2_{\text{lost}}, \tag{7.6}$$

where

$$\text{SNR}^2_{\text{det}} = 4\int_0^\infty \frac{\left|h_{\text{gw}}(f)\right|^2}{S_{\text{det}}(f)}\,df \tag{7.7}$$

and

$$\text{SNR}^2_{\text{lost}} = 4\int_0^\infty \frac{|C|^2\left|h_{\text{gw}}(f)\right|^2}{S_{\text{det}}(f)^2} S_c(f)\,df. \tag{7.8}$$

Adapting Eq. 7.2 gives the controls-noise-dependent expression for the detection range lost as

$$d_{\text{lost}} = \frac{d_{\text{gw}}}{2.26\rho_0}\frac{\text{SNR}^2_{\text{lost}}}{2\text{SNR}_{\text{det}}}. \tag{7.9}$$

The lost detection range can also be computed using the linearization. The point of linearizing the equation is to split it into terms that show the specific impact of

a particular noise, S_c, on the SNR and BNS range. In this equation, S_c can be any noise of interest, assuming it is small enough to keep the linearization accurate. The S_c term is inside an integral that includes a frequency-dependent weighting. That weighting term will later be useful to define a shaping filter for the controls optimization problem in terms of the lost SNR or range cost.

The RMS of a noise from channel N_c is indicated by $\tilde{R}_c$. In order to keep the interferometer locked, there is a maximum allowable RMS. If the noise in the interferometer exceeds that limit, the interferometer is unable to maintain operation, e.g., sufficient alignment.

$$\tilde{R}_c \equiv \sqrt{\int_0^\infty S_c(f)df} < \tilde{R}_{c,\text{unlock}}. \tag{7.10}$$

A fundamental requirement of control systems is to ensure that the controlled DOF does not exceed its lock-loss threshold. Thus, there is a tradeoff between operational requirements and noise injection.

There are two primary sources of controls noise: environmental noise and measurement noise. Environmental noise is any noise that acts on the states of the suspended interferometer plant. For example, seismic noise is an environmental noise that is of great concern for ground-based interferometers. Measurement noise is the difference between the measured states of a system and their true value, generally due to an imperfect intermediate sensor used to probe the states. Both the measurement and environmental noise pass through a control filter and then feed back into the system, but measurement noise would not otherwise affect plant states except through the control system actuating it into real plant noise.

Considering only two noise sources is a useful simplification. In reality, there are many noise sources that all contribute differently to every DOF in the interferometer. However, these noise sources can only enter the system in two places: either before the plant (environmental noise) or before the controller (measurement noise). This distinction also aligns with the assumptions and model of a typical optimization using classical control theory to shape the open-loop gain.

Without physically decreasing environmental or measurement noise, which much of the engineering resources of LIGO are dedicated to, the best way to decrease controls noise is to find an optimal controller, K, such that the minimum amount of noise is fed through into DARM while achieving a total RMS goal.

Two channels, N_p and N_a, are distinguished here. The former, which must meet a noise threshold, is the sum of environmental noise and controller actuation, filtered by the plant's transfer function. This noise, N_p, is the noise directly seen by the interferometer. The latter, N_a, is the signal sent by an actuator via its control system and shaped by the plant's transfer function, excluding any direct contributions from environmental noise. These two noises are presented visually in Fig. 7.1. The optimal controller must ensure that the plant's noise $\tilde{R}_p$ meets requirements by shaping its spectrum S_p through the environmental and measurement noise contributions.

$$S_p(f) = \left|\frac{P(f)}{1 - G(f)}\right|^2 S_{env}(f) + \left|\frac{G(f)}{1 - G(f)}\right|^2 S_{meas}(f), \tag{7.11}$$

where

$$G(f) = K(f)P(f) \tag{7.12}$$

is the open-loop gain of the full system, $P(f)$ is the transfer function of the plant component, either the input or output of $P(f)$ represents the degree of freedom to be controlled, $K(f)$ is the transfer function of the controller, $S_{env}(f)$ is the PSD of the environmental noise, and $S_{meas}(f)$ is the PSD of the measurement noise. Here, it is explicitly indicated that each of these expressions is a function of frequency; this dependence is continued but will not be explicitly stated in the following formulas.

The total plant noise, N_p, which is presented in Eq. 7.11, is the typical minimization metric for modern controls. It is the measure of all the noise coming out of a system, including all coherent cancellations and interactions. In a system with no cross-coupling to any other degree of freedom, minimizing S_p would minimize the sum of both measurement noise and environment noise injected into the system.

The definition of N_p is often presented as controls noise in modern controls. However, this differs from most previous work in gravitational wave interferometer sensing and control, where controls noise is defined as any noise that passes through the controller and becomes a new contribution through actuation on the DOF. The definition of actuation point controls noise, S_a, is

$$S_a = \left|\frac{PG}{1 - G}\right|^2 S_{env} + \left|\frac{G}{1 - G}\right|^2 S_{meas}, \tag{7.13}$$

and

$$N_a = \frac{PG}{1 - G} N_{env} + \frac{G}{1 - G} N_{meas}. \tag{7.14}$$

The actuation point controls noise definition is useful because optimizing to reduce it is more directly a means of penalizing actuation, as K is in the numerator of both environmental and measurement noise contributions to Eqs. 7.13 and 7.14. In large multi-DOF coupled systems like LIGO, it is often true that actuation on one DOF can inject noise into another. The PSD S_c, from Eq. 7.6, could be set equal to any noise of interest like S_p; however, in the present analysis $S_c = S_a$, as is appropriate for the alignment-control DOF investigated here and as is typical for GW controls. The noise from this metric, N_a, is then used as the noise, N_a, in Eq. 7.4. An approach more widely used in modern controls that only uses S_p as an optimization metric is discussed in Section 6.7.

In the present analysis, both S_a and S_p are used. The former is used for the control effect on SNR and BNS range, and the latter for the minimization of the total RMS, $\tilde{R}_p$ from Eq. 7.10, needed to maintain the operating point. Finding a "best" controller is difficult because of the trade-off between suppressing environmental noise in S_p with a large K and suppressing measurement noise in S_a with a small K. Finding the optimal controller requires a multi-objective optimization: the RMS controls noise, $\tilde{R}_p$, and the BNS detection range.

7.1.2 Application to Alignment Sensing Control System

Controls noise associated with the ASC system is one of the largest sources of controls noise at low frequencies [79, 133]. The ASC system is responsible for keeping the LIGO test masses aligned in their angular degrees of freedom. Each LIGO interferometer has eight principal control filters dedicated to maintaining angular stability in the arm cavities. These eight control filters are the product of three distinguishing degrees. Each system is either in pitch or yaw (P/Y) rigid-body rotation, differential or common motion between the two Michelson arms (C/D), and either HARD or SOFT spring reaction to motion.

HARD modes are when the two test masses that make up LIGO's Fabry-Pérot arm cavities rotate in opposite directions, e.g., one rotates very slightly up, and the other rotates very slightly down. Since LIGO's Fabry-Pérot cavities have a high g-factor, the differential rotation causes a rotation of the optical axis, and radiation pressure pushes the arms back towards their nominal angle. Thus, HARD modes increase suspension stiffness and occur at higher frequencies. SOFT modes are the opposite, where the mirrors move at a common angle, which translates the cavity axis. The translation moves the beam off center and produces a negative spring

constant (anti-spring).

The name of each mode is the composition of these distinguishing degrees; i.e., CSOFTY is the common soft mode in the yaw angular tilt. In this study, only the DHARD yaw controller is calculated. The choice of actuation point noise optimization for the DHARD yaw DOF presented here was motivated by the strong coupling between angular actuation and length in LIGO Hanford's arms.

7.2 Problem Formulation

The heart of the controls noise problem is straightforward to outline: too much measurement noise gets passed through the controller and is actuated spuriously into the DARM DOF, injecting noise. As with all LIGO systems, the goal is to minimize the noise that is fed into the readouts of the interferometer that contain information of astrophysical importance, mainly the DARM readout. The solution to this control problem is more difficult, as it must account for many different factors, including the plant's response, the measurement noise passed through the control system, and the system's stability and robustness.

The modern controls optimization method that is well-suited to this problem is LQG control. An LQG controller is the optimal linear controller for a system that is disturbed by Gaussian white noise and has a quadratic (RMS-like) cost function. Described in Sections 6.3.2 and 6.3.3, LQG controllers minimize a system's $\mathcal{H}_2$ norm, which is a measure of the RMS at system outputs from all of its noise inputs. The "$\mathcal{H}$" stands for solutions within a Hardy space, indicating that the system must be closed-loop stable. This is necessary to have well-posed noise in the time domain, and appropriate for implementable feedback controllers. The typical quadratic cost function for LQG is based on the total noise RMS, S_p, of a DOF. LIGO's performance relies on multiple criteria, so a weighted RMS that includes both the total noise and the lost detection range as figures of merit is used.

So far, the noise model of Section 7.1.1 has not included the nonlinear/bilinear aspects. It has simply been a model of direct, linear noise coupling, which should naively only be optimized for the BNS range while ignoring the DOF RMS. The following sections investigate how to incorporate a weighted RMS to balance noise suppression against control noise injection.

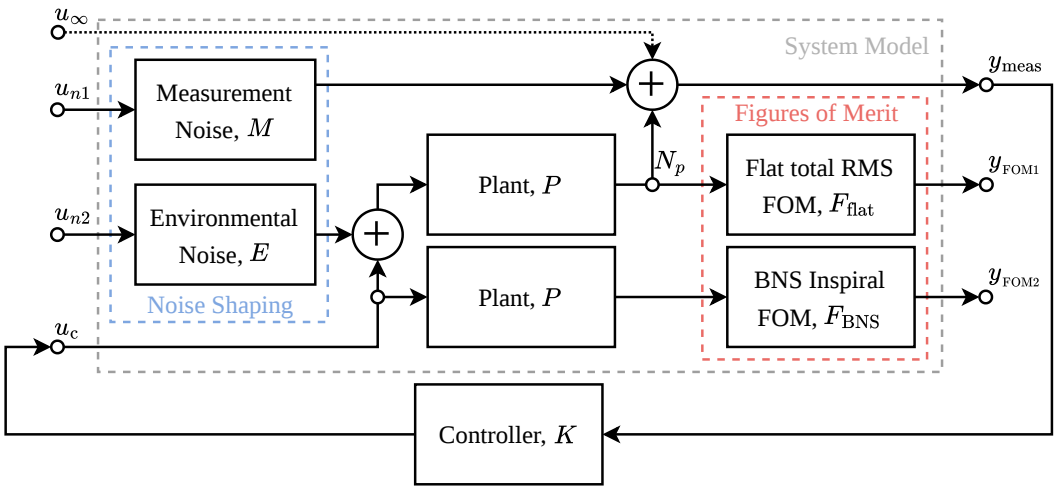


Figure 7.1: The layout of the modeled system with a controller attached. Here, the input white noise, u_{n1}, is being shaped to appear as environmental noise seen by the plant. The system has two FOMs, which are simply transfer functions that weigh the output noise: a broadband FOM outputs the unshaped noise, i.e., every frequency has the same weight, and a BNS inspiral FOM, which represents the band of detection for BNS signals. In this model, the two plant blocks are identical and serve as a way to represent S_a for actuation-point controls noise minimization. These FOMs are discussed further in Section 7.2.2. The input associated with u_∞, which has a dotted line, is used for the $\mathcal{H}_\infty$ minimization discussed in Section 7.5.1. Adapted from [100].

7.2.1 System Overview

When creating a SISO controller for a single DOF, a MIMO system is constructed that takes into account the multiple inputs that can affect the DOF, as well as multiple outputs to measure its performance. This system, whose block diagram is presented in Fig. 7.1, includes all of the information that is used to find the optimal controller, mainly of three types: the noise in the system, the plant of the system, and the FOMs of the system. The first two of these pieces are rather straightforward and regularly measured at the LIGO sites, while the motivation for the FOMs is given in Section 7.2.2.

The full system model, outlined by the dashed grey box in Fig. 7.1, includes four inputs, a control input, u_c, a measurement noise input, u_{n1}, a state noise input (environmental noise), u_{n2}, and an $\mathcal{H}_\infty$ disturbance input. Both noise inputs are white noise inputs that pass through shaping filters to model the measurement and environmental noise. The control input, u_c, is connected to the output of the controller and is not a white noise input. The $\mathcal{H}_\infty$ disturbance input, u_∞, serves as a port for the solver to inject unshaped signals into the system and is a necessity

for optimal control, as analyzed later. The system has two types of outputs: the figure of merit outputs and the measured output, y_{meas}. y_{meas} gives the state of the plant with added measurement noise. Measurement noise is only observed by the controller and represents an imperfect estimate of the plant's true state.

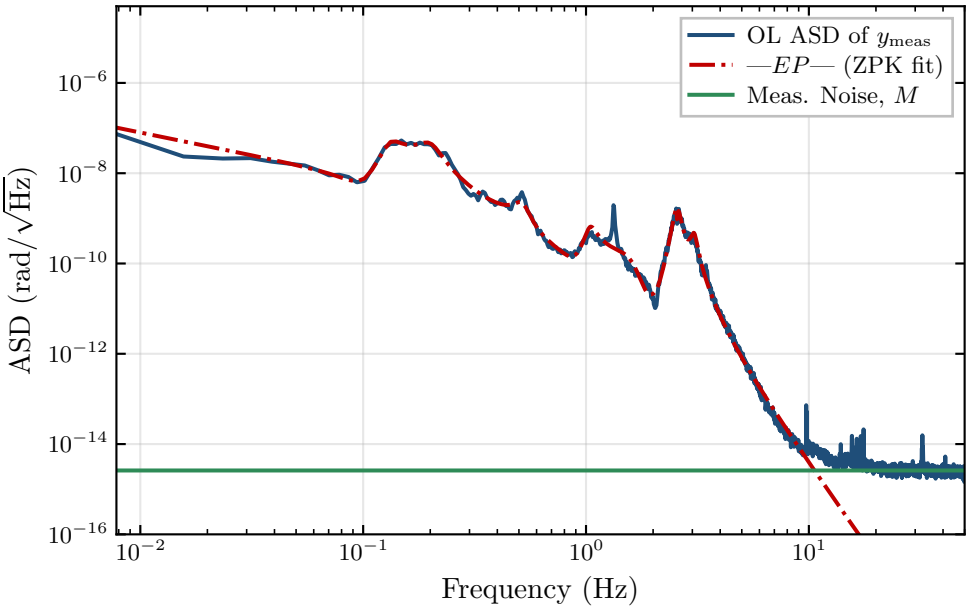


Figure 7.2: The noise spectrum of the DHARD yaw DOF at the LIGO Hanford Observatory decomposed into fits of measurement and environmental contributions. The spectra are fit to filters M and E, see relations of Eqs. 7.15 and 7.17. The measured ASD of y_{meas} is the measured spectrum of the cavity axis (see Fig. 7.1) and is equal to $|EP + M|$. This spectrum assumes no control, $K = 0$ (open-loop). After fitting a ZPK to $|EP|$, the contribution from E was isolated by dividing by the model for the plant (P). The filters are fitted using the Python package `wield.iirrational` to refine fits with gradient descent after making initial guesses using the AAA algorithm [140, 141]. Adapted from [100].

The system layout includes multiple blocks, each representing a transfer function. The plant, P, represents the physical system being simulated; in this case, it is the mechanism by which the actuator converts force/torque inputs into displacements/tilts. The measurement noise, M, is the shaping filter that turns the input Gaussian white noise into the appropriately shaped sensing noise, while the environmental noise, E, serves the same purpose for the state noise on the plant, both shown in Fig. 7.2. The transfer functions relate to Eqs. 7.11, 7.13 and 7.14 through

$$S_{\text{env}} = |E|^2 \, S_{n1}^u \tag{7.15}$$

and

$$S_{\text{meas}} = |M|^2 S^u_{n2}, \tag{7.16}$$

where

$$S^u_{n1}, S^u_{n2} \equiv 1/\sqrt{\text{Hz}} \tag{7.17}$$

and are uncorrelated unit white-noise PSDs. This indicates that these inputs are driven by Gaussian white noise that is then shaped by the relevant noise-shaping filters. The two FOM filters, F_{BNS} and F_{flat}, represent the transfer function that weights the frequency domain signal properly for two different measures of the controller's performance described in Section 7.2.2.

The two FOM outputs are free of any direct measurement noise, and they are each connected to a point representing the N_{p}, and N_{a} plant and actuator noise terms of Eqs. 7.11, 7.13 and 7.14. The frequency response of the plant must have two copies in this model to produce the different contributions of K (to be designed) in the expressions of N_{a} and N_{p}.

The FOM block includes a shaping filter for each path that weights the relative contribution of S_a or S_p in the optimization problem. The construction and weighting of the FOM filters is described in Section 7.2.2 and is fundamental to finding the Pareto front of optimal controllers for a given system. The absence of measurement noise indicates that the FOM outputs report the true value (up to weighting) of the DOF from which it is constructed and not what would be observed through an imperfect physical measurement.

For this alignment control application, the state of the plant at N_{p} represents the total angular motion of the mirror or cavity in its degree of freedom (differential, HARD, yaw). A flat filter representing the first FOM then has y_{FOM1}, which integrates into the total variance of the mirror motion. This is the typical objective that is minimized in LQG/$\mathcal{H}_2$ optimal control. The N_{a} connection to y_{FOM2} represents how the control actuation pushes on the mirror in pitch or yaw. If the beam is miscentered, the alignment actuation is converted into length noise. The BNS FOM then contains a weighting filter for how important this length noise conversion is for the SNR or range figure of merit.

7.2.2 Figures of Merit

As discussed in the previous sections, FOMs are measures of the performance of a closed-loop system. In the example presented here, an essential new complexity is

added: multiple FOMs.

In the model of the system, shown in Fig. 7.1, two FOMs are shown, which have been indicated by Eqs. 7.6 and 7.10. Shown inside the red dashed outline in Fig. 7.1 and represented by transfer functions, these two FOMs, F_{BNS} and F_{flat}, allow for different frequency weightings of the output signal. Figure 7.3 shows the magnitude of the two FOMs used in the augmented ASC system. The first broadband FOM, F_{flat}, is required to minimize $\tilde{R}_{\mathrm{p}}$ which is necessary to preserve the inequality in Eq. 7.10. The broadband FOM represents the unshaped total noise in the DOF seen after the plant, which affects the operational locking of the interferometer. This FOM is also necessary in order for the problem to be well-posed for optimal control, ensuring item 4 in Section 6.3.3.

The second FOM, F_{BNS}, measures how alignment-control actuation reduces the sky-averaged BNS detection range. For two 1.4 $M_{\odot}$ neutron stars, the range convention of Eq. 7.2 can be written as

$$d_{\mathrm{range}} = A_R \sqrt{\int_0^{\infty} \frac{f^{-7/3}}{S_h(f)}\,\mathrm{d}f}, \tag{7.18}$$

where $S_h(f)$ is the PSD of the total noise in the detector, and A_R is a constant that depends on the chirp mass of the binary system and the detection threshold. The constant A_R can be expressed as

$$A_R = \frac{2}{2.26\rho_0}\sqrt{\frac{5\pi}{24c^3}}\frac{(G\mathcal{M})^{5/6}}{\pi^{7/6}}, \tag{7.19}$$

where $\rho_0 = 8$ is the detection threshold, $\mathcal{M}$ is the chirp mass, and the factor 2.26 converts the optimal source distance to the sky-averaged range. For the present actuation-point objective, the detector-noise decomposition in Eq. 7.5 is specialized by setting $S_{\mathrm{c}} = S_{\mathrm{a}}$. For $|C|^2 S_{\mathrm{a}} \ll S_{\mathrm{det}}$, the first-order expansion

$$\frac{1}{S_{\mathrm{det}} + |C|^2 S_{\mathrm{a}}} \simeq \frac{1}{S_{\mathrm{det}}} - \frac{|C|^2 S_{\mathrm{a}}}{S_{\mathrm{det}}^2} \tag{7.20}$$

gives

$$d_{\mathrm{range}}^2 \simeq A_R^2 \int_0^{\infty} \frac{f^{-7/3}}{S_{\mathrm{det}}(f)}\,\mathrm{d}f - A_R^2 \int_0^{\infty} \frac{f^{-7/3}|C(f)|^2 S_{\mathrm{a}}(f)}{S_{\mathrm{det}}(f)^2}\,\mathrm{d}f. \tag{7.21}$$

The first term is d_{det}^2, the squared range set by the baseline detector noise without this controls-noise contribution. Applying $\sqrt{D-E} \simeq \sqrt{D} - E/(2\sqrt{D})$ to Eq. 7.21,

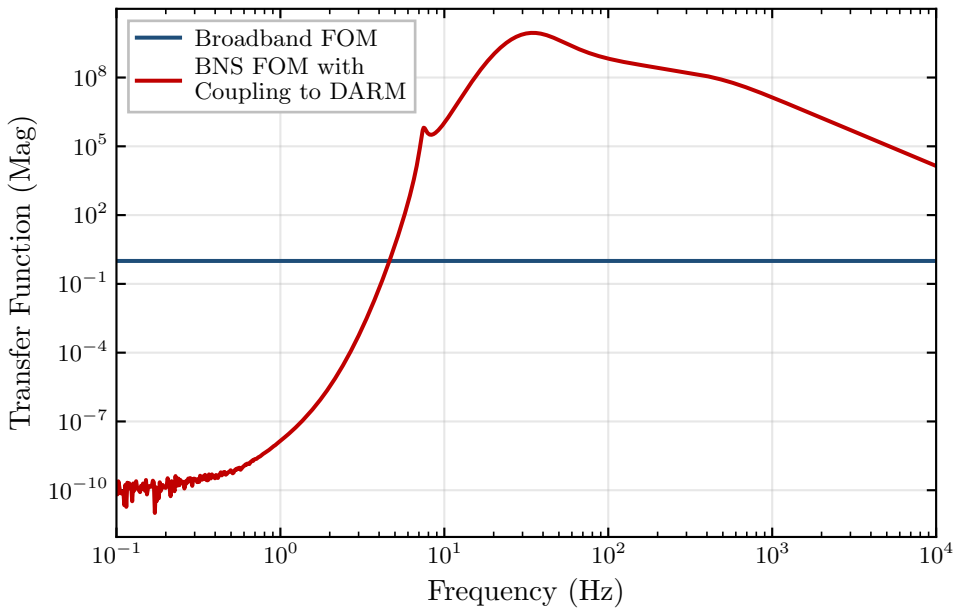


Figure 7.3: The magnitude of two FOM filters used in the augmented system, as shown in Fig. 7.1. Both of these FOMs shape either N_{p} or N_{a} to create metrics for the controller to minimize. The broadband FOM results in an unshaped signal that must be below a specified threshold to keep the interferometer locked. The BNS FOM shapes S_{a} allowing for the direct optimization of LIGO's BNS detection range. Adapted from [100].

where D and E denote the baseline and controls-noise terms, respectively, gives

$$d_{\mathrm{range}} \simeq d_{\mathrm{det}} - \frac{A_R^2}{2d_{\mathrm{det}}} \int_0^\infty \frac{f^{-7/3}|C(f)|^2 S_{\mathrm{a}}(f)}{S_{\mathrm{det}}(f)^2}\,\mathrm{d}f \equiv d_{\mathrm{det}} - d_{\mathrm{lost}}. \tag{7.22}$$

The $\mathcal{H}_2$ solver minimizes this integrated output PSD; equivalently, it evaluates the output RMS from the closed-loop Gramians using Eqs. 6.11 and 6.12. The BNS FOM is therefore a shaping filter, rather than the integral itself, with squared magnitude

$$|F_{\mathrm{BNS}}(f)|^2 = A_R^2 \frac{f^{-7/3}|C(f)|^2}{S_{\mathrm{det}}(f)^2}. \tag{7.23}$$

Connecting this filter to N_{a} makes the solver's integrated output variance equal to the numerator of the range-loss term. The physical range loss is obtained by dividing that variance by $2d_{\mathrm{det}}$. The filter in Fig. 7.3 spans more than 16 orders of magnitude, making the augmented system numerically ill-conditioned in double precision and motivating the numerical treatment described in the following sections.

The frequency-dependent coupling function $C(f)$ is measured through an excitation injected into the control loop [79]. When excited, S_{a} from Eq. 7.13 becomes

$$S_{\mathrm{a,\,exc}} = S_{\mathrm{a}} + \left|\frac{G}{1-G}\right|^2 S_{\mathrm{exc}}, \tag{7.24}$$

where S_{exc} is the PSD of the excitation signal summed in before the controller. When this excitation is added, S_h from Eq. 7.5 becomes

$$S_{h,\,\mathrm{exc}} = S_{\mathrm{det}} + |C|^2 \left(S_{\mathrm{a}} + \left|\frac{G}{1-G}\right|^2 S_{\mathrm{exc}}\right). \tag{7.25}$$

From Eqs. 7.24 and 7.25, $|C|^2$ can be calculated and fit from

$$|C|^2 = \frac{S_{h,\,\mathrm{exc}} - S_h}{S_{\mathrm{a,\,exc}} - S_{\mathrm{a}}}. \tag{7.26}$$

7.2.3 Pareto Front of Multiple FOMs for Bilinear Noise

To fully motivate the use of these two FOMs in the bilinear noise case, Eq. 7.6 can be rewritten to directly depend on both FOMs. Consider the noise coupling factor C from Eq. 7.8. For the multiplication effect that occurs for bilinear noise on a single DOF, a model is assumed where $C \propto \tilde{R}_{\mathrm{p}}$ represents an intermodulation effect. The effect of an intermodulation can be seen by rewriting Eq. 7.4 in the time domain, with the control noise term becoming bilinear

$$h_{\mathrm{signal}}(t) = h_{\mathrm{gw}}(t) + N_{\mathrm{det}}(t) + \chi N_{\mathrm{a}}(t) N_{\mathrm{p}}(t), \tag{7.27}$$

where χ is a scale factor for the cross-coupling. For the alignment example considered here, N_{p} is the residual angular motion of the mirror, while N_{a} is the actuated tilt. If the residual angular motion is non-zero, then the applied tilt causes a coupling to arm length.

When the N_{p} noise is dominated by low-frequency components, this term becomes $\chi \tilde{R}_{\mathrm{p}} N_{\mathrm{a}}(t)$ by linearizing. See Section 7.3 for a formal derivation. This substitution can then be used in Eq. 7.8 via $C(f) = \chi(f) \cdot \tilde{R}_{\mathrm{p}}$. In the alignment control example problem, this relation for constructing $C(f)$ yields the coupling factor called "A2L" for alignment-to-length. This factor is measured [79] and translates angular motion in the alignment system into length motion in the DARM gravitational wave channel. This form for C and similar models of RMS-dependent coupling factors indicates that one aspect of the range depends on the RMS as computed from the flat FOM, while the other depends on the RMS as computed from the shaped BNS FOM.

Assuming this form of the C coupling function, the SNR loss of Eq. 7.8 becomes

$$\mathrm{SNR}^2_{\mathrm{lost}} = 4 \cdot \tilde{R}^2_{\mathrm{p}} \cdot \int_0^\infty \frac{|\chi|^2 \left|h_{\mathrm{gw}}(f)\right|^2}{S_{\mathrm{det}}(f)^2} S_{\mathrm{c}}(f)\, \mathrm{d}f \tag{7.28}$$

$$= \underbrace{\int_0^\infty |F_{\mathrm{flat}}|^2 S_{\mathrm{p}}\, df}_{\tilde{R}^2_{\mathrm{flat}}} \cdot \underbrace{\int_0^\infty |F_{\mathrm{BNS}}|^2 S_{\mathrm{c}}\, \mathrm{d}f}_{\tilde{R}^2_{\mathrm{BNS}}/\overline{R}^2_{\mathrm{flat}}}, \tag{7.29}$$

where $\tilde{R}^2_{\mathrm{BNS}}$ denotes the RMS of the noise that comes out of the BNS FOM, which is output y_{FOM2} in Fig. 7.1, $\tilde{R}^2_{\mathrm{flat}}$ is the RMS of the total plant motion during control, and $\overline{R}^2_{\mathrm{flat}}$ is a constant for RMS at the time of the measurement of the A2L coupling factor $C(f)$.

As a minimization problem, a necessary condition for optimality can be established by taking the derivative with respect to the controller parameters $\mathbf{K}$, in bold to represent parameters in a state-space matrix form. Setting these derivatives to zero finds an extremal point of the noise as

$$0 = \frac{1}{4|\chi|^2} \frac{\mathrm{d}}{\mathrm{d}\mathbf{K}} \mathrm{SNR}^2_{\mathrm{lost}} \tag{7.30}$$

$$= \tilde{R}^2_{\mathrm{flat}} \frac{\mathrm{d}}{\mathrm{d}\mathbf{K}} \tilde{R}^2_{\mathrm{BNS}} + \tilde{R}^2_{\mathrm{BNS}} \frac{\mathrm{d}}{\mathrm{d}\mathbf{K}} \tilde{R}^2_{\mathrm{flat}}. \tag{7.31}$$

This relation can be satisfied through the equivalence

$$0 = \frac{\mathrm{d}}{\mathrm{d}\mathbf{K}} \tilde{R}^2_{\mathrm{syn}}, \tag{7.32}$$

when

$$\zeta \equiv \frac{\tilde{R}_{\mathrm{BNS}}}{\tilde{R}_{\mathrm{flat}}}. \tag{7.33}$$

The LQG and mixed-sensitivity approaches below find control systems that meet this necessary condition. By scanning over the weight parameter ζ, the Pareto front of optimal controllers is produced. The above derivation indicates that the bilinear-optimal controller exists along this Pareto front.

There are multiple reasons why the parameter ζ should be abstracted as an independent term for Eq. 7.34. The main reason is that, prior to optimization, the ratio of the RMS values of the two FOMs cannot be known. Thus, there is no presumptive ζ to target. Since LQG can only optimize over a single parameter, all of a system's FOMs must be combined. With ζ, a synthetic weighted RMS of the two FOMs can

be defined, which meaningfully relates to the above condition on the derivative of the SNR as

$$\tilde{R}^2_{\text{syn}} = \zeta^2 \tilde{R}^2_{F_{\text{BNS}}} + \tilde{R}^2_{F_{\text{flat}}}. \tag{7.34}$$

The weighted sum of the two FOM outputs creates a single synthetic output with its own RMS, $\tilde{R}_{\text{syn}}$. The problem now boils down to a problem to minimize $\tilde{R}_{\text{syn}}$ over many given weightings of ζ. It is advantageous to only weight one of the FOMs to achieve the desired relative weighting because leaving the weighting of the flat FOM untouched allows for consistent $\mathcal{H}_\infty$ norms across different weightings, which will be important in Section 7.5.1.

For values of ζ that are zero, or very large, the optimization of $\tilde{R}_{\text{syn}}$ will find the minimum RMS in one of the respective FOMs. Because the RMS values satisfy a minimization constraint, the optimized values of both $\tilde{R}_{\text{BNS}}$ and $\tilde{R}_{\text{flat}}$ are expected to be monotonic but not necessarily smooth or convex in ζ.

Together, this suggests that scanning ζ will allow one to find the global minimum of the lost SNR, but does not establish that the control-optimal SNR loss function is a convex optimization problem in ζ. For any more complex dependence of C on χ or $\tilde{R}_c$, the derivative argument above is expected to still apply, but the relationship of Eq. 7.32 will become more complex. Finally, the above derivation and example based on the C coupling parameter assume only a single DOF. More elaborate models of the C parameter are likely dependent on the RMS of multiple DOFs. Thus, the decision point of where to balance flat RMS versus the BNS-weighted RMS is, in general, more complex. These are major reasons that ζ should be treated abstractly to produce a Pareto front. The following plots show that the ζ parametrization clearly shows the boundary of the optimal space for the two major RMS contributions. Thus, scanning ζ to find a boundary that shows the specific tradeoff between flat-weighted RMS and the BNS-weighted RMS is the most generally useful expression and is achieved through the ζ parameter. The above argument is expected to apply to MIMO systems using a vector of ζ scales, but this is left for future work.

7.3 The Bilinear Noise Metric and Spectral Convolution

This section shows how a multiplicative, bilinear coupling affects the spectrum of a signal. The main result is that multiplication in the time domain produces convolution in the frequency domain. Start with two independent noise channels

N_A and N_B. Each has a spectral density defined as

$$\langle N_A(f)N_A(-f-f')\rangle \equiv \frac{1}{2}S_A(f)\delta(f') \tag{7.35}$$

and

$$\langle N_B(f)N_B(-f-f')\rangle \equiv \frac{1}{2}S_B(f)\delta(f'), \tag{7.36}$$

where their independence gives the property $\langle N_A(f)N_B(f')\rangle = 0$.

Now assume that a channel $N_x(t) \equiv N_A(t)N_B(t)$ is present. By the convolution theorem,

$$N_x(f) = N_A(f) \circledast N_B(f) \equiv \int_{-\infty}^{\infty} N_A(f')N_B(f-f')\mathrm{d}f'. \tag{7.37}$$

The spectrum of this bilinear product can now be derived through

$$\begin{aligned}
\frac{1}{2}S_X(f)\delta(f') &\equiv \langle N_X(f)N_X(-f-f')\rangle && (7.38)\\
&= \left\langle \begin{array}{l} \displaystyle\int_{-\infty}^{\infty} N_A(f'')N_B(f-f'')\mathrm{d}f'' \\ \quad \times \displaystyle\int_{-\infty}^{\infty} N_A(f''')N_B(-f-f'-f''')\mathrm{d}f''' \end{array} \right\rangle && (7.39)\\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \underbrace{\langle N_A(f'')N_A(f''')\rangle}_{\frac{1}{2}S_A(f'')\delta(f''+f''')} \\
&\quad \times \underbrace{\langle N_B(f-f'')N_B(-f-f'-f''')\rangle}_{\frac{1}{2}S_B(f-f'')\delta(f'+f''+f''')}\mathrm{d}f''\mathrm{d}f''' && (7.40)\\
&= \frac{1}{4}\delta(f')\int_{-\infty}^{\infty} S_A(f'')S_B(f-f'')\mathrm{d}f''. && (7.41)
\end{aligned}$$

Thus, for this type of bilinear noise,

$$S_X(f) = \frac{1}{2}S_A(f) \circledast S_B(f). \tag{7.42}$$

The derivation is now applied to the bilinear noise term, $N_X = \chi N_\mathrm{p}(t)N_\mathrm{a}(t)$ seen in Eq. 7.27.

It is first assumed that the RMS noise power of S_p is concentrated at low frequencies, i.e.,

$$S_\mathrm{p}(f) \approx \tilde{R}_\mathrm{p}\delta(f+\epsilon), \tag{7.43}$$

for some small ϵ.

Applying the convolution from this derivation then gives:

$$S_X(f) = \chi\frac{\tilde{R}_\mathrm{p}}{2}\left(S_\mathrm{a}(f+\epsilon) + S_\mathrm{a}(f-\epsilon)\right) = \chi\tilde{R}_\mathrm{p}S_\mathrm{a}(f). \tag{7.44}$$

This final result indicates that $|C| = |\chi|\tilde{R}_\mathrm{p}$ as used in Section 7.2.2.

7.4 LQG Optimization

The LQG problem is to find an optimal controller that yields the lowest $\tilde{R}_\mathrm{syn}$ for a given ζ. This is equivalent to finding a controller that minimizes the $\mathcal{H}_2$ norm of a *transfer-function-matrix*, $\mathbf{T}_2(f)$, from the u_{n1}, u_{n2} vector of inputs, to the $y_\mathrm{FOM1}, y_\mathrm{FOM2}$ vector of outputs, with a ζ scaling first applied to the y_FOM2 output. Here $\mathbf{T}$ is a set of transfer functions that define the system from Fig. 7.1. A mathematical description of this system using state space systems and a more rigorous definition of $\mathbf{T}$ can be found in Section 6.3.1.

By contrast, LQG is often taught and presented starting from the separation principle, combining a Kalman filter as a noise-optimal state linear-quadratic estimator with a linear quadratic regulator that minimizes an actuator cost function. The actuator cost is often described as fuel consumption, but when using the FOM weighting filters, it can also be interpreted as part of a noise-minimization procedure. This interpretation of LQG as an $\mathcal{H}_2$ noise minimization procedure can be established intrinsically through minimization [106], as well as through the case analysis of the separation principle [101].

The $\mathcal{H}_2$ optimization can prove difficult by hand or even with a gradient-descent numerical optimizer because of design trade-offs. Finding the perfect balance between suppressing environmental noise and introducing minimal measurement noise, all while weighting the output by two FOMs, is a job that optimal control can perform in a single algebraic solution step. Here, LQG is used to find the optimal controller to give the lowest possible $\tilde{R}_\mathrm{syn}$. The control transfer function, K, is determined by the state-space description of the control, $\mathbf{K}_\mathrm{LQG}$, using the method described in Section 6.3.3. The method of the numerical solver is not essential to understanding the results presented here, but an important point is that LQG optimization is

non-iterative. This is essential for speed, as the method requires a scan over the ζ parameter that indicates the tradeoff between two FOMs. Because of dynamic range and numerical challenges, an understanding of the numerical solver has been essential for implementing the study with a reliable software implementation, but ideally, it should not be required for end users.

7.4.1 LQG Optimal Controllers

Using the plant system, whose layout is presented in Fig. 7.1, a set of optimal LQG controller filters was calculated using the method described in Section 6.3.3. These controllers, corresponding to the filter, K, in Fig. 7.1, were calculated for different weights of the two FOMs corresponding to the BNS range and total RMS for a range of ζ values. The change in the weighting ζ causes the BNS FOM to increase its contribution to the $\tilde{R}_{\text{syn}}$. By calculating a number of the ζ values, it is possible to show the performance of those controllers in relation to the LIGO's early observing run O4 DHARDY open-loop gain with respect to any tradeoff between total noise and lost performance (observing range). Since the solver is always finding the lowest noise across different weightings, the results form a Pareto front, showing the boundary of all optimal linear controllers for this cavity-alignment degree of freedom.

Fig. 7.4 shows a comparison between several LQG optimal controllers for several ζ values. To calculate the controllers, a system scaled by ζ is used. During the solve, if ζ is too large, the solver will either fail or create sub-optimal controllers due to numerical error in eigenvalue computation and loss of the ability to order deflating subspaces. Once the controller is calculated, it is connected to an unscaled system, where $\zeta = 1$. The unscaled system models the controllers' performance, with a Lyapunov equation used to compute the lost BNS range from the RMS of the BNS FOM output, as described in Section 7.2.2. The two performance metrics are then plotted. Lower is better, representing either suppressed noise in the alignment plant or less loss of detection range. The LQG open-loop gain of the controllers is shown in comparison to a hand-tuned LIGO loop design for DHARD yaw used during LIGO's fourth observing run.

The open-loop gain of the plant and controller system for the same LQG controllers is shown in Fig. 7.5. There are several features and similarities demonstrated in both the optimal LQG controllers and the LIGO hand-tuned controller.

The closed-loop gain of $G/(1 - G)$, shown in Fig. 7.6, is the filter through which

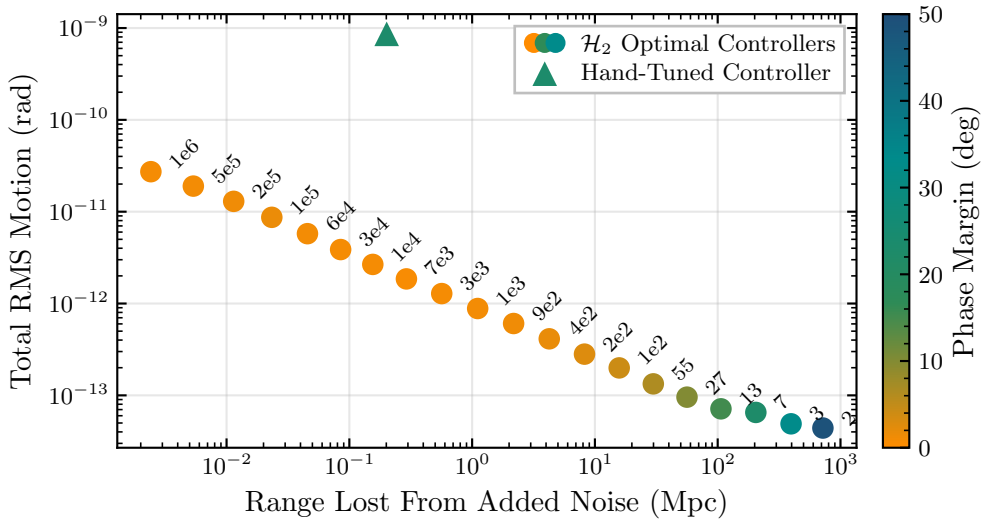


Figure 7.4: The RMS noise at the two FOM outputs given the white noise disturbances for the closed-loop system, including the system shown in Fig. 7.1 and optimal LQG controllers calculated for specified ζ values. Each point represents the performance of a single controller that is the LQG optimal solution for the plant with a different FOM weighting. The RMS value is shown on the axis of the corresponding FOM. The BNS FOM's RMS is calibrated to represent BNS range lost by the noise from a given controller, while the flat FOM represents the RMS noise in radians for ASC DHARD yaw. The number next to each point represents the ζ value used to acquire the desired weighting. The color of each point corresponds to the phase margin of the controller. The triangle represents the current LIGO hand-tuned controller's performance and phase margin. All of the LQG optimal controllers whose performance beats the hand-tuned controller in both FOMs have poor phase margins, and they get worse as the BNS range FOM is emphasized using a higher ζ factor. The margins can be improved by the LQG optimal solutions with an $\mathcal{H}_\infty$ performance bound discussed in Section 7.5 and presented in Fig. 7.7. Adapted from [100].

the noise S_{meas}, is seen in S_a and S_h as described by Eq. 7.11. When the peak of $G/(1-G)$ is high, the stability margins are poor, which is visible in Fig. 7.6. This suggests that if one can constrain the height, then one can enforce more acceptable margins.

The stability margins are computed for the open-loop gain, G, of the LQG controllers. The phase margin, which is related to the peak of $G/(1-G)$, is presented as the color of the points in Fig. 7.4. All of the gain and phase stability margins for the LQG controllers are unacceptably low, below 10 degrees, where 30 degrees is usually taken to be a lower limit. LIGO's control systems must be robust and remain stable when dealing with drifting operating conditions of the interferometer.

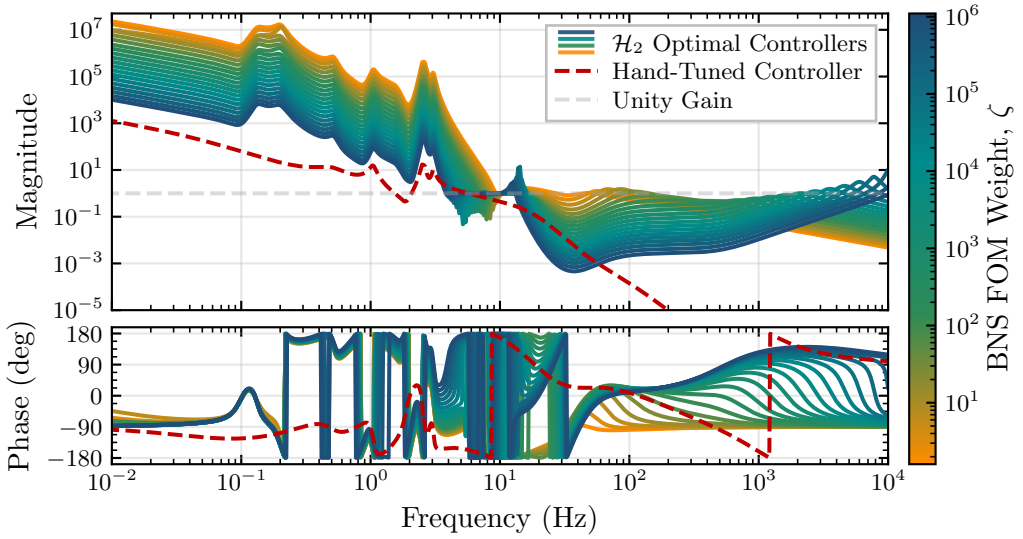


Figure 7.5: The open-loop gain, G, of the unscaled plant and controller system shown for different values of ζ. As ζ gets lower, the optimizer gives less weight to the $\tilde{R}_{\mathrm{BNS}}$. A hand-designed LIGO DHARD yaw controller is included for comparison. Each of these LQG controllers was optimized using an LQG solver with no enforced stability margins. The hand-tuned controller meets the stability requirements necessary for implementation in the interferometer. Adapted from [100].

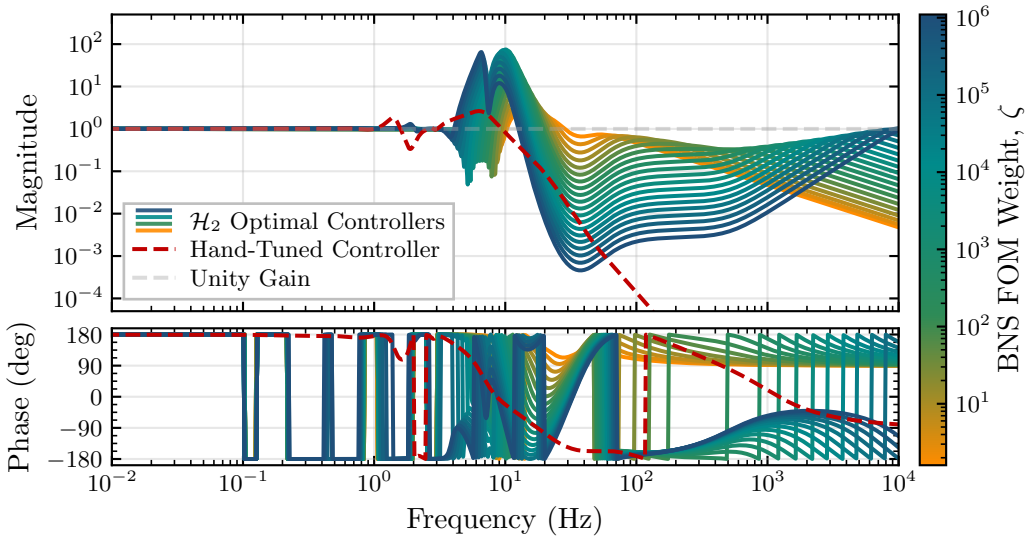


Figure 7.6: The closed-loop gain, $G/(1-G)$, of the plant and controller system shown for different values of the relative FOM weighting parameter, ζ. These are the same LQG controllers presented in Figs. 7.5 and 7.7. The current hand-tuned LIGO DHARD yaw controller is included for comparison. The large peaks seen around 10 Hz are gain peaking, which causes low stability margins by inflating the $\mathcal{H}_\infty$ norm of the closed-loop system. This is further discussed in Section 7.5. Adapted from [100].

If the gain of the optical or mechanical components of the interferometer plant is increased or decreased by a small margin, then the gain of G will also be affected. If the margins are smaller than the change in the gain or phase of the plant, then the system will become unstable.

7.4.2 Properties and Limitations Apparent in LQG Controllers

All of the calculated controllers aim to reduce the noise of the current hand-tuned LIGO DHARD yaw controller. Figures 7.4 and 7.6 show the RMS, open, and closed-loop transfer functions for the LQG controllers and for the hand-tuned LIGO controller. The properties exhibited by the LQG controllers should be related to the hand-tuned controller, but might also inform future design.

The pieces of information that motivate the hand-tuning of the controllers are generally the RMS of S_p, the open, and closed-loop transfer function to see the stability margins, as well as the noise spectrum presented in Fig. 7.2. The hand-tuned controller has an integrator under 1 Hz to suppress the most significant contribution to the total RMS: tilt noise. The unity-gain frequency of the open-loop gain is below 10 Hz, and the hand-tuned controller rolls off rapidly above 10 Hz to decrease the impact on the observing range. The roll-off enforces a hand-tuned version of the F_{BNS} weighting, ζ, but costs a lot of phase at the unity-gain frequency. The hand-tuned controllers sometimes are designed with multiple unity-gain frequency crossings to target noise peaks where needed, but these are challenging to keep stable, so hand design often just suppresses the lowest-frequency noise.

The LQG controllers have a much more aggressive open-loop gain. There is a strong roll-off after 10 Hz to minimize the impact on the BNS range, meaning the rate of the roll-off is highly dependent on ζ. When ζ is very low and thus the F_{BNS} does not contribute much to the total RMS of $\tilde{R}_{\mathrm{syn}}$, the roll-off is almost nonexistent. The limit on when the roll-off becomes non-negligible is when $\zeta\tilde{R}_{\mathrm{BNS}}$ becomes larger than the residual environmental noise $\tilde{R}_{\mathrm{p}}$. One of the most noticeable characteristics of the LQG controllers is that they are aggressive. The closed-loop gain, shown in Fig. 7.6, shows that the unchecked gain peaking can get to about 70 times its DC gain, which corresponds to an unacceptable phase margin of around a few degrees.

7.5 Solution With Custom Stability Margins

While the goal of many of LIGO's control loops is to minimize the $\mathcal{H}_2$ norm of the closed-loop system, optimally, for bilinear noise, the controllers must also be robust and stable. When hand-designing control loops, the measures of this robustness are

the gain and phase margins. The gain margin is the smallest gain shift that makes the system unstable, and the phase margin is the smallest amount of phase lag that causes the system to become unstable.

The pure LQG solutions for the system, presented in Section 7.4.1, show unacceptable stability margins and gain peaking. The best controller for many of LIGO's applications would be an LQG controller if there were a way to guarantee its stability margins. Unfortunately, to quote the entire abstract of J. Doyle's paper on guaranteed margins for LQG controllers, "There are none" [142].

A more complicated strategy must instead be employed. The phase margin for the system is directly linked to the peak of the closed-loop gain. This peak can be measured using an $\mathcal{H}_\infty$ norm and can be bounded. The following work mixes such a bound within the LQG solution and presents robust and noise-optimal controllers for the example alignment (ASC) system, and can be used for many other LIGO control systems.

7.5.1 $\mathcal{H}_\infty$ Optimization Problems

Before describing the mixed-sensitivity approach, it is worth first describing the pure $\mathcal{H}_\infty$ optimization approach and contrasting it to the $\mathcal{H}_2$ minimization performed by LQG. In changing an LQG problem to utilize an $\mathcal{H}_\infty$-optimal solver, the same transfer function $\mathbf{T}_2(f)$ of noise inputs to FOM outputs can be used. The optimization objective is then to find the controller that minimizes the maximum value of $\|\mathbf{T}_2(f)\|_2$ over all frequencies f (cf. Section 6.3.6). $\mathcal{H}_\infty$ optimization is typically claimed to find controllers more robust to model changes or adversarial (optimally bad) inputs; however, it requires its own precise and uniquely defined FOMs to be applicable or relatable to astrophysics metrics of gravitational-wave detectors. In particular, $\mathcal{H}_\infty$ makes no guarantees about minimizing RMS noise levels and is thus fundamentally unrelated to SNR and detection range metrics.

In effect, when using noise shaping filters and FOM weights, a simple reinterpretation of an $\mathcal{H}_2$ model into $\mathcal{H}_\infty$ neither improves robustness nor helps with the true noise optimization problem. One must instead use a different set of weights and shapes to define a $\mathbf{T}_\infty$ transfer function matrix that is dedicated to the $\mathcal{H}_\infty$ optimization. The following sections show the mixed-sensitivity approach that retains the $\mathbf{T}_2(f)$ of the $\mathcal{H}_2$ optimization while including a dedicated weighting to apply an $\mathcal{H}_\infty$ bound. In the following, $\mathbf{T}_\infty$ is implemented by utilizing the additional input u_∞ of Fig. 7.1.

What, then, is the difference between an $\mathcal{H}_\infty$ optimization and a bound? Both applications of $\mathcal{H}_\infty$ use an additional single parameter, γ, to apply the bounded real lemma [115] and produce algebraic Riccati equations that only have an admissible solution when $\|\mathbf{T}_\infty(f)\|_2 < \gamma$ for all f. The $\mathcal{H}_\infty$ optimization problem is to find the smallest γ, and associated controller, such that these equations hold. In finding the minimal admissible γ, there is no remaining freedom in the controller to also optimize an LQG $\mathcal{H}_2$ problem and minimize noise RMS due to the uniqueness of the optimal $\mathcal{H}_\infty$ controller. When a γ above the minimum value is chosen, there is a large class of $\mathcal{H}_\infty$-sub-optimal controllers [120], all related through γ-bounded linear fractional transformations [115], that can be considered as the admissible set for an $\mathcal{H}_2$ noise optimization problem. In this way, it is possible to find a solution with a minimal $\mathcal{H}_2$ norm within the set of $\mathcal{H}_\infty$-sub-optimal controllers.

7.5.2 LQG Optimal Solutions Given an $\mathcal{H}_\infty$ Performance Bound

As discussed in Section 7.5.1, placing a limit on a suitably defined $\mathcal{H}_\infty$ norm can enforce stability margins of a system. However, $\mathcal{H}_\infty$ solvers do not return controllers with acceptable noise. The ideal controller for most of LIGO's control loops would be the optimal LQG filter that has an $\mathcal{H}_\infty$ norm enforcing robustness. Here, a bound on the phase margin is derived, and results from the alignment control model are shown. Section 6.5 describes the numerical method for calculating the optimal solution [123, 124].

The phase margin, ϕ_m, of the system is expressed as the phase when the magnitude crosses the unity-gain frequency, or expressed as an equation,

$$\phi_m = \angle G(f_\mathrm{u}), \tag{7.45}$$

where f_u is the unity-gain frequency. As the unity-gain frequency is approached

$$\lim_{f \to f_\mathrm{u}} |G(f)| = 1, \tag{7.46}$$

meaning, at f_u, the closed-loop gain is at or near its maximum. If a bound is added to the closed-loop gain and the phase margin is considered in radians, then

$$\left| \frac{G(f)}{1 - G(f)} \right| \leq \gamma \tag{7.47}$$

and

$$\lim_{f \to f_\mathrm{u}} \left| \frac{G(f)}{1 - G(f)} \right| \approx \frac{1}{|\phi_m|}, \tag{7.48}$$

yielding

$$|\phi_m| \geq \frac{1}{\gamma}. \tag{7.49}$$

By limiting how large γ can be through the controller, the phase margin is bounded from below. For example, a $\gamma \approx 1.27$ limit enforces approximately a $\pi/4 = 45$ degree phase margin.

To implement this, the additional u_∞ input of Fig. 7.1 is employed. By using the transfer function matrix from the input u_∞ to the y_{FOM1} output to define $\mathbf{T}_\infty(f)$, the ideal $\mathcal{H}_\infty$-bound can be established in the definitions above.

The results, presented in Figs. 7.7 to 7.9, show a number of controllers of varying ζ and γ values. Shifting these two parameters allows a group of controllers to be calculated. A single preferred controller can then be post-selected from the group to meet the desired outcome. These plots look similar to those in Section 7.4, but include a varying γ bound as a parameter.

The chosen configuration of $\mathbf{T}_\infty(f)$ to measure the closed-loop gain has zero norm for the trivial no-control solution, $G = 0$. In practice, no useful control loops that influence or optimize the RMS can be created as $\gamma \to 1$. Numerically, the solver becomes increasingly fragile in this limit. The procedure is to first choose a moderately large γ value, and then solve for a controller. If this solution is successful, then a smaller value is chosen and the solve is repeated. γ is successively reduced towards $\gamma = 1$ until reliable convergence can no longer be achieved. For this ASC example, phase margins of around 45 degrees are achieved.

The controllers' RMS performance, presented in Fig. 7.7, shows that the current LIGO controller for DHARD yaw can be improved in both the BNS range as well as the overall RMS in the loop while keeping the same phase margin. In the lost range due to the added noise, an order-of-magnitude improvement can be made. The total RMS can only be improved by about four times for controllers that would be implemented while maintaining the controller's stability.

Each group of controllers in Fig. 7.7 that is connected by a light dotted line represents controllers with the same ζ, displayed at the top of each branch, for a varying γ. The phase margin is shown by the color of each point. It is clear that as the phase margin increases, so does the RMS noise. In effect, every choice of phase margin produces a Pareto front.

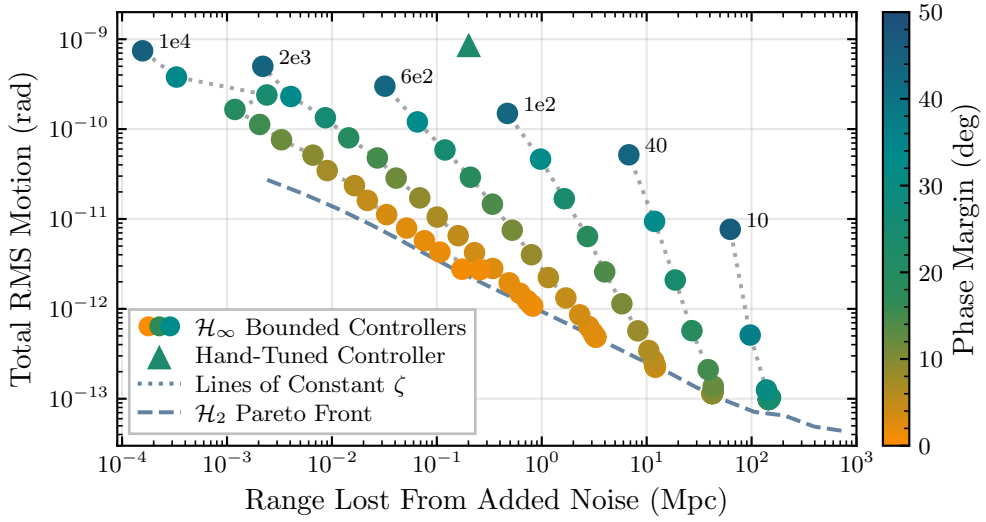


Figure 7.7: The RMS motion of the optic in radians and the range lost from added noise, which are the outputs from the two FOMs. The plot shows the phase margin of all the controllers as a color-coded representation. The phase margin is related to the gain margin monotonically, and their relation as well as their relation to γ is presented in Fig. 7.10. Each branch of the plot, shown by the light dotted line, has the same ζ value with different γ values. The number at the top of each branch is the ζ value for that branch. The dark dashed line shows the $\mathcal{H}_2$ Pareto front, which is the limit to the controller's noise performance, as it represents the performance of the LQG controllers from Fig. 7.4, which are RMS optimal. Adapted from [100].

The open-loop gain of a single branch of these constant ζ controllers is presented in Fig. 7.8 and compared to the current LIGO DHARD yaw controller. By plotting only one ζ value, it is possible to show the effect that γ has on the controllers. As γ is decreased, the controller becomes more robust, and the overall amplitude decreases, especially at low frequencies, making the controllers less aggressive. This is also true for the other controllers at other ζ values. However, at around 7 Hz, the response stays almost constant for the varying γ values. The fact that the controller is trying to stay at or above unity-gain shows that it is not merely pushing down the gain everywhere but actually trying to maintain the noise performance while improving stability. In comparison to the non-$\mathcal{H}_\infty$ norm limited controllers in Fig. 7.5, these are much less aggressive at low frequencies. Between 0.1 Hz and 1 Hz, the gain of the optimal controllers is still sometimes significantly larger than the current LIGO controller. In Fig. 7.2, the environmental noise shaped by the plant is shown to peak around 0.15 Hz. The controller's aggressiveness is trying to suppress the environmental noise as seen at the plant output.

The closed-loop transfer function of $G/(1-G)$ for the system is shown in Fig. 7.9.

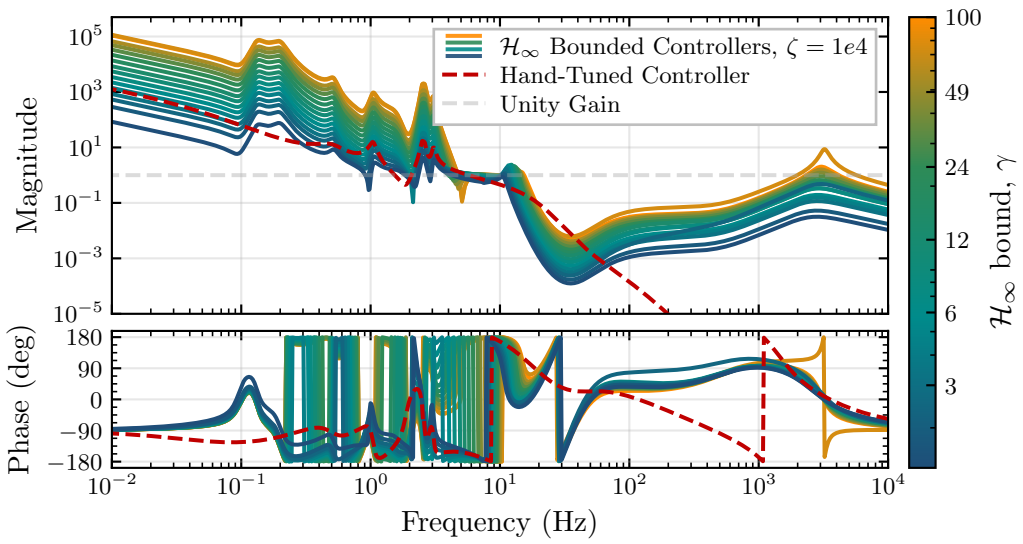


Figure 7.8: The open-loop gain, G, of controllers from Fig. 7.7 with a constant ζ value but with varying γ values. As γ decreases, the robustness of the controller, as measured by the stability margins, increases. These calculated controllers are compared to the current LIGO controller for DHARD yaw. Adapted from [100].

The $\mathcal{H}_\infty$ norm for each closed-loop system is the maximum of the closed-loop magnitude. Since the $\mathcal{H}_\infty$ norm must always be less than γ, as γ is decreased, the peak of the magnitude of the closed-loop transfer function is forced down. As shown in Eq. 7.49, this sets a limit on the phase margin.

While the phase margin is directly related to γ, the gain margin is an important element of the controller's stability. Figure 7.10 shows that the gain margin and phase margin are monotonic. By putting a bound on γ, it is possible to also create a controller with both acceptable gain and phase margins.

7.5.3 Properties and Limitations of $\mathcal{H}_\infty$ Norm Limited LQG Controllers

The first and most important change from the pure LQG controllers to the controllers presented in Fig. 7.7, is the improved phase margin. At the lowest γ values, the phase margin can beat the margins of the current hand-tuned controller.

As seen in the closed-loop transfer function, the $\mathcal{H}_\infty$ norm is being suppressed as the γ value decreases, enforcing an upper limit. The limiting of the $\mathcal{H}_\infty$ norm is not seen in the plot of the closed-loop gain of the pure LQG controllers in Fig. 7.6, since there is no $\mathcal{H}_\infty$ boundary enforced. The enforcement of the limit directly leads to more robust controllers as measured by the gain margin and phase margin.

While the LQG results in Section 7.4.1 show controllers that do not show all of

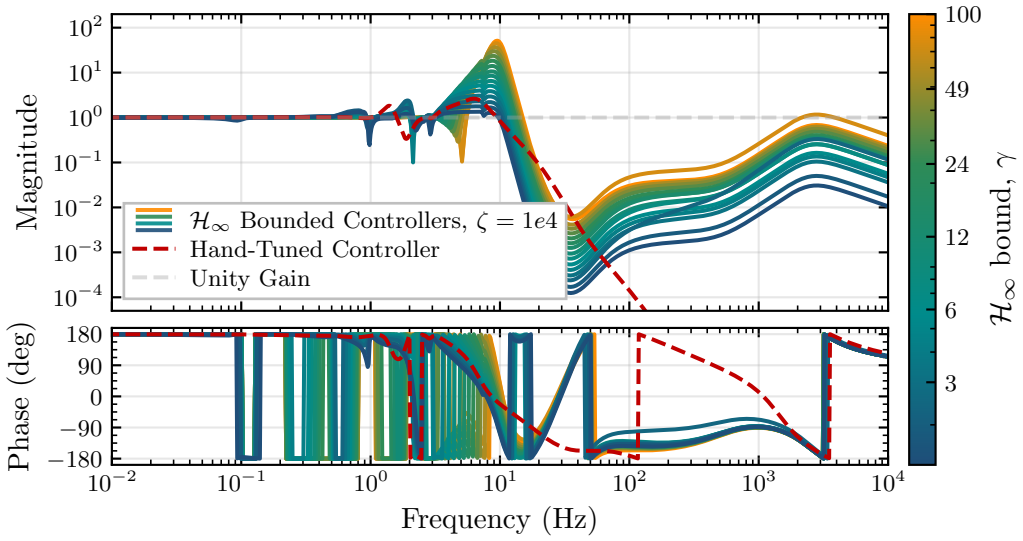


Figure 7.9: The closed-loop gain, $G/(1-G)$, of controllers from Fig. 7.7 with a constant ζ value but with varying γ values. As γ decreases, the robustness of the controllers increases. The phase margin of the controller is directly linked to the height of the maximum peak of the closed-loop gain. As γ decreases, this peak is suppressed, as shown here, and thus, the stability margins increase. This relation is shown in Fig. 7.10. These calculated controllers are compared to the closed-loop gain of the current LIGO controller for DHARD yaw. Adapted from [100].

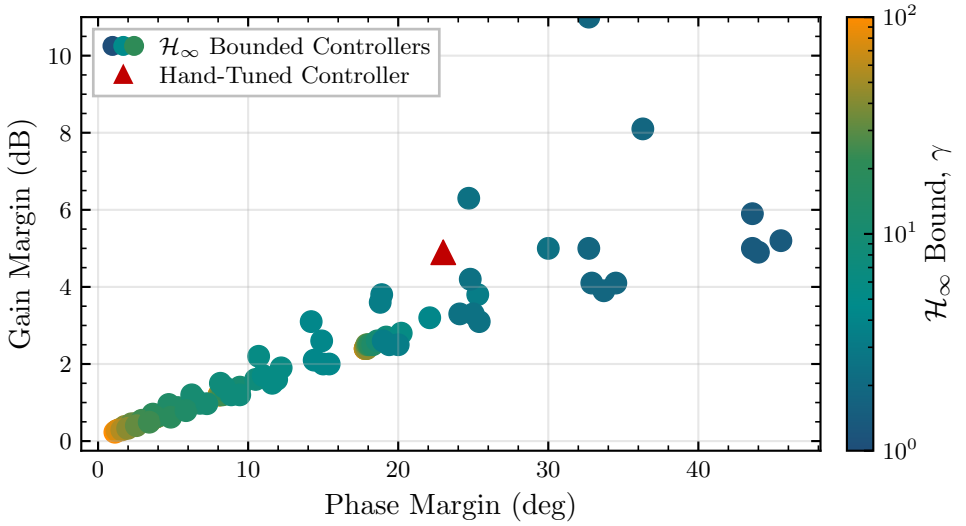


Figure 7.10: The gain margin, phase margin, and γ for the LQG controllers with $\mathcal{H}_\infty$ performance bounds displayed in Fig. 7.7. The phase margin, gain margin, and γ are monotonic. The stability margins are also shown for the current LIGO controller for DHARD yaw, which has no gamma value since it was hand-tuned and not created by an $\mathcal{H}_\infty$ minimization algorithm. Adapted from [100].

the same features of the current hand-tuned LIGO controller, the $\mathcal{H}_\infty$ bounded controllers show many of the same features. In the open-loop transfer function

(Fig. 7.8), the roll-off of the hand-tuned and $\mathcal{H}_\infty$ bounded controllers is very similar in frequency and magnitude. This similarity appears even though $\mathcal{H}_\infty$ bounded controllers' RMS is calculated from N_a rather than u_c, as is often used in LIGO's controller optimization. This similarity may suggest that this distinction is not as important. At very low frequencies, the $\mathcal{H}_\infty$ bounded controllers are actually slightly less aggressive than the hand-tuned controller. As γ decreases, the $\mathcal{H}_\infty$ bounded controllers get less aggressive.

A notable difference between the current hand-tuned LIGO controller and the calculated $\mathcal{H}_\infty$ bounded controllers is their behavior around unity gain ($G = 1$). The hand-tuned controller causes many unity-gain crossings in the open-loop gain as the gain peaks over the line to target noise peaks. Only the $\mathcal{H}_\infty$ bounded controllers with the lowest γ values exhibit this behavior.

It is important to note that the solver employed does not exactly implement the gain-peaking constraints that enforce phase margin. The implementation is slightly more restrictive. Current numerical solution methods for the mixed-sensitivity problem cannot fully independently define $\mathbf{T}_2$ and $\mathbf{T}_\infty$. Therefore, the method also has to include the y_{FOM2} output, which instead implements the constraint

$$\left|\frac{G(f)}{1-G(f)}\right|^2 \leq \frac{\gamma^2}{|F_{\text{flat}}(f)|^2 + \zeta^2 |F_{\text{BNS}}(f)|^2}. \tag{7.50}$$

This could be corrected using a compensating weighting filter on the u_∞ input, but this increases the number of system states, requires more computation, and requires a more advanced set of numerical equations. The existing results are satisfactory, though slightly more constrained than necessary to enforce a stability requirement.

7.6 Comparison with AI-Informed Controller Synthesis

Since the process demonstrated in this chapter is limited to creating optimal linear controllers, it is useful to compare its performance to other methods of controller synthesis, such as those explained in [143]. In that work, a similar LIGO ASC problem was posed as reinforcement learning with frequency-domain rewards. A neural-network policy was trained against a linear stochastic model identified from plant measurements and was then exported as a drop-in SISO controller. The deployed target was the LIGO Livingston CHARD pitch loop rather than the DHARD yaw loop considered in this chapter. The control trade-off, however, is the same. Low-frequency control authority must be kept large enough to reject seismic and optical-spring disturbances, while actuation noise in the 10–30 Hz observation band

must be minimized. The deep loop shaping (DLS) policy was reported to reduce control noise in the observation band by more than a factor of 30, with reductions up to a factor of 100 in sub-bands [143].

The DLS reward construction is closely related to the FOM construction, laid out in Section 7.2.2. In both cases, the control objective is specified by spectral weights that separate the broadband stabilization requirement from the narrowband noise-injection penalty. The principal distinction is that DLS searches over a nonlinear neural-network policy using reinforcement learning, while the method developed above solves for the optimal linear controller associated with the specified FOMs. As a result, the linear method provides an explicit Pareto boundary for the modeled problem, while the neural-network method must be benchmarked after training.

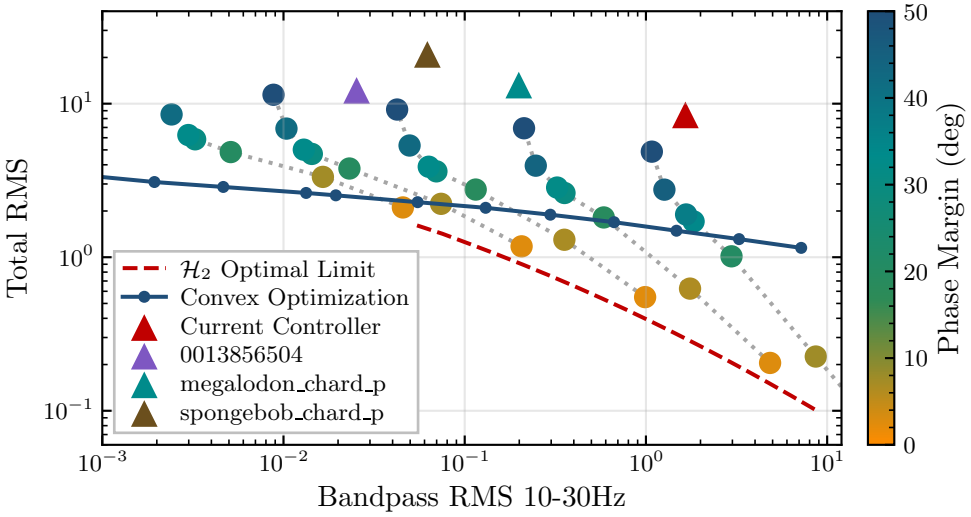


Figure 7.11: Comparison of the optimal linear controllers with the DLS neural-network policies of [143] in a two-objective performance plane. The horizontal axis shows the 10–30 Hz bandpassed RMS, and the vertical axis shows the total RMS. Lower values on both axes indicate better noise performance. The DLS neural-network policies improve on the current linear controller in the 10–30 Hz band, while the optimal linear controllers beat both the DLS and hand-tuned controllers while maintaining an approximately 30° phase margin.

Fig. 7.11 places the DLS policies on a similar two-objective plane used in Fig. 7.7: total RMS motion versus a frequency weighted bandpass RMS (10–30 Hz). The incumbent hand tuned linear controller is improved by the neural-network policies, consistent with the published DLS result. In the same plane, the optimal linear controllers occupy a lower-noise boundary. For comparable 10–30 Hz bandpassed RMS, the robust $\mathcal{H}_{2/\infty}$ controllers have lower total RMS than the DLS policies; equivalently, for comparable total RMS, smaller bandpassed RMS is achieved. This

is achieved for linear controllers that maintain a $> 30^\circ$ phase margin. From the numerical comparison, the DLS policies lie roughly 3–10 times above the algebraic Pareto front in total RMS and fixed bandpassed RMS. The absolute values are uncalibrated so they do not represent an absolute predicted performance like those in Fig. 7.7, but the relative performance is meaningful. The DLS policies are not on the Pareto front, but they are also not far from it, which is a promising result for a nonlinear policy that was not trained to be optimal for the linear model. The DLS policies do appear to be on a similar performance trend as the algebraic controllers, which suggests that they are improving on the same underlying control trade-off. The algebraic controllers, however, are able to achieve better noise performance for acceptable stability margin.

The DLS controllers can not be directly compared to the optimal linear controllers because as nonlinear policies, they are not guaranteed to be limited by the same linear performance bounds. The DLS policies could, in principle, outperform the optimal linear controllers if they are able to leverage nonlinearities in the plant or policy to achieve better noise performance. However, the fact that the DLS controllers were trained on a linear model means that it is almost impossible for the DLS policies to infer the nonlinearities of the plant and thus to outperform the linear controllers for the same model. The DLS policies are most likely limited by the same linear performance bounds, and their distance from the algebraic Pareto front is a measure of their sub-optimality for the modeled problem.

7.6.1 Relation to Convex Optimization

Buchli et al. also produced controllers using convex optimization as a simulation-based linear benchmark [143]. Those optimized linear controllers were reported to have similar predicted performance to the neural-network policies, but were not considered fit for deployment because they were open-loop unstable and had highly aggressive disturbance-rejection behavior. This comparison separates two questions: whether a trained neural-network policy can outperform the incumbent hand-designed controller, which was demonstrated by DLS, and whether a nonlinear policy outperforms the best linear controllers for the same, or similar, frequency-domain objectives, which is the comparison addressed by the algebraic Pareto front.

In this setting, convex optimization begins by choosing a finite set of controller parameters. The controller search is then written as an optimization problem whose objective and constraints are convex in those parameters. The objective can be made

to resemble the same frequency-weighted noise cost used in earlier sections, while additional inequalities can be added to limit loop gain, sensitivity, complementary sensitivity, control effort, or other sampled frequency-domain quantities. This flexibility is the main advantage of the approach: practical requirements can often be added directly as constraints without deriving a new set of Riccati equations for each design choice. The caveat is that the guarantee applies to the optimization problem that was posed, not automatically to the physical loop. Controller order, frequency grid, weighting functions, and robustness constraints are all part of the formulation, and the final controller still has to be checked against the continuous-frequency plant. In addition, this is a more traditional optimization problem, where the solution is found through iterative numerical methods rather than algebraic synthesis. The convex-optimization approach is therefore more flexible and requires fewer constraints, but is less transparent than the algebraic method.

For gravitational-wave control models the plants, noise weights, and FOMs can span many orders of magnitude, making numerical optimization challenging. Dense frequency grids or large LMIs constraints can make the numerical problem expensive, and practical stability-margin constraints may be conservative, approximate, or insufficiently tied to the eventual open-loop shape. This computational cost would also make convex optimization difficult to use in a dynamic implementation that repeatedly updates the controller as the interferometer operating conditions change, as discussed in Section 7.4.2. By contrast, the Riccati method demonstrated above is solved algebraically for each value of ζ, and the mixed-sensitivity extension exposes the robustness trade-off through γ. The resulting controllers therefore carry both a noise-optimal interpretation and a direct stability-margin interpretation.

The solid blue convex-optimization controllers in Fig. 7.11 lie below the incumbent and learned-policy points but remain above the dashed $\mathcal{H}_2$ optimal limit. They also pass through the family of $\mathcal{H}_\infty$-bounded algebraic solutions, indicating that the two linear synthesis methods recover a similar trade-off between broadband motion and 10–30 Hz noise. The remaining separation from the $\mathcal{H}_2$ limit can reflect robustness constraints, the selected controller parametrization, or the numerical formulation of the convex problem. However, the agreement in this two FOM picture does not imply that the resulting controllers are identical; it establishes that they attain comparable values of the two noise metrics.

The magnitude comparison in Fig. 7.12 provides a stronger test of this agreement than the two scalar noise metrics alone. Despite being obtained through different

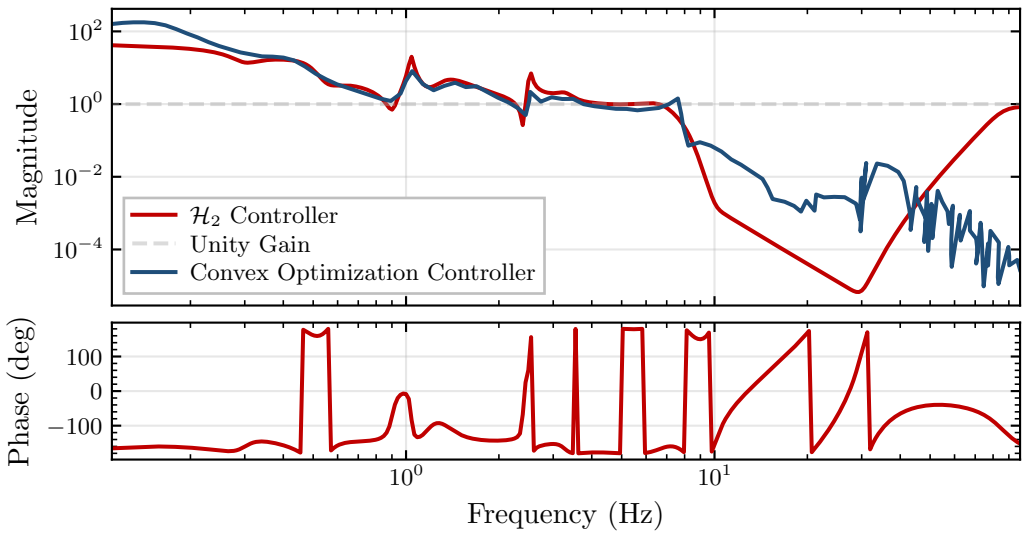


Figure 7.12: Open-loop comparison of two independently synthesized linear controllers. The red trace is the algebraic $\mathcal{H}_2$ controller, and the blue trace is the convex-optimization controller from [143]. Despite the different synthesis methods, their magnitudes reproduce the same broad gain, resonant features, and unity-gain structure throughout the main control band below approximately 10 Hz. The responses separate at higher frequencies, where the algebraic controller initially rolls off more rapidly and the convex controller retains additional structure. The phase of the convex optimization controller was not available.

synthesis procedures, the convex and algebraic controllers have remarkably similar loop shapes throughout the main control band. They reproduce the same broad gain, resonant features, and unity-gain structure below approximately 10 Hz. This independent recovery of nearly the same response suggests that the loop shape is set primarily by the plant and spectral objectives, and it therefore strengthens confidence that the algebraic solution is capturing the relevant optimum rather than an artifact of its particular synthesis method. The traces do not agree exactly, especially at higher frequencies, where the algebraic $\mathcal{H}_2$ controller initially rolls off more rapidly and the convex-controller response contains additional structure. Part of this difference may arise from how the traces were obtained: the algebraic response is calculated from the model, whereas the convex-controller loop was measured from the model. The rougher high-frequency convex trace may therefore include reduced measurement signal-to-noise, imperfect coherence, or finite averaging in addition to any intrinsic difference between the controllers.

Taken together, the agreement in both the performance plane and the open-loop response provides independent validation of the algebraic synthesis developed in this chapter. For the same Livingston CHARD pitch degree of freedom, convex

optimization reaches nearly the same high-performing region and recovers the principal loop-shape features predicted by the algebraic solution. Both linear methods attain lower values of the two modeled noise objectives than either the current controller or the DLS policies, even though the DLS policies still substantially improve performance relative to the incumbent design. These CHARD-pitch results are distinct from the DHARD-yaw controllers developed in the main sections, demonstrating that the algebraic procedure is applicable to many different DOFs. More importantly, the algebraic construction provides an explicit $\mathcal{H}_2$ performance limit, a Pareto front of robustness-constrained solutions, and a direct connection between γ and stability margin. The result is therefore more than an improved controller: it establishes a quantitative benchmark for what linear control can achieve for a system and its objectives while preserving the required robustness. That benchmark shifts the central question from whether a controller improves on the incumbent to how closely it approaches the best robust linear performance permitted by a system.

Part III

Commissioning for GQuEST

Chapter 8

OPTICAL FILTER CAVITY CHAIN FOR PHOTON-COUNTING

"Water, water, every where,
Nor any drop to drink"

— S. T. Coleridge, *The Rime of the Ancient Mariner*

The photon-counting readout described in Chapter 3 requires an optical filter that is much more selective than the output optics of a conventional interferometer. The geontropic signal appears as incredibly weak optical sidebands at MHz offsets from the laser carrier, while contrast defect, residual motion, scattered light, and imperfect spatial-mode matching send much larger carrier-frequency fields toward the output port. The OFCs provide the last frequency-selective element before the SNSPD: they reject light at the carrier frequency ν and transmit a narrow band near the selected sideband frequency $\nu + \epsilon_r$.

The first OFC was built as the construction and controls test bed for the rest of the output-filter chain. That cavity established the monolithic body design, the mirror-mounting and cleaning procedure, the 1550 nm finesse measurement, the 775 nm auxiliary lock, and the locked-noise measurement workflow. The full photon-counting readout uses four OFCs in series; the first chained systems were used to demonstrate the alignment and control procedure before the full four-cavity readout was operated as a single filter.

8.1 Output Filtering for Photon-Counting Readout

The output filters translate the photon-counting idea into a difficult optical requirement: transmit the small signal sidebands while rejecting the much larger carrier field. The solution is four OFCs all locked to one another to create a narrow pass band at the selected sideband frequency. Keeping these cavities aligned and locked together is a major technical challenge.

Another photon counting experiment has used coupled cavity chains [144]. That experiment used four cascaded free-space Fabry-Perot cavities with linewidths near 30 kHz to isolate weak mechanical sidebands from a much stronger pump field.

The cavities were actively aligned with an auxiliary locking beam, but the locking light had to be temporarily disabled during photon counting to avoid saturating the detector. The measurement relied on the already aligned filter remaining usable while the lock light was off before the cavities fell out of resonance. The photon-counting readout here asks for a more persistent version of the same problem: several independent cavities must be chained together, locked with auxiliary fields, mode matched in series, and aligned so that their 1550 nm pass bands overlap on the same MHz-offset signal sideband while their carrier resonances remain rejected. This section lays out the frequency-selective role of the OFCs, beginning with the sideband readout picture and then turning to the carrier-rejection, linewidth, and pass-band alignment requirements that set the cavity design.

8.1.1 Signal Sideband Readout

Photon counting changes the measurement from a continuous optical-power readout to a count-rate measurement in a narrow optical band. As reviewed in Section 3.3.4, a differential interferometer length fluctuation creates optical sidebands around the laser carrier. In a homodyne readout those sidebands beat with carrier-frequency local-oscillator light. In the photon-counting readout, the interferometer is instead operated close to destructive interference and the output optics remove the carrier as completely as possible. Photons that survive the output filters are then counted directly with an SNSPD.

The frequency selectivity of the OFCs is central to the readout. The selected sideband frequency is near the expected geontropic signal band and near a relatively clean region of the high-frequency thermal-noise model seen in Fig. 3.11. The reference GQuEST design uses a frequency detuning, ϵ_r, of 17.6 MHz with practical operation over a wider MHz tuning range. The geontropic signal is broadband, but the photon-counting filter is intentionally narrow so that the count rate can be measured in a frequency band where classical backgrounds, carrier leakage, and higher-order-mode transmission are all minimized. In Fig. 3.11, the offset sideband falls between the dominant thermal noise peaks, which are the result of the mechanical resonances of the interferometers' end mirrors, discussed in Section 3.3.3.

This makes the OFCs complementary to the LFC, discussed in Chapter 9. The LFC is locked to the carrier and uses its passive low-pass response to remove laser-noise sidebands before the interferometer. The OFCs do the opposite job at the output:

they deliberately reject the carrier and pass only a narrow MHz-offset band. The control problem is different as well. The LFC can use the carrier as the resonant reference, while the OFCs must be held with an auxiliary field so that the carrier remains off resonance during the measurement.

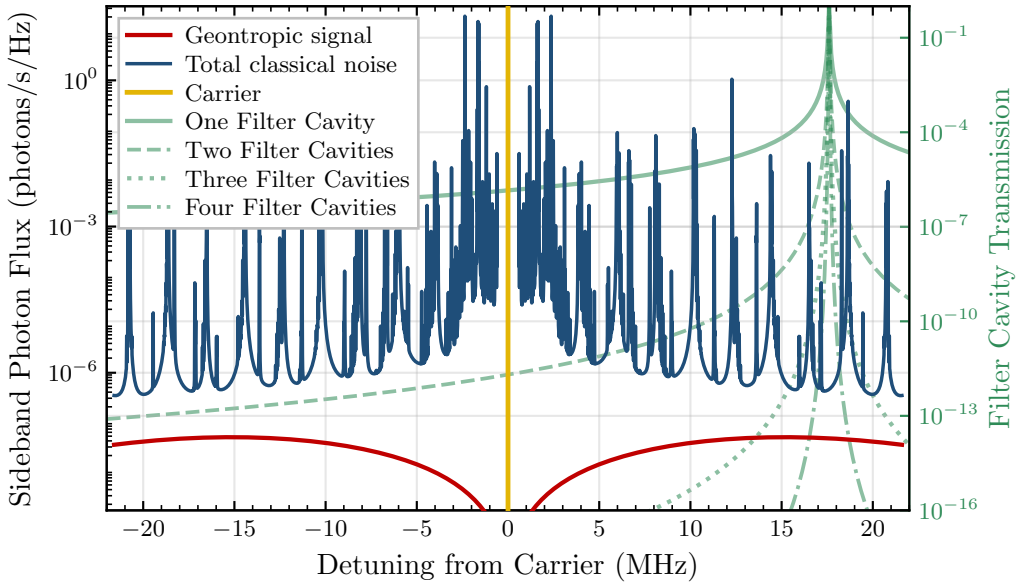


Figure 8.1: Signal-sideband transmission through the output filter chain. The red trace is the geontropic signal sideband photon-flux spectral density inferred from the GQuEST signal model, and the blue trace is the total classical-noise photon-flux spectral density from the GQuEST noise budget shown in Fig. 3.11. The narrow spikes in the blue trace are the high-frequency thermal-noise resonances of the interferometer optics, discussed in Section 3.3.3. The spectra are mirrored about the carrier so that the upper and lower optical sidebands can be compared on the same detuning axis. The yellow vertical line marks the carrier. The green curves, plotted against the right axis, show the transmission of one through four cascaded OFCs centered at the nominal ϵ_r pass band. Cascading the cavities narrows the transmitted band around the selected sideband while strongly suppressing carrier-frequency light and nearby off-resonant backgrounds before the SNSPD.

8.1.2 Carrier-Rejection Requirement

The carrier-rejection requirement is set by the contrast between the expected signal photon flux and the carrier-frequency light leaving the interferometer. In the GQuEST design, contrast defect and low-frequency interferometer motion can send order 0.1-1 W of carrier-frequency light toward the readout. The target signal and classical-noise count rates after filtering are on the 10^{-3}-10^{-2} Hz scale. Since one watt of 1550 nm light corresponds to order 10^{18} photons per second, more than 200 dB of carrier-power suppression is required before the SNSPD path.

The design value adopted for the output-filter chain is roughly $A_{\mathrm{carrier,chain}} \gtrsim 200\,\mathrm{dB}$, with the nominal four-cavity design providing about 230 dB at the ϵ_r offset [1]. This level puts fundamental-mode carrier leakage below the dominant expected photon-counting backgrounds in the reference design. This means each cavity must provide on the order of 50-60 dB of carrier suppression, which is a challenging but achievable level for a single moderate-finesse cavity. The four-cavity chain then provides the full suppression.

This requirement is a power-suppression requirement for the spatial mode that actually enters the cavity chain. Light that is scattered around the mode, coupled into higher-order modes, or generated downstream of the first filter does not necessarily receive the same rejection. For that reason the carrier-rejection calculation is only the starting point. The experimental validation also has to measure the transmitted carrier in the fundamental mode, the residual higher-order-mode transmission, and any background photons produced by the auxiliary-control system or by scatter near the detector path.

8.1.3 Single-Cavity Finesse and Bandwidth

Each OFC is a moderate-finesse cavity rather than an extremely narrow single filter. For a single Lorentzian cavity with FWHM linewidth Δf_1, the off-resonant carrier transmission at a detuning $\epsilon_r \gg \Delta f_1$ is approximately

$$T_1\,(0) \simeq \left(\frac{\Delta f_1}{2\epsilon_r}\right)^2. \tag{8.1}$$

For $\Delta f_1 = 42$ kHz, this gives about 58 dB of carrier-power suppression per cavity. Four cavities in series then give about 232 dB of carrier suppression.

The same linewidth must still transmit enough of the signal sideband to keep the count rate high enough. The chosen single-cavity linewidth is near

$$\Delta f_1 \simeq 42\ \mathrm{kHz}, \tag{8.2}$$

corresponding to a cavity pole near 21 kHz. For a round-trip length near 2.4 m, the FSR is about 125 MHz, so the design finesse is

$$\mathcal{F}_{\mathrm{OFC,design}} \simeq \frac{125\ \mathrm{MHz}}{42\ \mathrm{kHz}} \simeq 3.0 \times 10^3. \tag{8.3}$$

This value is close to the measured finesse of the first cavity after optical cleaning, discussed in Section 8.6.1.

The choice of a moderate finesse is a deliberate compromise. Increasing the finesse would improve the single-cavity carrier rejection, but it would also make the pass band narrower and make four-cavity alignment more fragile. Decreasing the finesse would make it easier to keep the cavities on resonance together, but it would require more cavities for the same suppression or would leave too much carrier light for the SNSPD path. The current design is a balance between those two tradeoffs and has more forgiving optical requirements because it multiplies moderate filters rather than relying on one extremely narrow cavity.

8.1.4 Four-Cavity Pass-Band Alignment

The four-OFC chain must have all four pass bands aligned with one another near the selected signal-sideband frequency. A residual relative lock-point error reduces the signal transmission even if each cavity remains close to resonance. For one cavity with FWHM linewidth Δf_1, the normalized power transmission near resonance is

$$T_i\left(\delta f_i\right) = \left[1 + \left(\frac{2\delta f_i}{\Delta f_1}\right)^2\right]^{-1}. \tag{8.4}$$

Only the relative detunings enter this signal-transmission loss. If all four cavity pass bands shift together, the peak of the passband shifts in frequency, but the transmission is unchanged. Thus the common detuning is removed before estimating the pass-band overlap. Defining σ_f as the comparable RMS scale of the remaining three relative detuning modes, the chain transmission can be estimated as

$$T_{\text{chain}} \simeq \left[1 + 4\left(\frac{\sigma_f}{\Delta f_1}\right)^2\right]^{-3}. \tag{8.5}$$

Requiring less than a 10% loss of signal power, or $T_{\text{chain}} > 0.9$, gives

$$\sigma_f < \frac{\Delta f_1}{2}\sqrt{0.9^{-1/3} - 1} = 0.0945\,\Delta f_1. \tag{8.6}$$

For the design linewidth $\Delta f_1 = 42$ kHz, this gives

$$\sigma_f \lesssim 4.0 \text{ kHz RMS}. \tag{8.7}$$

This corresponds to a small equivalent cavity-length error. Using $f_{\text{FSR}} = 125$ MHz and $\lambda = 1550$ nm,

$$\sigma_{L,\text{rt}} = \lambda\frac{\sigma_f}{f_{\text{FSR}}} \lesssim 0.049 \text{ nm RMS}. \tag{8.8}$$

Thus the four-cavity signal-transmission requirement is that the relative pass-band alignment be held within a few kilohertz RMS, or about 5×10^{-2} nm RMS in equivalent round-trip length, after subtracting the common motion of the four cavity resonances.

The caveat is that the common motion cannot be arbitrarily large. The OFCs are placed in a low-noise frequency bin between the high-frequency thermal-noise features shown in Figs. 3.11 and 3.13. Those features are separated by order 0.5-1 MHz near the readout band. A common drift of all four cavities by a comparable amount would preserve the four-cavity product transmission but would move the photon-counting band out of the intended thermal-noise valley. The absolute common-mode pass-band drift must therefore remain small compared with this MHz-scale spacing, while the much tighter few-kilohertz requirement applies to the relative overlap of the four OFC pass bands.

This tolerance includes both static and dynamic errors. Static errors come from setting the wrong acousto-optic modulator (AOM) frequency or choosing the wrong auxiliary resonance. Dynamic errors come from residual cavity-length motion after the PDH loop is closed. The practical goal of the calibration and noise measurements is to lock each cavity and verify that the locked pass bands remain overlapped under this threshold during the photon-counting measurement.

Table 8.1 summarizes the reference design values used for the OFC chain. The measured values for OFC 1 are given later in Section 8.6; the table here collects the design scale that sets the optical layout, carrier rejection, and auxiliary lock requirements. The curved-mirror radius of curvature (ROC) is listed as the common nominal value, but small cavity-to-cavity changes in R_c are also a possible way to shift the transverse-mode spectrum; their effect on the four-cavity higher-order mode (HOM) product transmission is discussed in Section 8.5.2.

8.2 Frequency Alignment of the OFC Chain

The cavity-chain frequency alignment is the practical implementation of the filtering requirement from Section 8.1.4. Each cavity has to be held so that the same 1550 nm signal sideband is transmitted, but the carrier and unwanted neighboring resonances are not transmitted through the full chain. This section describes the frequency picture used to choose the 1550 nm pass band, the 775 nm auxiliary lock points, and the intentional length offsets between cavities.

Quantity	Symbol	Design value
Number of OFCs	N	4
Signal wavelength	λ	1550 nm
Auxiliary-control wavelength	λ_{ctrl}	775 nm
Nominal signal-sideband offset	ϵ_r	17.6 MHz
Round-trip optical path length	L_{rt}	2.4 m
Free spectral range	f_{FSR}	125 MHz
1550 nm design finesse	$\mathcal{F}_{1550}$	3.0×10^3
1550 nm FWHM linewidth	Δf_1	42 kHz
1550 nm cavity pole	f_{pole}	21 kHz
Carrier suppression per cavity		$\sim$ 58 dB
Four-cavity carrier suppression		$\sim$ 232 dB
Relative pass-band alignment requirement	σ_f	$\lesssim$ 4.0 kHz RMS
Equivalent round-trip length requirement	$\sigma_{L,\mathrm{rt}}$	$\lesssim$ 0.049 nm RMS
Curved mirror ROC		-1.6 m
Curved-mirror AOI		3°
1550 nm input/output coupler reflectivity	R_{1550}	99.91%
1550 nm input/output coupler transmissivity	T_{1550}	883 ppm
775 nm auxiliary finesse	$\mathcal{F}_{775}$	$\sim$ 300
Representative AOM drive frequencies	f_{AOM}	200-208 MHz
Minimum PDH modulation-frequency separation		$\sim$ 1 MHz

Table 8.1: Reference design values for the OFCs. These values define the nominal four-cavity photon-counting filter and are distinct from the measured OFC1 characterization results in Section 8.6.

8.2.1 1550 nm Carrier Rejection and Sideband Transmission

The 1550 nm frequency alignment has two simultaneous constraints. The desired sideband frequency $\nu + \epsilon_r$ must be resonant in every cavity, while the carrier frequency ν must remain far from resonance in every cavity. If the cavities had identical lengths, neighboring longitudinal and transverse resonances would also line up. That would create multiple passbands. If another passband overlapped with an interferometer end mirror resonance, visible in Fig. 3.11, then a sideband generated by the thermal noise of that resonance could leak through the cavity chain and be detected by the SNSPD. The solution is to stagger the cavity lengths, as in Fig. 8.2, so that the FSR is not identical from cavity to cavity. The shifted neighboring resonances in Fig. 8.2 indicate the intentional length differences between cavities.

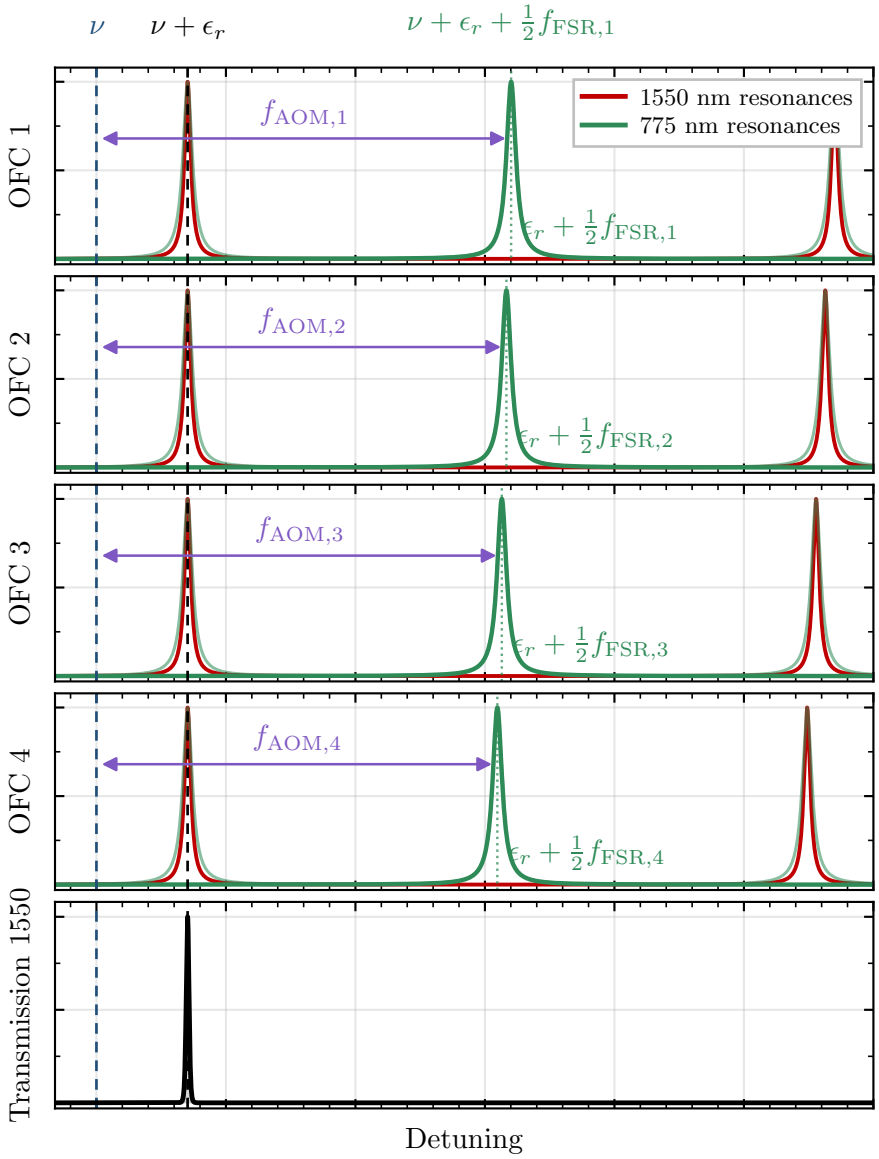


Figure 8.2: Frequency alignment of the four OFCs on an equivalent 1550 nm frequency axis. The red traces mark the 1550 nm resonance combs, with the selected sideband $\nu + \epsilon_r$ aligned to the common pass band while the carrier frequency, ν, is rejected. The green traces mark the 775 nm auxiliary resonance combs, which appear near the half-FSR points on this axis. For each cavity, the AOM frequency places the auxiliary field on the 775 nm resonance one half-FSR above the selected 1550 nm pass band. Solving for the RF drive gives $f_{\mathrm{AOM},i} = (2/N_{\mathrm{pass}})[\epsilon_r + (n_i + 1/2) f_{\mathrm{FSR},i}]$, which reduces to $f_{\mathrm{AOM},i} = \epsilon_r + (n_i + 1/2) f_{\mathrm{FSR},i}$ for the double-pass paths used here. The lock points use $n_i = 1$, placing the AOM drives in the 200-208 MHz range, and the staggered round-trip lengths give each row its own $f_{\mathrm{FSR},i}$. The length stagger and auxiliary-lock offsets are visually enlarged here so that the shifted neighboring resonances from cavity to cavity, and the bottom panel shows the product of the four 1550 nm transmissions. The resonance widths, visualized offset spread, and transmission are not to scale.

8.2.2 775 nm Auxiliary Resonances

The operating lock is made with 775 nm light. Because 775 nm light has twice the optical frequency of the signal light, its resonance comb is spaced by half the 1550 nm FSR when plotted on an equivalent 1550 nm frequency axis. This second resonance is used to hold the cavity steady while not allowing the carrier frequency to pass through. The control resonance used for each cavity must be selected carefully so that a stable auxiliary lock also produces the desired 1550 nm pass band.

Each 775 nm path contains an AOM. On the equivalent 1550 nm detuning axis, the carrier is at $\Delta f_\nu = 0$ Hz and the desired 1550 nm pass band is at ϵ_r. The 775 nm auxiliary light has resonances halfway between the 1550 nm resonances on this same axis.[1] Because the four cavities have slightly different lengths, each cavity has its own FSR, $f_{\mathrm{FSR},i}$. The AOM frequency for cavity i is chosen to place the delivered 775 nm light one half-FSR beyond the selected 1550 nm pass band,

$$f_{\mathrm{AOM},i} = \frac{2}{N_{\mathrm{pass}}}\left[\epsilon_r + \left(n_i + \frac{1}{2}\right) f_{\mathrm{FSR},i}\right], \tag{8.9}$$

where n_i is an integer. For the double-pass AOM paths in Fig. 8.8, $N_{\mathrm{pass}} = 2$, and Eq. 8.9 reduces to $f_{\mathrm{AOM},i} = \epsilon_r + (n_i + 1/2) f_{\mathrm{FSR},i}$. Changing n_i chooses which 775 nm longitudinal resonance is used for the auxiliary lock. This does not change the modulo-FSR filtering condition, but it does change the required AOM drive frequency. The $n_i = 0$ solution would put the drive near 80 MHz, while $n_i = 1$ puts it in the 200-208 MHz range, in the practical operating range of the AOMs and RF chain. For the model round-trip lengths listed in Table 8.2, $f_{\mathrm{FSR},i}$ spans 124.91 to 123.61 MHz, giving the listed $f_{\mathrm{AOM},i}$ values for $\epsilon_r = 17.6$ MHz, $n_i = 1$, and $N_{\mathrm{pass}} = 2$. The AOM frequency is thus chosen to place the 775 nm lock point on the center of a half-FSR auxiliary resonance, where the cavity can be held with PDH locking while the 1550 nm carrier remains off resonance. The relevant quantity is the offset modulo the cavity FSR; the absolute optical frequency shift can be hundreds of megahertz while the detuning that matters to the filter is the placement of the 775 nm lock relative to the 1550 nm sideband pass band. The double-pass geometry also reduces beam-pointing changes when the AOM frequency is changed by using a curved mirror to retroreflect the beam back through the same AOM crystal.

[1] Strictly, the AOM-shifted 775 nm field is not exactly twice the 1550 nm carrier frequency. If the AOM path shifts the 775 nm optical frequency by $N_{\mathrm{pass}} f_{\mathrm{AOM},i}$, the half-FSR term is multiplied by $\eta_i = 2\nu/(2\nu + N_{\mathrm{pass}} f_{\mathrm{AOM},i}) \simeq 1 - N_{\mathrm{pass}} f_{\mathrm{AOM},i}/(2\nu)$. For the double-pass paths and $f_{\mathrm{AOM}} \sim 205$ MHz, this changes the lock point by about 2.0×10^{-4} MHz, so it is neglected in the MHz-scale of Eq. 8.9, Fig. 8.2, and Table 8.2.

In actuality, the exact AOM frequencies are adjusted during lock acquisition to optimize the pass-band alignment. The model values in Table 8.2 give a starting point for the AOM frequencies, but the final values are found empirically by adjusting the AOM frequencies while monitoring the cavity transmission and error signals.

8.2.3 Length Offsets Between Cavities

The cavity lengths were intentionally offset from one another so that the FSRs are not identical. The design round-trip lengths are listed in Table 8.2 and are the same lengths used in the HOM model in Section 8.5.1. The total span is one inch, chosen to provide adequate suppression of the second resonant mode, as plotted in Fig. 8.2, while still small compared with the 2.4 m cavity length.

Cavity	L_{rt}	ΔL_{rt}	f_{FSR}	f_{AOM}
OFC 1	2.40000 m	0 mm	124.91 MHz	204.97 MHz
OFC 2	2.40847 m	8.47 mm	124.47 MHz	204.31 MHz
OFC 3	2.41693 m	16.93 mm	124.04 MHz	203.66 MHz
OFC 4	2.42540 m	25.40 mm	123.61 MHz	203.01 MHz

Table 8.2: Round-trip length offsets, corresponding FSRs, and $n_i = 1$ AOM frequencies used for the four-OFC chain model with $\epsilon_r = 17.6$ MHz. The offsets are measured relative to OFC 1. Each cavity is designed to be exactly 1/4 inch longer than the previous one, giving a total span of one inch.

These offsets were chosen to make unwanted resonances move by much more than a cavity linewidth. Adjacent cavities differ in FSR by about 440 kHz, which is more than ten times the 42 kHz design linewidth. The desired 1550 nm pass band is recovered with the 775 nm auxiliary lock points and AOM offsets, but neighboring longitudinal resonances and many higher-order spatial-mode resonances no longer line up from cavity to cavity.

The size of this effect can be estimated from the next longitudinal resonance. If one cavity is resonant one FSR above the desired pass band, the other cavities are detuned by the differences in their FSRs. Using a Lorentzian power transmission,

$$T_i(\delta f) = \frac{1}{1 + \left(\frac{2\delta f}{\Delta f_{\mathrm{FWHM}}}\right)^2}, \tag{8.10}$$

with $\Delta f_{\mathrm{FWHM}} = 42$ kHz, the worst-case product transmission of the three off-resonant cavities at a neighboring longitudinal resonance is about 3.1×10^{-9}. This corresponds to roughly 85 dB of additional rejection compared with four identical

cavities whose adjacent FSR resonances would all overlap. For the overlap of the edge cavities in this length sequence, OFC 1 and OFC 4, the corresponding rejection is about 95 dB. The controls align the one resonance that is wanted, while the mechanical offsets keep accidental resonances from being transmitted by the full chain.

8.3 Optical and Mechanical Design

The optical design of each cavity has to satisfy two requirements at the same time. It must provide a stable, reproducible passband for the signal sideband for photon counting, while also leaving enough access for the 775 nm auxiliary lock and the mechanical actuation. The design combines a bow-tie optical layout, dichroic mirror coatings, a rigid cavity body, passive damping, and a piezoelectric transducer (PZT) length actuator.

8.3.1 Bow-Tie Geometry and Cavity Mode

Each OFC is a four-mirror traveling-wave bow-tie cavity, depicted in Fig. 8.3. The traveling wave geometry separates the reflected and transmitted beams and gives independent access to the 1550 nm signal path and the 775 nm control path. A linear cavity would make these paths harder to separate and would complicate the injection of the auxiliary control field. The bow-tie geometry also avoids direct back-reflection toward the upstream optics and leaves physical room for photodiodes (PDs), steering mirrors, and fiber couplers. The cavity also allows for the control field to counter-propagate with the signal field, which minimizes the possibility of back-scatter from the control field reaching the SNSPD path.

Only M3 is curved, with a ROC of -1.6 m, and the remaining mirrors are flat. This single curved optic provides the transverse stability of the cavity mode while keeping the mode structure easy to interpret. The angle of incidence (AOI) is approximately $3°$, which separates the beams while keeping the astigmatism from the curved mirror small enough for efficient mode matching and predictable transverse-mode spacing. The aluminum cavity body fixes the nominal bow-tie geometry, and the deliberate round-trip length offsets described in Section 8.2.3 are then introduced with shims on the mirror mounts. Those offsets are part of the filter design rather than construction errors to be tuned away.

The nominal geometry was modeled in the same way as for the LFC design [145]. The resulting eigenmode has a waist radius of 584 μm on M1 and reaches a maximum beam radius of 1.17 mm at the concave curved M3 mirror, as plotted in Fig. 8.4.

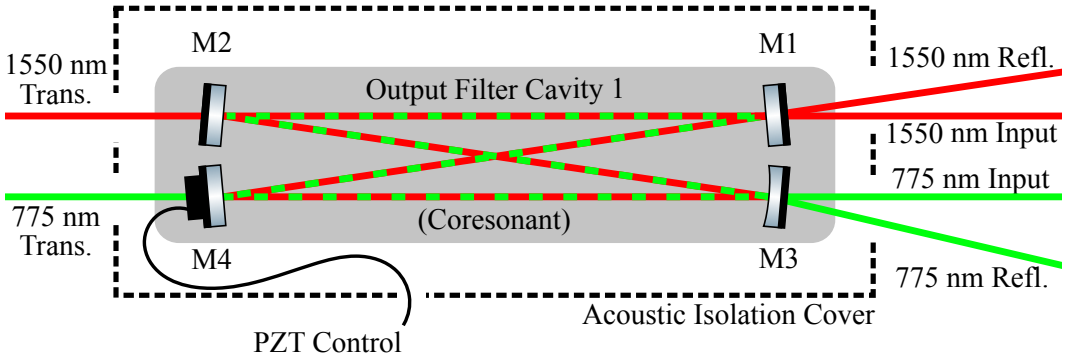


Figure 8.3: Optical layout of a single four-mirror traveling-wave OFC. The 1550 nm signal field is shown in red and the 775 nm auxiliary control field is shown in green. In this layout both fields are co-resonant in the bow-tie cavity, but they use different input and reflection ports: the 1550 nm light enters M1 and transmits toward the left side of the cavity, exiting at M2, while the 775 nm light enters near M3 and transmits through M4. The only curved mirror is M3, with a ROC of -1.6 m. The separated reflection and transmission ports make it possible to monitor the 1550 nm and 775 nm fields independently during construction, alignment, and locking. The dashed outline indicates the acoustic isolation cover, which doubles as a protective shield that reduces scattered light, and the PZT actuator on M4 provides the length-control input for holding the cavity on the selected auxiliary resonance.

These were the mode-matching targets for both input beams, although the input beam had to match the beam size at its relevant input coupler. For the 1550 nm signal path, the input coupler is M1, so the mode-matching target was to put a 584 μm waist on M1. For the 775 nm control path, the input coupler is M3, so the mode-matching target was to put a 1.17 mm beam on M3, but the 775 nm beam is converging after M3 so the input beam had to be converging with a size of 1.7 mm on M3 and be converging to a 584 μm waist at M1.

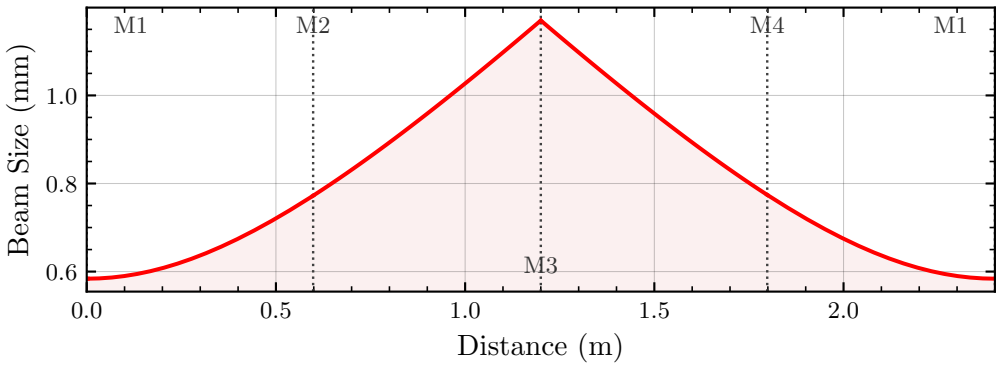


Figure 8.4: Modeled Gaussian beam radius around the unfolded OFC round trip. The cavity eigenmode has a waist radius of 584 μm near M1 and reaches a maximum radius of 1.17 mm at the curved M3 mirror before returning to M1.

The optical model gives a round-trip Gouy phase of 120.1°, plotted in Fig. 8.5. This value is close to one third of a full transverse-mode cycle, so the third-order transverse-mode family lies close to the next longitudinal resonance, with the sixth- and ninth-order families repeating near later longitudinal resonances. This is similar to the design of the LFC cavity, which also had a round-trip Gouy phase near 120° to keep the higher-order modes away from the suppression offset.

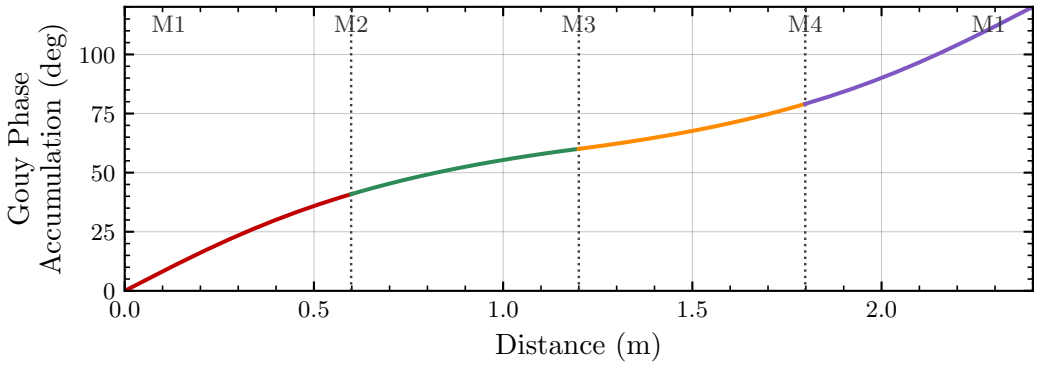


Figure 8.5: Accumulated Gouy phase around the unfolded OFC round trip. For the nominal 2.4 m geometry and the −1.6 m curved mirror ROC used in the optical model, the round-trip Gouy phase is 120.1°, close to one third of a full transverse-mode cycle.

This transverse-mode placement is clearer in Fig. 8.6. It is not as much of a problem if the input beam is well mode matched, but because the control resonance will be in the center of the FSR, that area needs to be kept clear of HOMs. A single OFC can exhibit small transverse-mode resonances near a fundamental peak, while the full chain relies on the length offsets and independent auxiliary lock points to keep those unwanted resonances from overlapping in all four cavities.

8.3.2 Mirror Coatings and Optical Functions

The mirror coatings define two wavelength-dependent optical systems inside the same mechanical cavity. At 1550 nm, M1 and M2 are the input and output couplers, with transmissions near 883 ppm. M3 and M4 are high reflectors at 1550 nm, so the signal field sees a high-finesse four-mirror filter cavity. At 775 nm, the optical roles are reversed: M3 and M4 provide the auxiliary input and output coupling, while M1 and M2 are highly reflecting. This dichroic design allows the 775 nm control field to be resonant and measurable even when the 1550 nm carrier is rejected. The nominal coating functions are summarized in Table 8.3.

The corresponding design finesse at 1550 nm is about 3.0×10^3, matching the

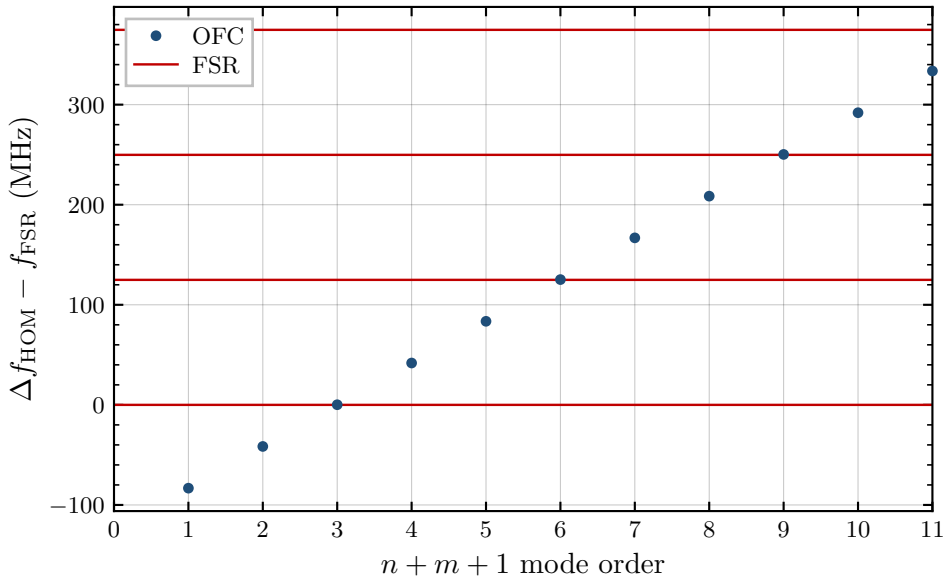


Figure 8.6: Higher-order transverse-mode spacing from the OFC optical model. The markers give the signed offset $\Delta f_{\mathrm{HOM}} - f_{\mathrm{FSR}}$ for mode order $n + m + 1$. The horizontal lines mark zero offset and one, two, and three FSRs. Because the round-trip Gouy phase is close to 120°, the third-, sixth-, and ninth-order families fall near successive longitudinal resonances.

Mirror	1550 nm coating	775 nm coating	Cavity role
M1	$T \simeq 883$ ppm	$T + L \lesssim 1$ ppm	1550 nm input coupler
M2	$T \simeq 883$ ppm	$T + L \lesssim 1$ ppm	1550 nm output coupler
M3	$T + L \lesssim 1$ ppm	$T \simeq 10^4$ ppm	775 nm coupler, curved optic
M4	$T + L \lesssim 1$ ppm	$T \simeq 10^4$ ppm	775 nm coupler, length actuator

Table 8.3: Nominal coating functions for the OFCs, written in transmission units. The 1550 nm field is coupled through M1 and M2, while the 775 nm auxiliary field is coupled through M3 and M4. M3 is the curved mirror with an ROC of −1.6 m in the cavity-design sign convention (negative for convex).

bandwidth requirement in Eq. 8.2. The 775 nm finesse is lower, with a representative value near $\mathcal{F}_{775} \simeq 300$. This lower auxiliary finesse helps the control loop because it gives a robust reflection signal and a wider capture range, while the narrower 1550 nm resonance supplies the photon-counting filter.

The coating choices also define which ports serve as diagnostics. The 1550 nm transmission through M2 directly monitors the photon-counting pass band and is the natural signal for 1550 nm finesse measurements, transmission-peak checks, 775 nm light leakage, and chained transmission measurements. The reflected 775 nm light from the auxiliary input side is instead the operating error signal. Keeping these two functions on different wavelength-dependent ports is what allows the cavity to be locked continuously without intentionally transmitting the carrier-frequency light that the filter is designed to reject.

8.3.3 Cavity Body, Rubber Isolation, and PZT Actuation

The cavity bodies were built as solid machined aluminum enclosures rather than as collections of post-mounted mirrors. The monolithic body fixes the nominal bow-tie geometry, reduces relative motion between mirrors, and makes later cavities reproducible once the first cavity assembly was understood. Before the deliberate shim offsets are added to give the shift in round trip length, the nominal OFC geometry uses two 0.598 m short sections and two 0.602 m diagonal sections, giving a 2.400 m round-trip length. These are the same segment lengths used in the HOM calculation. Each body sits on rubber isolation spacers that provide passive damping between the cavity and the optical table. This support is more elaborate than the vacuum suspension of the LFC and provides a less stiff mechanical suspension.

When in operation the cavities are covered with lids and internal baffles to reduce acoustic pickup, dust exposure, and room-light scatter. The covers[2] are made of aluminum-covered board with four holes placed to allow access to the 8 cavity ports and the PZT feedthrough.

One mirror in each cavity is mounted on a PZT actuator. The PZT provides the length actuation for acquisition, scans, and the locked operating state. It is a fine actuator around the fixed mechanical length of each cavity; it is not used to make the four FSRs identical. The actuator range must cover slow thermal drifts, while the actuator bandwidth must be high enough to stabilize the cavity length against table motion and acoustic pickup. The first cavity established the practical mounting

[2] Called earmuffs

preload, cable routing, and loop-sign conventions used for the later OFCs.

The spacer stack and cavity body set the mean FSR, including the deliberate offsets from Table 8.2 and are static. The PZT then holds the cavity near the selected auxiliary resonance and removes residual motion about that operating point. In this sense the actuator is not a coarse length-trimming mechanism; it is the dynamic part of the auxiliary lock. A stable chain requires both reproducible mechanical construction and enough control range to keep each intentionally different cavity on its own calibrated lock point.

The in-air mechanical implementation also sets the noise environment for the auxiliary locks. Acoustic pickup, table motion, and resonances of the actuated mirror can all appear as equivalent detuning noise of the 1550 nm pass band. The relevant mechanical measurements are the same ones that matter for operation: the PZT-to-length transfer function, the locked residual noise spectrum, the effect of acoustic covers or lids, the mirror finesse, and the stability of the lock over the time scale of a photon-counting run. The first cavity was used to develop the mechanical design and the control procedure, and the later cavities were built to match that design as closely as possible.

8.3.4 Construction of the First OFC

OFC 1 was used to develop the construction sequence. The mirrors were installed and aligned first at 1550 nm, where the cavity transmission could be scanned directly and the fundamental mode was easy to identify. The cavity was cleaned and reassembled a number of times, and finesse measurements were performed to check the design parameters. After initial testing the cavity was disassembled and the suspension was installed. The lid was then used to reduce room-light pickup, dust exposure, and scatter from other paths.

Figure 8.7 presents the optical layout used for this first-cavity commissioning. The 1550 nm path comes from the power distribution center, where the laser light is split so that the OFC can be tested independently of the interferometer input chain. This direct 1550 nm input passes through power-selection optics and an electro-optic phase modulator (EOPM), which adds the PDH sidebands used for diagnostic locks, linewidth checks, and ringdown measurements. An input-side flipper mirror then makes the setup reversible: the same input optics can be used either as a local OFC 1 test path or as the path that sends light toward the interferometer input. A pickoff of the same laser is sent to second-harmonic generation (SHG) to make the 775 nm

auxiliary field. That field has its own mode-matching path and diagnostics before it is injected into the same cavity mode. The layout keeps the signal-wavelength and auxiliary-control measurements connected to the same physical cavity while leaving enough independent PDs and cameras to diagnose them separately. In its testing setup, the upstream light did not come from the interferometer. This allowed for independent testing of OFC 1 without worrying about the PRC or the other cavities.

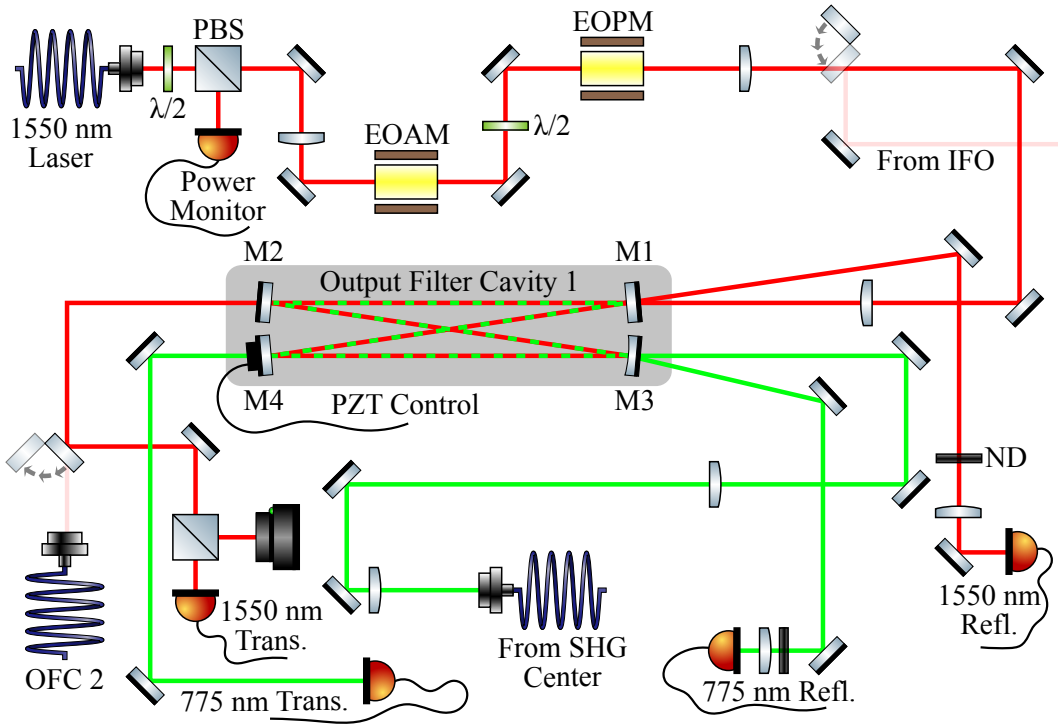


Figure 8.7: Optical layout of the OFC 1 test setup used during construction and commissioning. The red path is the directly injected 1550 nm laser field, with selectable input power and an EOPM used to add PDH sidebands for diagnostic locking and linewidth scans. An input-side flipper mirror allows the same input optics to be switched between the local OFC 1 test path and the interferometer-input path. The green path is the 775 nm auxiliary field generated by the SHG stage and injected through the separate M3 input coupler. At the 1550 nm cavity output, a second flipper mirror selects either the local diagnostic path, where the transmitted beam is sent to the 1550 nm transmission photodiode and camera, or the chain path, where the OFC 1 output is coupled toward OFC 2. The 775 nm output has its own transmitted photodiode and camera (not shown) for auxiliary-lock diagnostics.

The input optics in Fig. 8.7 were chosen to make OFC 1 a capable stand-alone test cavity before it was integrated into the full chain and interferometer. On the 1550 nm path, a polarizing beam splitter (PBS) sets the power incident on the cavity, while the electro-optic amplitude modulator (EOAM) and EOPM provide controlled

amplitude and phase modulation for test measurements. This made it possible to drive transmission scans, measure ringdowns, and test error-signal generation without changing the upstream laser preparation through the LFC, described in Chapter 9. The 1550 nm beam then enters the cavity through M1 and exits through M2. The input flipper mirror was included so that this diagnostic input path could later be exchanged for the interferometer-input path without rebuilding the OFC 1 input optics.

The 775 nm input path is deliberately simpler on all of the OFCs. The auxiliary light is already phase modulated at a common frequency immediately after the SHG stage, before it is split into the individual cavity paths seen in Fig. 8.8. As a result, the board-local 775 nm optics only need to mode match and steer the auxiliary beam into the cavity through M3, after which it exits through M4. This common upstream modulation means that the field-programmable gate arrays (FPGAs) used for the 775 nm PDH loops must use synchronized clocks, so that the demodulation phases remain well defined across the four cavities.

The output optics provide both local diagnostics and the handoff to the next cavity. After M2 and M4, respectively, the transmitted 1550 nm and 775 nm beams are sent to their transmitted photodiodes, with a small pickoff from each path sent to a camera. These camera pickoffs helped during construction because they allowed for monitoring of the spatial mode exiting the cavity. This also helps with alignment, where one can walk the beam based on which resonances are seen flashing on the transmission camera. On the 1550 nm output, a second flipper mirror selects between the local diagnostic path and the chain path. In the diagnostic position, the transmitted light is sent to the 1550 nm transmission photodiode and camera. In the chain position, the same cavity output is coupled toward OFC 2. This allowed the same cavity to be characterized by itself and then connected to the next filter cavity without rebuilding the output optics.

One design note for the first OFC is that the 775 nm path is not counter propagating with the 1550 nm path. The auxiliary light is injected through M3 and exits through M4, while the signal light enters through M1 and exits through M2. This means that the two beams are not traveling in opposite directions, which could have been done by injecting the 775 nm light through M2 and having it exit through M1. For OFCs 2, 3, and 4, the 775 nm path should be counter propagating with the 1550 nm path. The first cavity was built to test the mechanical design and the optical alignment procedure, but it is possible to see the 775 nm auxiliary light in

the transmitted 1550 nm path on the camera. This light is likely to be suppressed by the later cavities, but the counter-propagating geometry is preferred to reduce the possibility of back-scatter from the auxiliary light reaching the SNSPD path.

8.3.5 Alignment of The Output Filter Cavities

The initial alignment of an OFC is a geometric alignment of two different beams into the same four-mirror cavity. The 1550 nm beam enters through M1 and exits through M2. The 775 nm auxiliary control beam enters through M3 and exits through M4. With the beam already mode matched, the first goal is simply to make both beams follow the intended cavity path and hit the centers of the relevant mirrors. A custom alignment tool was used for this step. The tool fits in front of the mirror and is indexed by the geometry of the cavity body in front of the mirror. The tool has a small hole in the center that allows the beam to pass through if it is on the way to hitting the center of the mirror. The tool also has a different hole that allows the beam to be reflected towards the next mirror. Without the tool, the alignment could still be done by visually making sure the beam is centered on the mirrors, but that method is much less precise.

This procedure was written for a cavity where all of the mirrors have angular adjustment, but the first OFC was built with M4 fixed in angle. The compensation for this is described in the relevant steps below. Understanding the process for this alignment procedure will allow for easy adjustments to slight differences in the cavities, such as the fixed M4 angle in OFC 1, or counter propagating paths.

The practical alignment sequence is:

1. Put the cavity body in its final position and set the input beam heights. Start with all of the mirrors removed from the cavity. Use low optical power and block the other wavelength while aligning one beam.

2. Set the input mirrors for both paths. Make sure that the steering mirrors have ample space between them and are adequately separated from the cavity. Align each path's second steering mirror (the last mirror that the beam hits before entering the cavity) so that it is centered along the cavity axis. Imagine a line that goes through the centers of the cavity mirrors and extends beyond the cavity. The last steering mirror before the cavity should be centered on this line. To do this efficiently a laser level can be used to draw a straight line

down the cavity. Both steering mirrors should be at the same height as the cavity.

3. Align both input beams so that they hit the centers of both steering mirrors. This allows for maximum adjustment of the input beam position and angle.

4. Block the 775 nm beam.

5. Start with the 1550 nm path. Install the alignment tool in the M1 position and, using the second steering mirror, steer the incoming 1550 nm beam until it is centered on the M1 reference, so that it hits the center of M1. This fixes the input position of the signal beam.

6. Add the alignment tool to the M2 position. If you did a good job of aligning the second steering mirror, the beam should go through the second alignment tool. If the beam is a little off, use the two steering mirrors together to walk the beam so that it is centered on both alignment tools. This step sets the angle of the beam as it enters the cavity, which defines the path of the beam through the cavity and which parts of the mirror surfaces it hits. The steering mirrors will not be adjusted again until the final alignment of the cavity.

7. Use this opportunity to install the output optics for the 1550 nm path, especially a camera to monitor the transmitted beam. The camera should be placed as close to the cavity output as possible to get the widest possible view of the beam. Center the camera on the beam.

8. Block the 1550 nm beam and unblock the 775 nm beam.

9. Repeat Steps 5–7 for the 775 nm path, using the M3 and M4 alignment tools, and the 775 nm path steering mirrors. This sets the input position and angle of the auxiliary control beam.

10. Block the 775 nm beam and unblock the 1550 nm beam.

11. Remove the alignment tool from M2 and install M2. Double check to see that the 1550 nm beam still looks to be centered on M2.

12. Adjust the angle of M2 so that the reflected beam goes through the M3 alignment tool. This step sets the angle of M2. From this point on, M2 is fixed and should not be adjusted.

13. Block the 1550 nm beam and unblock the 775 nm beam.

14. The 775 nm path should be hitting the center of M4. Remove the M4 Alignment tool and install M4. Adjust the angle of M4 so that the 775 nm reflected beam goes through the M1 alignment tool, meaning that the beam is hitting the center of M1. For OFC 1, M4 cannot be angled directly, so this adjustment is instead made with the mirror immediately in front of M4. In that case M4 is treated as a fixed reference surface, and the upstream mirror is used to set the ray that reflects from M4 toward M1.

15. This step is optional, but helpful for the later steps. Omit this step if you don't have a way of preserving M1's angle. Install M1 and add the M2 alignment tool back. The 775 nm beam should still be hitting the center of M1. Adjust the angle of M1 so that the reflected 775 nm beam hits the center of the M2 alignment tool. Since M2 is installed, if everything has been done precisely, the 775 nm beam should be reflecting to hit the center of the M3 alignment tool and overlapping with the input beam. This step is valuable because it sets the angle of M1, which, when reinstalled, will make the final alignment easier. Remove M1 in a way where you can preserve its angle, and leave the M2 alignment tool in place.

16. Block the 775 nm beam and unblock the 1550 nm beam. The 1550 nm beam should still be hitting the center of M2 and reflecting through the center of the M3 alignment tool.

17. Remove the M3 alignment tool and install M3. Adjust the angle of M3 so that the reflected 1550 nm goes through the center of the M4 alignment tool. This step is the last time when M3 will be adjusted.

18. The 1550 nm beam should now be following the intended cavity path and hitting the centers of M1 (when it is reinstalled), M2, M3 and M4. Make sure that the beam is going back through the M1 alignment tool.

19. Use the laser card to trace the entire path of the 1550 nm beam through the cavity, making sure that it is not reflecting off any of the walls of the cavity body. If the beam is hitting the wall, back track in these steps and start again from the last step where the beam was still clear of the walls.

20. If the transmitted beam has moved on the camera, move the camera so that it is centered again. The beam will be much dimmer because it is passing through M2, which is highly reflective.

21. Remove all alignment tools.

22. Install M1 and increase the exposure time on the camera so that you see the first pass 1550 nm beam.

23. If the angle of M1 was preserved from the optional Step 15, the reflected 1550 nm beam should be very close to overlapping with the input beam and creating a cavity. If this is the case, the camera will see a bright flashing pattern. If the camera does not see a flashing pattern, the angle of M1 needs to be adjusted. Adjust the angle of M1 until a very faint second pass is seen on the camera. The exposure time may need to be increased to see the second pass. Once the second pass is seen, adjust the angle of M1 until the second pass overlaps with the first pass. This will create a cavity and the camera will see a bright flashing pattern.

24. Start to scan the cavity length with the PZT. This will cause the camera to see flashes more frequently helping with alignment.

25. Once the pattern is flashing, use M1 and M4 to walk the beam until the flashing pattern moves to lower order modes, as seen in Fig. 3.5. For OFC 1, use the mirror immediately in front of M4, which is M3, in place of M4 itself. Do not adjust M2 or M3. Adjusting M2 or M3 will change the cavity axis to translate or tilt, meaning that the 775 nm beam aligned earlier will no longer be aligned. Moving the cavity axis also causes the beam to move out of the center of the optics causing potential clipping. The modes that are most likely flashing are Hermite-Gaussian modes. If the pattern is taller than it is wide, the beam needs to be walked in the vertical direction. If the pattern is wider than it is tall, the beam needs to be walked in the horizontal direction. The goal is to get the flashing pattern to be as close to a circular zeroth-order mode as possible. If you see Laguerre-Gaussian modes, the mode-matching probably needs to be adjusted. Continue to walk the beam until the flashing pattern is as close to a circular zeroth-order mode as possible. The flashing pattern will be very bright on the zeroth-order mode; adjust the camera's exposure time accordingly.

26. Once this flashing pattern is seen, set up the 1550 nm reflection PD. The reflection PD will be used to generate the error signal for the PDH lock. On this PD, dips will be seen for every flash on the camera. Keep adjusting the

angle of M1 and M4 until the dips are as deep as possible. For OFC 1, again use the mirror immediately in front of M4 in place of M4.

27. Note where the 1550 nm beam is hitting the camera by adding a crosshair to the camera image. You can do this by blocking the 1550 nm beam inside the cavity just before M4. The crosshair will be used to align the 775 nm beam to the same cavity axis.

28. Block the 1550 nm beam and unblock the 775 nm beam. The 775 nm beam should be following the intended cavity path and hitting the centers of M1, M2, M3 and M4.

29. Look at the 775 nm beam on the 1550 nm camera. With the camera's exposure time greatly increased, the 775 nm beam should be visible on the camera. If you are adjusting these instructions and using this for a counter-propagating 775 nm path, the 775 nm beam will not be visible on the camera. Use the second 775 nm steering mirror to move the 775 nm beam to the same position on the camera as the crosshair, where the 1550 nm spot was. If the 775 nm beam was aligned correctly this angle should not need to be adjusted much.

30. Move the 775 nm camera to see the transmitted 775 nm beam in the center of the camera.

31. Use the two 775 nm steering mirrors to walk the 775 nm beam around while keeping the beam on the 1550 nm crosshair. On the 775 nm camera, a second pass beam should eventually be seen. Adjust the 775 nm steering mirrors until the second pass is overlapping with the first pass on the 775 nm camera which will cause the cavity to start flashing.

32. Still using only the 775 nm steering mirrors, repeat the walking procedure in Step 25, but without touching the cavity's mirrors. The goal is to get the flashing pattern to be as close to a circular zeroth-order mode as possible. If you see Laguerre-Gaussian modes, the mode-matching possibly needs to be adjusted. Continue to walk the beam until the flashing pattern is as close to a circular zeroth-order mode as possible.

33. Once the flashing pattern is as close to the zeroth-order mode as possible, set up the 775 nm reflection PD. The reflection PD will be used to generate the error signal for the PDH lock. On this PD, dips will be seen for every flash

on the camera. Keep adjusting the steering mirrors until the dips are as deep as possible.

34. Block the 775 nm beam and unblock the 1550 nm beam. The flashing should appear on the 1550 nm camera.

35. Repeat the walking procedure that was just done for the 775 nm beam but for the 1550 nm beam using only the steering mirrors. Maximize the depth of the dips on the 1550 nm reflection PD by walking the input beam with the two steering mirrors.

36. To fine-tune the alignment of either path even more, install the transmission PD for the path in question. Lock the cavity to that wavelength using the PDH reflection signal, and look at the transmitted beam on the transmission PD. While monitoring the transmission PD, walk the input beam with the two steering mirrors to maximize the transmitted power. The cavity lock will be very sensitive to changes in the input alignment, so this should be done very slowly. Occasionally it will be necessary to re-lock the cavity and/or maximize the reflection on the reflection PD. Repeatedly walk the beam in small increments until the signal is maximized on the transmission PD.

The same procedure can be adapted when one of the cavity mirrors cannot be angled. The central point is to preserve the defined cavity axis. If M4 is fixed, the adjustment that would normally be made with M4 is moved to the mirror immediately in front of M4. This gives up the clean overlap between the cavity axis and the 775 nm input-beam but if all the steps are followed the compensation needed from the 775 nm steering mirror should be minimal. The same logic applies if M1 is fixed instead of adjusting M1 and M4, now only M3 and M4 are used.

For a counter-propagating auxiliary beam, the same alignment logic applies, but the order of the alignment tools is reversed for that wavelength. For example, if the co-propagating auxiliary path is $M3 \rightarrow M4 \rightarrow M1 \rightarrow M2 \rightarrow M3$, then the counter-propagating path is $M4 \rightarrow M3 \rightarrow M2 \rightarrow M1 \rightarrow M4$. Use the corresponding pair of alignment tools for the actual auxiliary input side.

In this counter-propagating configuration, the 1550 nm transmission camera should not be used as a direct overlap reference for the 775 nm beam, because the counter-propagating beam exits from a different port on that mirror. Instead, check the 775 nm beam with a camera or transmission PD on its own transmission side. After

the 1550 nm alignment has defined the cavity axis, do not move the cavity mirrors. Walk only the auxiliary input steering mirrors until the transmitted 775 nm mode is as close as possible to the zeroth-order mode and the reflection PDH dips are maximized. The counter-propagating version is not a different alignment strategy; it is the same mirror-centering and mode-walking procedure applied to the same cavity axis in the opposite propagation direction.

The LFC alignment is the single-wavelength version of the same procedure. Since the LFC only uses the 1550 nm resonant path, the 775 nm steps are omitted. Instead of using the OFC alignment tools, the beam is checked directly against the centers of the cavity mirrors. For the initial alignment, the mirrors in the M1 and M2 mirror mounts can be temporarily replaced with 1 in. spacers with holes through their centers. If the beam passes through both holes, the input position and angle are centered on the first two cavity mirrors. The remaining steps then follow the same logic as the OFC procedure: define the cavity axis with the fixed mirrors, then use the input steering mirrors and the remaining adjustable cavity mirrors to walk the transmitted pattern toward the lowest-order mode.

8.4 Auxiliary 775 nm Control System

The OFC chain cannot be operated by locking the 1550 nm carrier to the cavities, because that is the light the filter must reject. The auxiliary control system solves this by locking each cavity with frequency-shifted 775 nm light while using the AOMs to choose where the corresponding 1550 nm pass band falls, as seen in Fig. 8.2. Since the four cavities have intentionally different FSRs through their varying lengths, the auxiliary frequencies are cavity-specific rather than copied from one path to the next. The optical preparation of the control fields, the sensing scheme, and the operating states that connect diagnostic measurements to the final auxiliary locks are essential tests of the control system design.

8.4.1 SHG and AOM Sled

The auxiliary control system starts with a pickoff of the main 1550 nm laser. That light is frequency doubled to 775 nm and distributed to the individual cavity paths, as described by Fig. 8.8. After the SHG stage, the light is phase modulated for PDH sensing, split into cavity-specific paths, shifted by AOMs, and coupled into fibers that deliver the control fields to the OFCs. The AOMs are the key part of this sled because they set the auxiliary lock frequency for each cavity and set which 1550 nm sideband is transmitted when the PDH loop is closed. For the initial commissioning

of OFC 1, the 775 nm light was picked off before the LFC system and was not stabilized, but for the design sensitivity the 775 nm light will be derived from the same filtered laser path that feeds the interferometer, after the LFC stage.

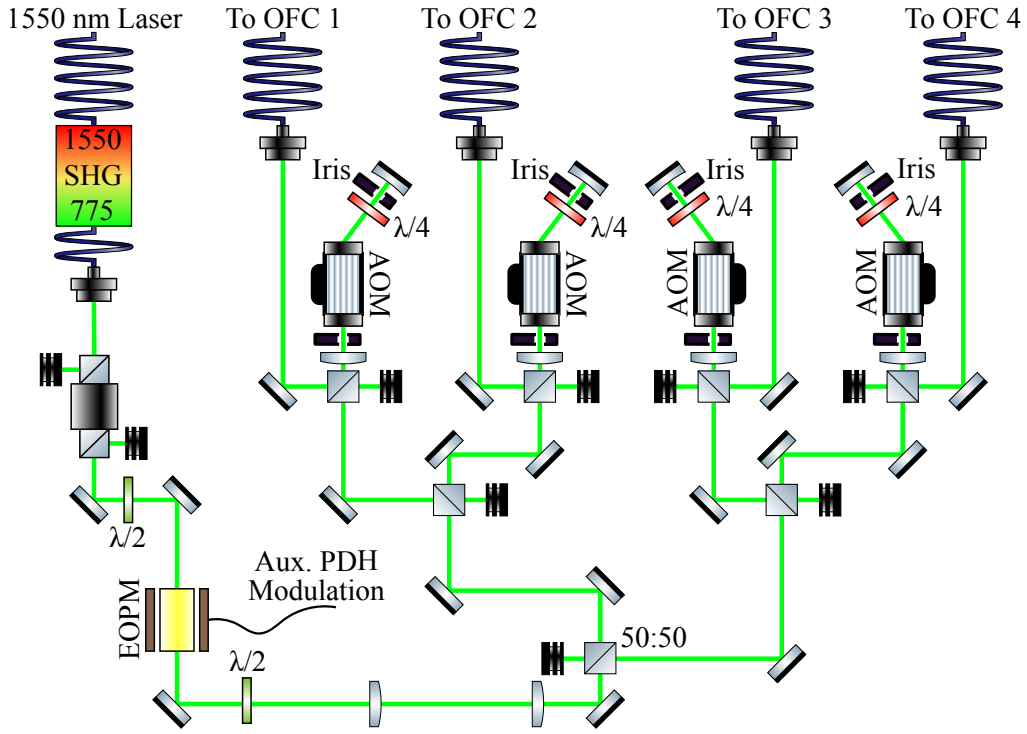


Figure 8.8: Auxiliary 775 nm preparation for the output filter cavities. A pickoff of the main 1550 nm laser is sent through the SHG stage to generate the 775 nm control field, which is shown in green. This output then passes through the common auxiliary PDH modulation path giving each of the auxiliary signals PDH sidebands, before being divided among the four cavity channels. In each channel the beam is mode matched into a double-pass AOM: the first-order diffracted beam is selected with an iris, sent through a quarter-wave plate, retroreflected, and sent through the quarter-wave plate and AOM a second time. The second pass doubles the frequency shift while the polarization rotation separates the returning beam from the input beam at the PBS. The shifted beam is then coupled into a fiber and delivered to the corresponding OFC. The four AOM drives are independent, so each auxiliary beam can be placed on the cavity-specific lock point required to align the common 1550 nm signal-sideband pass band.

Before going through the AOM, the 775 nm light passes through a PBS and a lens to set the beam size for the AOM. Once through the AOM, the beam passes through a quarter-waveplate. Then the zeroth-order beam is blocked by an iris and the first-order beam is sent through. The first-order beam then hits a mirror with a ROC of the same distance as the AOM crystal, which retro-reflects the beam back through the AOM even if the drive frequency changes. Before re-entering the AOM,

the beam passes the quarter-waveplate again, which rotates the polarization by 90° and allows the beam to be separated from the incoming beam with a PBS after the second pass through the AOM. This double-pass configuration is a common way to get a large frequency shift while keeping the beam pointing stable as the drive frequency changes. The final output beam is then coupled into a fiber that delivers it to the cavity's 775 nm input optics.

Since each AOM path is double passed, the delivered frequency shift is twice the AOM drive frequency. The nominal RF drive frequencies for the $n_i = 1$ auxiliary lock points are in the 200-208 MHz range, and laid out in more detail in Section 8.2.2. Only the frequency modulo the cavity FSR matters for the cavity resonance condition, so the large absolute shift is used to place the 775 nm lock point at the required relative detuning.

Because the four cavities have different FSRs, the four AOM paths should be thought of as independent frequency actuators rather than as a shared common offset. A common shift would not place all four auxiliary beams on the lock points needed for the same 1550 nm sideband. During commissioning, the nominal frequencies from Table 8.2 provide the initial settings, and the optical calibration in Section 8.4.5 supplies the final offsets.

8.4.2 PDH Sensing at 775 nm

Each cavity is locked with a PDH signal derived from its reflected 775 nm field. The electro-optic modulator (EOM) on the auxiliary sled applies phase modulation before the cavity-specific paths are separated. The reflected light from each cavity is detected on the reflection PD, demodulated at the common PDH modulation frequency, and filtered to produce a length-control signal.

A future upgrade may be that the common 775 nm modulation is replaced by separate EOM paths for each cavity, so that the modulation frequencies can be different. This would allow for more flexible demodulation and reduce the possibility of cross-talk between the loops. For the initial commissioning of OFC 1, the same modulation frequency was used for all cavities, but the demodulation phases were adjusted to give the maximal lock sign for each cavity. The design choice would be to have different modulation frequencies for each cavity, which would require separate EOM paths but would reduce potential cross-talk. A possible separation frequency for the EOM modulation frequencies would be at least 500 kHz between modulation frequencies.

The PDH loop actuates on the cavity's PZT. When the loop is closed, the cavity length follows the selected 775 nm resonance. Because the 775 nm resonance has been shifted by the AOM, the same lock also places the 1550 nm sideband resonance at the desired offset. This is the core control idea of the OFC chain: the auxiliary field controls the residual length motion, while the AOM determines which 1550 nm frequency is transmitted by that mechanically offset cavity.

The 775 nm lock can also be separated into high frequency and low-frequency control. The low frequency control is sent to the PZT and compensates for long timescale shifts in cavity length from processes like temperature change. The high-frequency control can be sent to the AOM and can change the offset frequency to cancel out any high frequency noise in the loop at a higher bandwidth than is possible with the PZT. This option matters because the PZT has a limited bandwidth, and the AOM can provide a much faster response to high-frequency noise. This division into a low-frequency path to the PZT and a high-frequency path to the AOM is very similar to the control scheme used for the LFC length control, where the low-frequency control is sent to the cavity length and the high-frequency control is sent to the laser frequency, described in Chapter 9.

8.4.3 1550 nm Test Lock

The 1550 nm lock is a diagnostic state used for scans, mode matching, sign checks, alignment, and calibration. In this configuration, the 1550 nm field is resonant with the cavity and a conventional reflected-light error signal is used to lock or scan the cavity. During this operation the 775 nm light may be incident on the cavity or not and the 775 nm light could be locked or unlocked. This state is not the photon-counting operating state, since the final readout requires the carrier to be rejected, as seen in Fig. 8.2. It is nevertheless the simplest way to measure the fundamental-mode transmission, identify higher-order modes, and calibrate the relation between PZT drive and optical frequency before the auxiliary lock point is set with the AOM.

The first cavity measurements used this diagnostic state to establish the 1550 nm finesse and to check the initial alignment and the alignment after removing or adjusting the cavity optics. The same state also provides a cross-check for the auxiliary lock: after the 775 nm loop is closed, the 1550 nm scan or error signal verifies whether the desired signal-sideband resonance is where the AOM setting predicts. It also allows for the 1550 nm to be held on resonance while the 775 nm lock is closed, which is a practical way to see if the 775 nm lock is drifting in relation

to the carrier laser, i.e., checking if ϵ_r is drifting or being held steady.

8.4.4 775 nm Operating Lock

The normal operating state uses the 775 nm PDH loop. In this state the 1550 nm carrier is not used as the lock reference. The cavity is held by the auxiliary field, and the 1550 nm readout light samples the filter transmission set by the AOM offset. This separation is essential because the filter must reject carrier light while staying resonant for the signal sideband.

For the full chain, the relevant lock metric combines individual lock stability with overlap of the four locked pass bands. A residual detuning that is harmless for one cavity can become a measurable signal-transmission loss after four Lorentzian filters are multiplied. This is why the locked spectra are converted into equivalent 1550 nm detuning and compared to the few-kilohertz requirement in Eq. 8.7.

8.4.5 AOM-Assisted Calibration

The AOM frequency provides a clean calibration handle because it offers an absolute detuning reference. The voltage-controlled oscillator (VCO) that controls the AOM's detuning allows for a Volts-to-Hertz calibration of the frequency shift to be set. The double-pass configuration of the AOMs means that the delivered optical frequency shift is twice the AOM drive frequency.

This means that the auxiliary frequency shift can be treated as a calibrated actuator without first sweeping across the optical resonance. Once the VCO tuning coefficient is known, a voltage change on the AOM path corresponds to a known frequency change in the delivered 775 nm light. The double pass simply doubles that conversion. This is the calibration used to convert the AOM error-signal spectra in Section 8.6.3 into equivalent cavity motion.

The 1550 nm diagnostic signal is then used as a consistency check rather than as the source of the calibration. If the cavity is held on the selected 775 nm lock point, the 1550 nm transmission or error signal verifies whether the corresponding signal-sideband pass band is at the expected detuning. The nominal AOM frequencies from Table 8.2 set this operating point from the designed lengths and FSRs. Small adjustments can still be made during commissioning, but they are corrections to a known frequency setting, not a separate scan calibration.

For OFC 1, this check was implemented in the dual-resonant testing state in Fig. 8.9. In this state the 775 nm and 1550 nm fields are both resonant with the cavity. The

cavity length can be controlled from either the 775 nm PDH signal or the 1550 nm error signal, while the unused error signal remains available as an out-of-loop diagnostic of the other wavelength's resonance condition. The AOM is driven by a VCO, so the auxiliary lock point can be set or held by the AOM controller without changing the optical alignment. This state is a commissioning and calibration state, not the final photon-counting operating state.

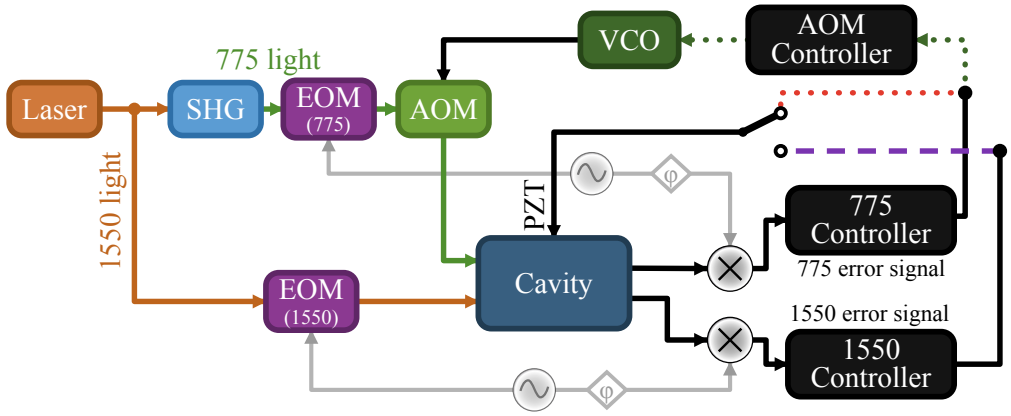


Figure 8.9: Control layout used for OFC 1 in the dual-resonant testing state, where the 775 nm and 1550 nm fields are both resonant with the cavity. Either error signal can be selected to drive the cavity PZT, while the unselected error signal remains available as an out-of-loop diagnostic. The AOM controller drives the VCO that sets the auxiliary frequency, allowing the selected 775 nm lock point to be checked against the measured 1550 nm resonance.

8.4.6 Control Configurations

The commissioning control states share the same hardware but use different references and actuator paths. The 1550 nm test lock, from Section 8.4.3 is used to characterize the optical filter directly. The 775 nm operating lock, from Section 8.4.4 is used for sideband filtering during full operation. The calibration state is the dual-resonant OFC 1 layout in Fig. 8.9: it combines the two readouts so that the AOM offset can be adjusted with respect to the measured 1550 nm or the 775 nm resonance condition. The final operating diagram, shown in Fig. 8.10, is the state in which the 775 nm loops hold the cavities and the 1550 nm carrier is rejected.

In practice these states are used as an ordered commissioning sequence. The 1550 nm state identifies the fundamental spatial modes in scans and verifies the optical linewidth. The calibration state then checks that the selected AOM frequency places the auxiliary lock point at the expected 1550 nm detuning. Finally, the operating state leaves the cavity locked on 775 nm while the 1550 nm field is used only

as the filtered signal path for commissioning. This sequence keeps the diagnostic measurements tied to the final photon-counting configuration.

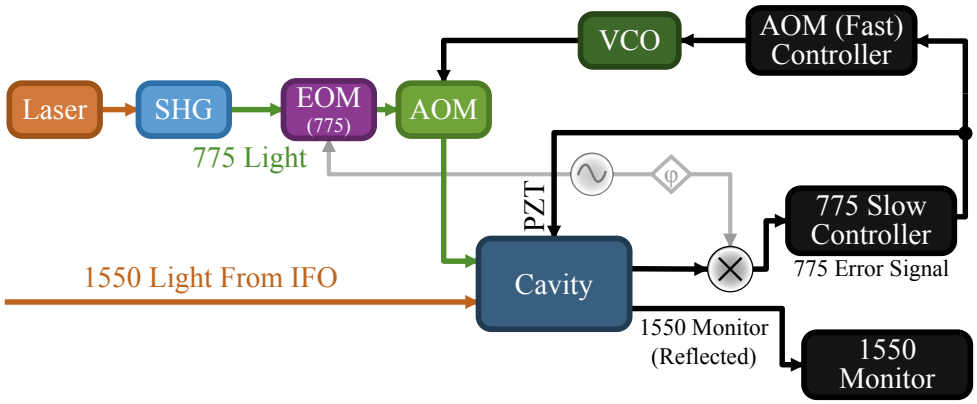


Figure 8.10: Operating control configuration for OFC 1. The 775 nm auxiliary field is prepared and frequency shifted by the system in Fig. 8.8 before entering the cavity. The reflected auxiliary field is demodulated to form the 775 nm error signal, which is divided between two feedback paths, as in the LFC control scheme shown in Fig. 9.2. The slow controller actuates on the cavity PZT to correct long-timescale length motion, while the fast controller drives the VCO and therefore the AOM frequency to suppress laser fluctuations beyond the PZT bandwidth. The 1550 nm field enters directly from the interferometer and is not used as the operating lock reference; its reflected signal is retained as a monitor of the carrier-rejection condition. The same architecture is repeated for the remaining OFCs, with an independently chosen AOM frequency for each cavity so that the four signal-sideband pass bands overlap.

8.4.7 CDS and FPGA Control

Bench-top controllers were used for early locking and loop development, while the final experiment benefits from a more reproducible and standardized control interface. The CDS/FPGA implementation moves the demodulation, filtering, lock-state sequencing, and monitoring signals into a unified experimental control system. This is primarily an operational improvement rather than a change in the optical principle. The same PDH error signals and PZT actuation paths are used, but the lock acquisition and state transitions become repeatable across all cavities.

That repeatability becomes more critical as the chain grows. Four cavities require four demodulation channels, four PZT feedback loops, four AOM frequencies, and enough diagnostic channels to see whether a loss of transmission is caused by detuning or a local lock failure in a specific cavity, or just a misalignment of the cavities to one another. Moving these states into a unified control system

provided by CDS also makes it possible to save exact experimental settings and loop configurations.

8.5 Suppression and Higher-Order-Mode Modeling

The design values above describe the individual cavities, but the photon-counting readout depends on the product of four cavities. When working together the interactions between the OFCs matter. The OFCs length offsets, transverse-mode structure, and mode mismatch combine when the cavities are chained together. This is especially relevant for carrier light in higher-order spatial modes, since a mode that leaks through one cavity is only dangerous if it also overlaps the pass bands of the remaining cavities.

Modeling the full chain, in the frequency domain, is done using the Finesse package, and gives a more complete picture of which offsets are safe and which offsets require more careful HOM or carrier-leakage measurements [145]. The model also provides the context for interpreting the single-cavity linewidth and locked spectra presented in the following section.

8.5.1 Higher-Order-Mode Generation and Transmission

Higher-order-mode leakage was modeled with a full four-cavity calculation. Each cavity is a detuned OFC design represented as a four-mirror bow-tie with input and output couplers near 883 ppm at 1550 nm, high-reflecting end mirrors, a 3° incidence angle, and the same nominal curved mirror ROC of -1.6 m. The four cavities are assigned slightly different round-trip lengths, matching the offsets in Table 8.2, so that their longitudinal spectra do not overlap perfectly. The transverse-mode spectra then shift through the changed FSRs and the resulting Gouy-phase placement of each mode family.

The length offsets alone do not create higher-order-mode power. They determine where the transverse-mode resonances land in frequency once some light has been coupled into those modes. To populate the modes in the model, the input field is deliberately mismatched to the cavity eigenmode at the input node. The misalignment consisted of a 10% waist-size mismatch and a -20% waist-position mismatch. This is a controlled stand-in for imperfect mode matching or alignment of the injected beam: it puts a small amount of carrier power into higher-order spatial modes so that their resonant leakage can be tracked. The calculation includes transverse modes through $m + n \leq 10$.

The resulting transmission of each cavity is then multiplied to obtain the full-chain transmission. The product is what matters experimentally because a photon reaches the detector only if it is transmitted by every cavity in series. In this version of the model, the mirror radii of curvature are kept common across the cavities, so the rejection in the product trace is primarily the consequence of the assumed input mode mismatch, the length staggering, and the cavity-specific auxiliary lock points.

For a stable cavity, the transverse-mode frequencies can be written schematically as

$$f_{qmn} = f_{\mathrm{FSR}} \left[q + \frac{(m + n + 1)\psi_{\mathrm{rt}}}{2\pi} \right], \tag{8.11}$$

where q is the longitudinal index, m and n are transverse indices, and ψ_{rt} is the round-trip Gouy phase. This equation illustrates why length staggering helps even when the mirror radii are the same. Changing L_{rt} changes f_{FSR}, and it also changes where a fixed transverse-mode family lands on the frequency axis. The AOM offsets then recover only the desired fundamental pass band; they do not force the neighboring longitudinal and transverse resonances to overlap.

The result supports the frequency-allocation strategy discussed in Section 8.2.3. The lower edge of the practical tuning range is set by the need for enough carrier rejection. The upper edge is set by the growing risk that carrier-frequency light in higher-order modes overlaps the output pass band. For the approximately 125 MHz FSR design, this gives a practical tuning range of roughly 8-40 MHz. Within that range, the selected readout offset is chosen to avoid both large interferometer thermal-noise peaks and strong filter-chain leakage features.

8.5.2 Effect of Varying Mirror Radius of Curvature

The baseline HOM model uses the same curved-mirror radius, $R_c = -1.6$ m, in all four cavities. This is the simpler design because it makes the cavities replicas apart from the deliberate round-trip length offsets. It also makes the origin of the rejection easier to interpret: the total trace in Fig. 8.11 is mainly set by the different FSRs and by the AOM offsets that recover the desired fundamental pass band.

To implement this in an adjustable way, the curved mirrors could be mounted inside compression rings that can tighten or loosen the mirror substrate to change the effective ROC. This is a common technique for tuning the ROC of cavity mirrors, and it can be used to intentionally shift the transverse-mode spectrum of each cavity [146–148]. The same effect could also be achieved by using thermal expansion to slightly curve the mirrors [149] or by using different curved mirrors

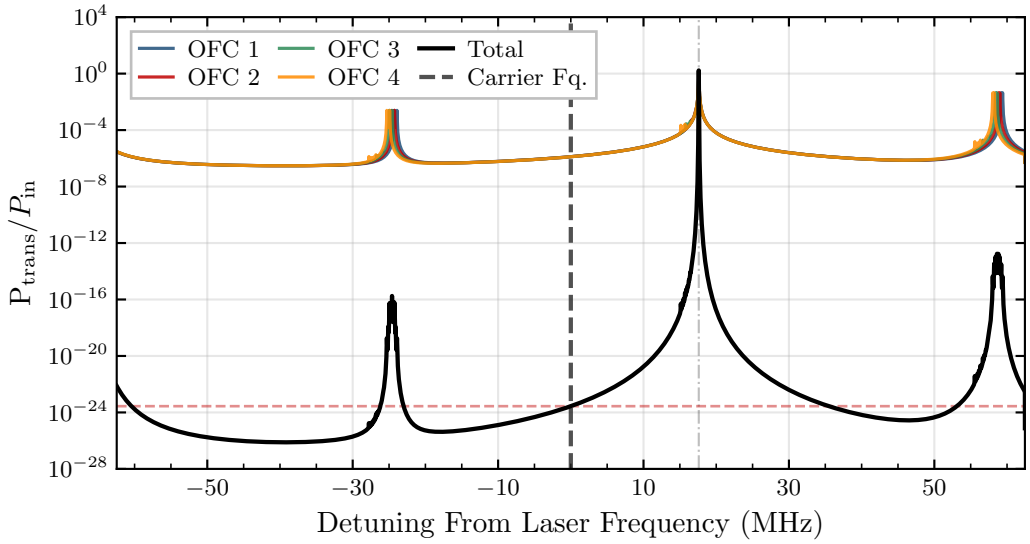


Figure 8.11: Finesse simulation of HOM leakage through the four-OFC chain with transverse modes included through $m + n \leq 10$. The input field is deliberately mismatched to the cavity eigenmode with a 10% waist-size mismatch and a −20% waist-position mismatch so that higher-order spatial modes are populated. Colored traces give the transmission of the individual OFCs for the constant-$R_c = -1.6$ m model; the black trace is their product. The vertical dotted line marks the carrier frequency, and the red dashed line marks the calculated chained carrier suppression at the carrier, 2.8×10^{-24}. Length staggering keeps many single-cavity transverse-mode resonances from lining up in the full chain, but any residual peaks in the product trace represent offsets where carrier light in HOMs can leak through the filter system.

in each cavity, but the adjustable option allows for an adjustable effect on the HOM features.

Adding a variable ROC allows for greater control of the HOM separation. To study this effect, a second HOM model was made with small cavity-to-cavity changes in R_c. The values used for this comparison are listed in Table 8.4. This is not a separate baseline design; it is a sensitivity study of how intentional changes in the curved mirrors would move the transverse-mode spectrum. Changing R_c changes the round-trip Gouy phase, so the transverse-mode families shift relative to the fundamental resonance. The accumulated round-trip Gouy phases in Table 8.4 were calculated from the round-trip ABCD matrix using the convention in Eq. 3.29. Because the bow-tie has a nonzero angle of incidence, the weak astigmatism gives slightly different phases in the two transverse planes. In the full chain this can either help or hurt, depending on whether the shifted HOMs move away from or toward another cavity's transmitted modes.

Cavity	L_{rt}	R_c	ζ_x	ζ_y
OFC 1	2.40000 m	−1.6000 m	120.14°	119.86°
OFC 2	2.40847 m	−1.5957 m	120.76°	120.48°
OFC 3	2.41693 m	−1.5915 m	121.38°	121.10°
OFC 4	2.42540 m	−1.5873 m	122.01°	121.73°

Table 8.4: Round-trip lengths, curved-mirror radii, and accumulated round-trip Gouy phases used in the varying-R_c HOM study. These values are a comparison set for testing curvature variation rather than the baseline OFC design. The Gouy phases were calculated from the round-trip ABCD matrix using Eq. 3.29; the two phase columns correspond to the weakly astigmatic transverse planes of the 3° bow-tie geometry.

Figure 8.12 compares the two chain transmissions. The fundamental pass band remains aligned in both models, as expected, while the nearby HOM product feature changes when R_c is varied. This comparison separates the rejection that is robust to the intentional length staggering from the leakage features that depend on the detailed transverse-mode spectrum of the built cavities.

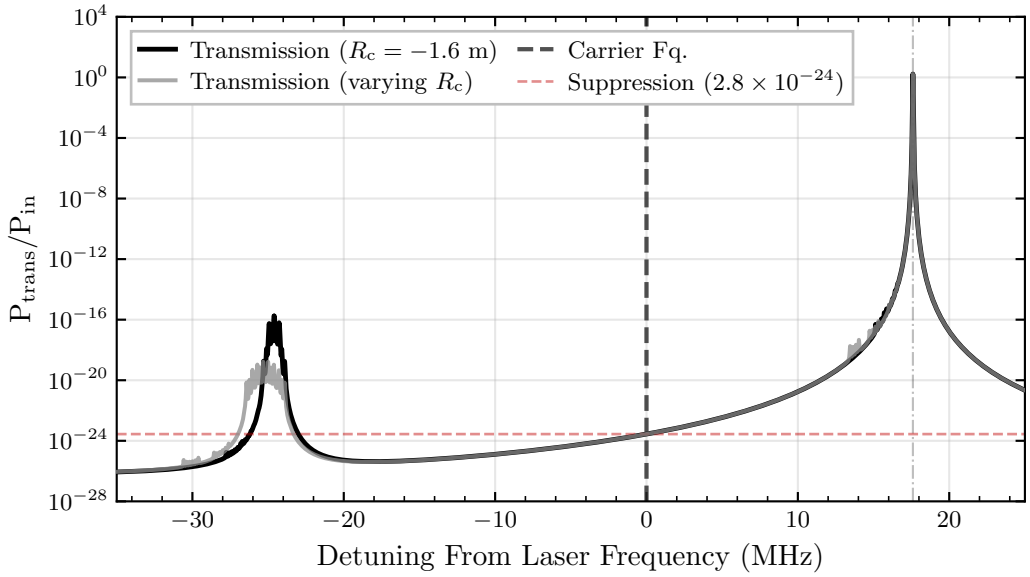


Figure 8.12: Product transmission for the four-OFC chain. The black trace is the baseline model with a common curved-mirror radius $R_c = -1.6$ m; the grey trace uses the first varying-R_c sensitivity model with the comparison values from Table 8.4. The individual cavity transmissions are omitted.

Figure 8.12 indicates that the added ROC variation can move the nearby HOM feature by a few megahertz. This is a significant shift relative to the measured 36 kHz linewidth of OFC 1, and it is large enough to change the expected leakage

at the chosen readout offset. For the varying-ROC values used here, the HOM peak is more spread out and shifted slightly away from the readout offset, reducing the expected leakage at that frequency. However, since the noise spectra in Fig. 3.11 contain many strong peaks, spreading the band out is not necessarily better than keeping it more compact with a higher transmission. The HOM features are sensitive to the cavity geometry, but the length staggering is the more robust way to keep them from lining up in the full chain. The ROC variation is a possible upgrade path if the measured HOM features are too close to the readout offset, but it is not part of the baseline design because it adds complexity and because the length staggering already provides a significant amount of rejection.

8.6 OFC 1 Characterization

The first cavity was used to turn the design values into measured operating numbers. Its measurements establish the achieved 1550 nm finesse, the scan-based mode-matching diagnostics, and the residual motion seen by the auxiliary lock. These measurements are not the full four-cavity readout, but they test the ingredients that matter for the chained system: the linewidth of a single pass band, the ability to identify the fundamental mode, and the ability of the auxiliary lock to hold that pass band at a chosen detuning.

8.6.1 Finesse Measurements

The first performance measurement of OFC 1 was the 1550 nm finesse. The cavity was locked, the input light was shifted rapidly by sending an impulse to the laser frequency, and the transmitted decay was fit with an exponential. The conversion from optical-power decay time to finesse follows from the storage-time relation in Eq. 3.22; the corresponding linewidth follows from the general cavity relation in Eq. 3.24.

The post-cleaning ringdowns gave

$$\mathcal{F}_{1550,\mathrm{OFC1}} = 3472 \pm 13. \tag{8.12}$$

For $f_{\mathrm{FSR}} \simeq 125$ MHz, this corresponds to

$$\Delta f_{\mathrm{FWHM,OFC1}} = 35.98 \pm 0.14\ \mathrm{kHz}. \tag{8.13}$$

The measured linewidth is slightly narrower than the 42 kHz design linewidth used in the photon-counting filter estimates. This is favorable for carrier rejection because the carrier is farther out on the Lorentzian tail, but it also makes the pass-band alignment requirement slightly more demanding.

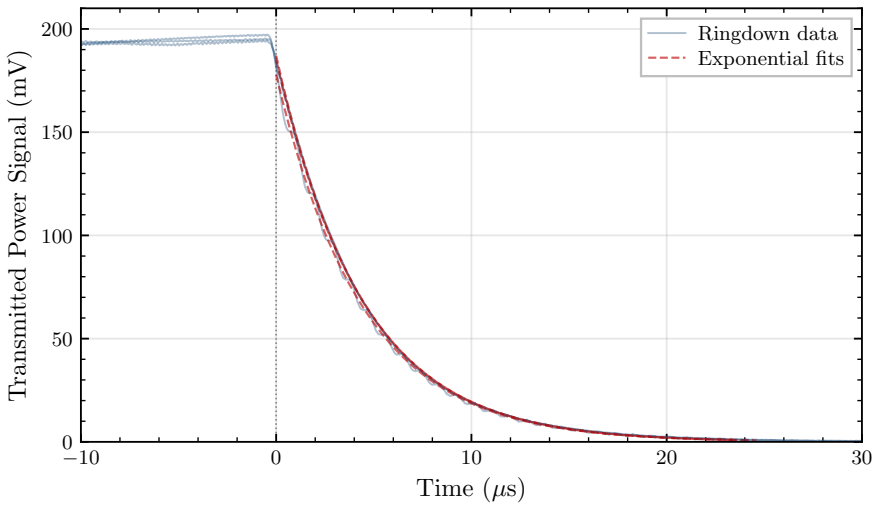


Figure 8.13: Measured 1550 nm ringdowns of OFC 1 after cleaning with First Contact. The decay of transmitted power is fit with an exponential and converted to finesse using the storage-time convention in Eq. 3.22. The three fits give a mean power decay time of 4.42 μs, corresponding to $\mathcal{F}_{1550} = 3472$ and a linewidth of 35.98 kHz for a 125 MHz FSR.

The lower-finesse 775 nm cavity response was measured separately because it sets the auxiliary-lock slope and capture range rather than the photon-counting pass-band width. The scan in Fig. 8.14 was calibrated with the 775 nm FSR from a cavity scan, and an isolated peak was fit with a Lorentzian. The fit gives

$$\Delta f_{775,\mathrm{OFC1}} = 393\ \mathrm{kHz}, \tag{8.14}$$

corresponding to

$$\mathcal{F}_{775,\mathrm{OFC1}} = 318. \tag{8.15}$$

This agrees with the coating expectation: the two 1% loss couplers give a finesse near 310 for a cavity with small additional round-trip loss. The auxiliary resonance is therefore much broader than the 1550 nm signal resonance, as intended for a robust operating lock.

The ringdown is the most direct linewidth measurement because it does not rely on the linearity of a voltage scan. Once the input light is extinguished, the decay time is set by the total round-trip loss of the cavity. This makes the measurement sensitive to the same mirror transmission, absorption, scatter, and clipping losses that determine the filter bandwidth during operation. In contrast, a driven cavity scan has to be calibrated from voltage to frequency and can be distorted by thermal response or by a changing scan speed. For this reason, the ringdown value is used as the linewidth of OFC 1 and the scan fits are used mainly as consistency checks.

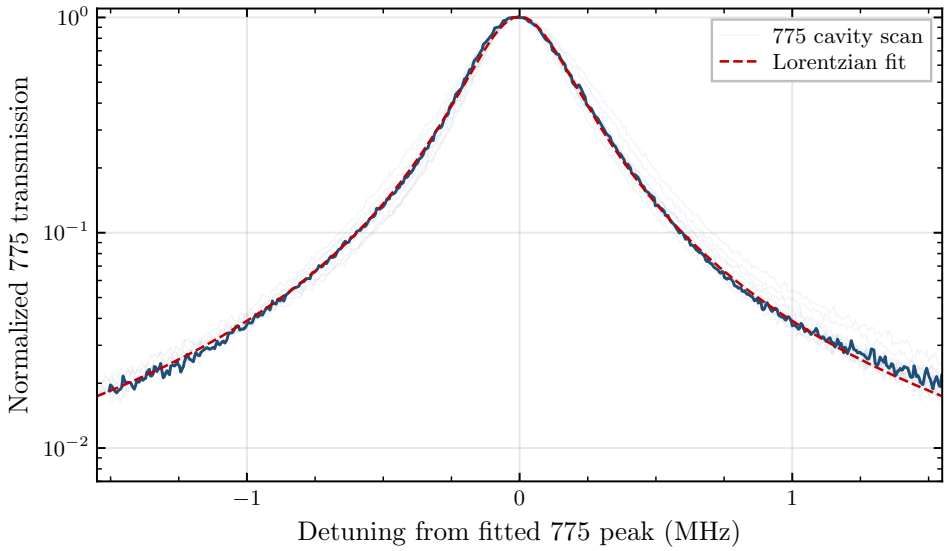


Figure 8.14: Measured 775 nm auxiliary-cavity transmission peak for OFC 1. The light blue traces are repeated falling-sweep scan segments after calibrating the scan axis with the 775 nm FSR. The dark blue trace is a representative scan, and the dashed red curve is a Lorentzian fit to the peak. The fit gives a FWHM linewidth of 393 kHz, corresponding to $\mathcal{F}_{775,\mathrm{OFC1}} = 318$ for the 125 MHz FSR.

8.6.2 Peak Scans and Mode Matching

Peak scans were used to identify the fundamental resonance, check the alignment, and measure the relative coupling to higher-order modes near the operating mode. A clean peak scan contains an isolated fundamental peak with smaller neighboring transverse-mode peaks. These measurements are strongest as alignment and mode-matching diagnostics; the ringdown measurement remains the cleaner finesse measurement because it does not require a calibrated sweep rate. The operating frequency calibration used for locked residual-motion measurements comes from the AOM path rather than from a full-cavity scan.

The scan in Fig. 8.15 plots the 1550 nm transmission of OFC 1 near the fundamental resonance. The scan voltage was calibrated using the observed spacing of repeatable resonances, with the 1550 nm FSR corresponding to about 125 MHz for the 2.4 m round-trip length. The median scan peak has an apparent FWHM of about 0.91 nm in round-trip length, or about 74 kHz. If this scan width is converted directly to a finesse, it gives approximately $\mathcal{F} \simeq 1700$. This is broader than the ringdown result in Fig. 8.13. The ringdown measurement is used as the cavity linewidth, while the scan linewidth is treated as a diagnostic of the sweep calibration, thermal response, and mode shape during a driven scan.

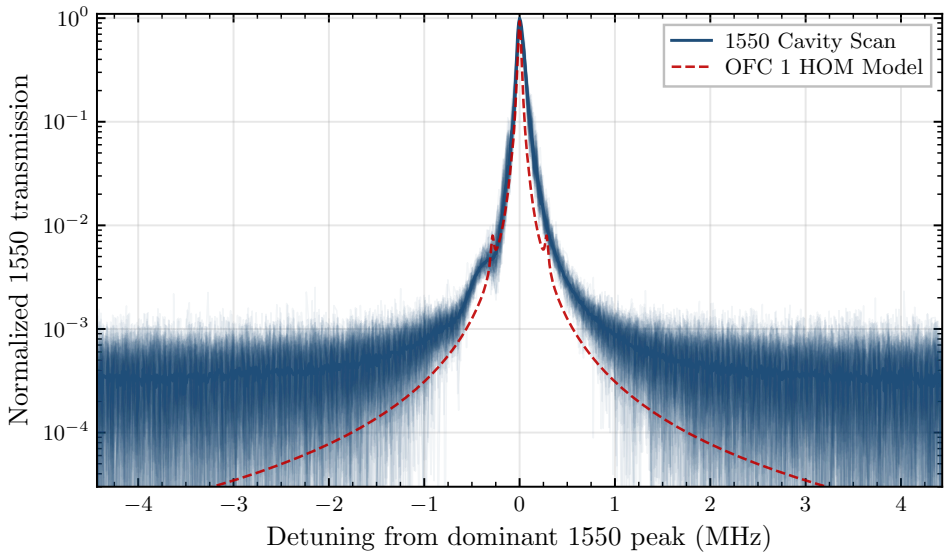


Figure 8.15: Measured 1550 nm cavity scan of OFC 1 near the fundamental resonance. The blue traces are repeated scan segments after converting the PZT sweep to detuning using the measured FSR. The dashed red curve is the OFC 1 higher-order-mode model from the Finesse calculation discussed in Section 8.5.1. The scan linewidth serves as a consistency check, while the ringdown in Fig. 8.13 gives the cleaner finesse measurement.

The logarithmic vertical scale in Fig. 8.15 is notable because it makes the weak transmission away from the fundamental peak visible. The repeated scans overlap well at the central TEM_{00} resonance, indicating that the peak position was stable enough for the scan to be used as a mode-matching diagnostic. The small structure on the negative-detuning side of the main peak is asymmetric. It appears where the single-cavity HOM model predicts nearby transverse-mode transmission, and it is not mirrored with the same amplitude on the opposite side of the resonance. That asymmetry is consistent with higher-order spatial modes rather than a simple broadening of the fundamental Lorentzian. The scan demonstrates where imperfect mode matching can leak carrier-frequency light through a single cavity.

The model curve in Fig. 8.15 is the single-cavity version of the full-chain HOM calculation in Fig. 8.11. In the model, a deliberate mode mismatch populates transverse modes so that the expected HOM leakage can be compared to the measured scan. The comparison uses both the linewidth of the fundamental peak and the positions of the small neighboring spatial-mode resonances relative to the carrier. These scans explain why mode matching matters for the full chain. Carrier-frequency light in higher-order spatial modes can pass through one cavity at a different detuning than the fundamental carrier. The output-filter chain relies on both good mode matching

and the four-cavity product transmission. A mode that is transmitted by one cavity is not a problem unless it happens to overlap with a resonance of the later cavities.

The small shoulders near the fundamental peak provide information about the cavity rather than being dismissed as imperfections in an otherwise clean scan. Their positions check the Gouy-phase model, while their amplitudes give a practical measure of the mode mismatch entering the cavity. The comparison to the Finesse model is intentionally qualitative at this stage, because the exact HOM amplitudes depend on alignment and injected beam shape. The more robust information is whether the observed HOMs appear on the side of the fundamental peak where the model predicts them, and whether their detunings are close enough to matter for the chosen readout offset.

8.6.3 Locked Noise Spectra

A locked spectrum of OFC 1 was measured to evaluate the stability of the auxiliary control loop and the residual length noise of the cavity. The most relevant quantity is the residual detuning expressed in units of the 1550 nm linewidth, since this determines how much the sideband pass band moves during operation. The locked spectrum also identifies mechanical resonances in the cavity body, mirror mount, and PZT assembly that can limit the control bandwidth.

Figure 8.16 plots the 775 nm error signal measured before the controller while the cavity was locked to the auxiliary field. The calibration comes from the AOM, which provides an absolute frequency reference. Because the beam is double passed through the AOM, the measured voltage was converted using 80 MHz/V and then expressed as an equivalent single-mirror displacement. To compare with the cavity-length requirement from the beginning of the chapter, this displacement is multiplied by two because the cavity length changes by twice the mirror displacement.

The integrated residual motion is directly comparable to the scale set in Section 8.1.4.

The relative pass-band requirement in Eq. 8.7 corresponds to about 0.049 nm RMS in equivalent cavity length, or roughly 4 kHz RMS at 1550 nm. The measured 775 nm residual is 0.048 nm RMS in the same cavity-length convention, corresponding to 3.9 kHz RMS of 1550 nm pass-band motion. This is at the design alignment scale for one cavity, but it is not yet the final four-cavity relative stability measurement.

Since the measurement is made with the 775 nm auxiliary field, the 1550 nm pass band should follow the same cavity-length motion. In practice the 1550 nm resonance doesn't track the 775 nm lock point perfectly because the 775 nm light

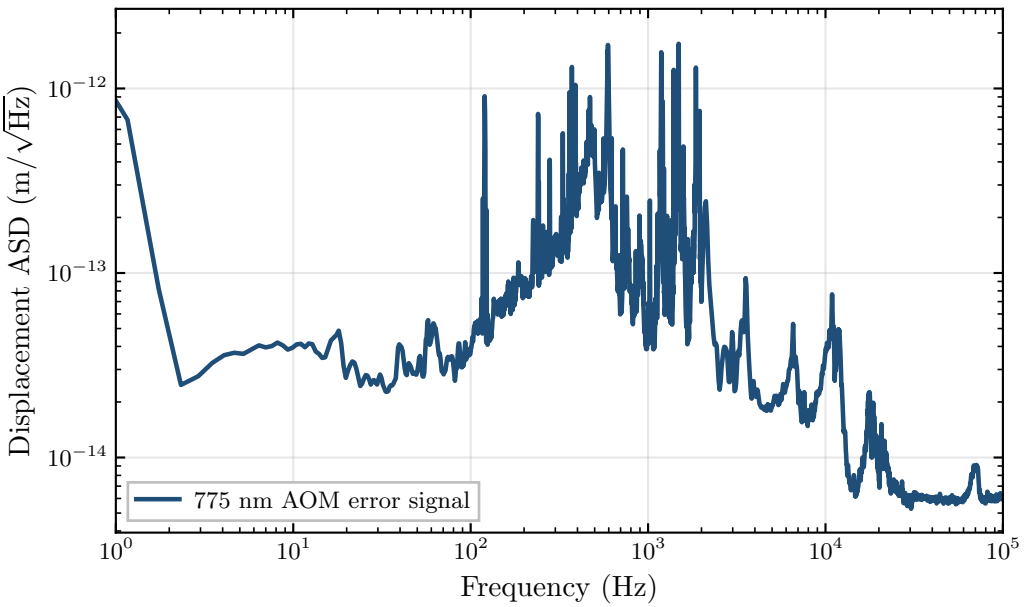


Figure 8.16: Pre-controller 775 nm error signal for OFC 1 while the cavity was locked to the auxiliary field with the PZT. The trace is calibrated with the double-pass 80 MHz/V AOM frequency reference and converted to single-mirror displacement. Integrating the spectrum and converting from mirror displacement to cavity-length motion gives 0.048 nm RMS.

can pick up more frequency noise as it goes through slightly different preparation stages. The result should therefore be interpreted as a calibrated single-cavity lock measurement: it demonstrates that the auxiliary lock is quiet enough to be in the correct range, while the final chained OFC measurement must still verify the relative overlap of the four 1550 nm pass bands directly.

Since the AOM selects the mean detuning and the PZT loop suppresses motion about that detuning, the static frequency-alignment model and the dynamic locked spectrum are complementary parts of the same commissioning check.

8.7 Integration with the Photon-Counting Readout

After the individual cavities are built, characterized, and modeled, the remaining question is how the chain connects to the photon-counting measurement. The practical integration problem is to align the cavities in series, maintain the auxiliary locks, preserve the transmitted spatial mode, and deliver the final filtered light to the detector path without reintroducing carrier leakage or control-light backgrounds. The frequency strategy developed above is what makes this possible: the mechanical lengths keep unwanted resonances separated, while the AOMs move only the selected auxiliary lock points needed to recover a common 1550 nm pass band.

8.7.1 Cascaded OFC Operation

The output of each OFC is mode matched into the next cavity. The chain is aligned by first optimizing the transmission of each cavity individually and then closing the auxiliary locks in sequence. Once the locks are closed, the cavity-specific AOM offsets are adjusted so that the 1550 nm pass bands overlap at the desired sideband frequency. The measured pass-band alignment, carrier suppression, and transmitted mode shape then determine whether the chain is ready for photon-counting operation.

The cascaded geometry has a diagnostic advantage: reflected light from the first cavity remains available for classical readout and interferometer control, while transmitted light from the last cavity is sent toward the photon-counting detector. This separation preserves the ability to control the interferometer without sending carrier-frequency light to the SNSPD.

Commissioning the chain is a staged measurement rather than a single lock acquisition. Each cavity is first treated as an individual filter so its linewidth, mode matching, and auxiliary lock can be checked. Adjacent cavities are then locked together to verify that their pass bands overlap and that no large HOM feature appears at the chosen offset. The full chain is the final product measurement: the desired sideband should remain transmitted, while carrier-frequency light and nearby transverse modes should be suppressed by the multiplication of the four cavity responses.

8.7.2 Interface to the SNSPD Path

The final optical mode transmitted by the OFC chain is the input to the single-photon detector path. The transmitted beam must be coupled cleanly into the detector mode, with minimal room-light, scatter, and blackbody leakage. For this reason the final cavity may have to be mounted in a vacuum chamber with a direct fiber coupling to the SNSPD. The fiber coupling provides spatial filtering that can further suppress any residual carrier leakage or higher-order-mode content, while the vacuum environment reduces the risk of dust and scatter near the detector.

The cavity lids and acoustic enclosures reduce optical pickup around the cavities from scattered light, while the downstream detector system provides the cryogenic filtering and shielding required for low dark-count operation.

The main calibration for the detector interface is the conversion from transmitted optical power to equivalent photon flux in the selected pass band. That conversion uses the same filter transmission and photon-flux normalization described in the

background and LFC chapters. Reporting the OFC result as an equivalent count rate makes the optical characterization directly comparable to the GQuEST photon-counting noise budget.

8.7.3 Background and Leakage Checks

Several leakage mechanisms have to be checked before the OFCs can be treated as a photon-counting filter rather than only as a set of optical cavities. Fundamental-mode carrier leakage is the most direct one: it can be measured by changing the amount of carrier-frequency light incident on the chain and measuring the transmitted power or downstream count rate. Higher-order-mode leakage is tested differently, by deliberately mismatching or scattering the input beam and measuring whether any transverse-mode features survive the full four-cavity product transmission. These measurements are complementary because the first tests the Lorentzian carrier rejection, while the second tests whether light has found another spatial mode through the chain.

The auxiliary-control system introduces a different class of backgrounds. The 775 nm light should not reach the SNSPD path, and it should not generate correlated 1550 nm photons through scatter, fluorescence, or nonlinear conversion in nearby optics. A direct check is to vary the 775 nm control power while keeping the cavities locked and look for a correlated change in the downstream photon rate. Room light, blackbody photons, and scattered 1550 nm light are checked with the cavity lids, baffles, and detector shielding in their final configuration. The goal is to measure these backgrounds in the same state used for data taking, because a background that is harmless during an open-table alignment can become significant after the signal path is coupled to the SNSPD.

8.7.4 Upgrades and Operating Upgrades

The first OFC established the cavity build and control procedure, while the chained cavities establish the readout margin. The main operating requirements are the carrier-rejection requirement, the pass-band alignment requirement, the higher-order-mode leakage requirement, and the control-light background requirement. The first three are set by the cavity linewidths, fixed length offsets, and calibrated AOM settings. The last is set by the optical isolation between the 775 nm auxiliary system and the 1550 nm detector path.

The natural upgrades are not necessarily optical: improved scatter control around the first cavity, baffling and scattered light reduction near the final detector path, more

reproducible AOM state control, and a full CDS/FPGA automation implementation of the auxiliary locks. These upgrades do not change the optical concept. They improve the stability and repeatability of the already demonstrated cavity chain. Transitioning from a single cavity test environment where the main purpose was to test and rapidly prototype the locking and measurement techniques, to a full four-cavity chain where the main purpose is to maintain stable performance over long time scales and produce repeatable results, naturally shifts the focus from optical design to operational stability.

The primary near future upgrade is transitioning control to a centralized CDS system controlling many individual FPGA systems. The same error signals and actuators are used, but the lock acquisition, state sequencing, and monitoring become more repeatable across all cavities. This repeatability is essential for maintaining the delicate frequency alignment of the four-cavity chain, especially as the number of control channels grows. The control systems will need to offer better control loops for the most critical DOFs using advanced control software like Buzz, discussed in Chapters 6 and 7 [150]. This would allow the control loops to be more robust against the mechanical resonances seen in the locked spectra, and it would allow the loop shapes to be optimized for the best noise suppression in the photon-counting band.

8.8 Summary

The OFCs provide the narrowband optical filter that makes photon-counting readout possible in GQuEST. Four bow-tie cavities in series transmit the selected signal sideband with a detuning of ϵ_r while suppressing carrier-frequency light by more than 200 dB. The design uses four approximately 2.4 m round-trip lengths, a linewidth of 42 kHz, and a relative four-cavity pass-band alignment requirement of about 4 kHz RMS. The cavities are intentionally length staggered so that neighboring longitudinal and transverse resonances do not line up across the chain and create a single passband optical filter.

OFC 1 established the construction and measurement bases for the other three. Through ringdown measurements it gave a 1550 nm finesse of 3472 ± 13, corresponding to a linewidth of 35.98 ± 0.14 kHz. A separate 775 nm scan gave an auxiliary finesse of 318, consistent with the broader resonance expected from the 775 nm couplers. The same cavity established the 775 nm auxiliary PDH lock, the AOM-set detuning procedure, and the locked-noise measurement used to verify

pass-band stability. The four-cavity frequency-alignment and higher-order-mode models explain how fixed length offsets and independent auxiliary lock points allow the desired sideband resonance to be aligned while unwanted carrier and transverse-mode resonances remain rejected.

Chapter 9

LASER FILTER CAVITY

"The rest is silence"

— William Shakespeare, *Hamlet*

The LFC is the input optical filter used to suppress laser amplitude and phase noise before the light reaches the GQuEST interferometer. The need for this cavity comes directly from the photon-counting readout. The OFCs reject carrier light at the laser frequency, but they transmit light near the signal sideband frequency $\nu + \epsilon_r$. Laser phase noise at that same sideband offset would pass through the output filters and produce photon counts that look like signal photons.

For the GQuEST laser system, amplified spontaneous emission produces a broadband white phase-noise floor [1, 69] of order

$$S_{\phi,\mathrm{in}}^{\mathrm{LP}} \sim \left(10^{-7}\,\frac{\mathrm{rad}}{\sqrt{\mathrm{Hz}}}\right)^2. \tag{9.1}$$

Without filtering before the interferometer, this laser phase noise would produce a photon-flux rate near 2×10^8 Hz after the output filters, far above the expected signal photon flux [1]. These photon-flux rates, written in Hz, use the relative-power-to-photon-flux conversion and output-filter pass-band integral in Eqs. 9.16 and 9.17. The LFC alone reduces the laser-noise photon flux to about 2×10^2 Hz, while the PRC alone reduces it to about 2×10^2 Hz. Using both stages reduces the expected laser phase-noise photon flux to about 1×10^{-4} Hz, below the signal flux for α of order unity [1]. The corresponding laser-noise displacement budget appears in Fig. 9.1. The unattenuated laser-noise trace is far above the geontropic target band, while the LFC and PRC attenuated traces push the laser contribution below the geontropic curve near the operating frequency $\nu + \epsilon_r$.

The vertical axis in Fig. 9.1 is an equivalent displacement ASD. A laser phase-noise ASD can be converted to an optical path displacement ASD using

$$\sqrt{S_x} = \frac{\lambda}{2\pi}\sqrt{S_\phi}. \tag{9.2}$$

For a 1550 nm laser with $\sqrt{S_\phi} = 10^{-7}$ rad/$\sqrt{\text{Hz}}$, this gives

$$\sqrt{S_{x,\text{raw}}} = \frac{1550\text{ nm}}{2\pi}\left(10^{-7}\frac{\text{rad}}{\sqrt{\text{Hz}}}\right) \simeq 2.5\times10^{-14}\frac{\text{m}}{\sqrt{\text{Hz}}}, \tag{9.3}$$

before any interferometric suppression or optical coupling factors are included. The unfiltered laser-noise trace, plotted in Fig. 9.1, is closer to 10^{-16} m/$\sqrt{\text{Hz}}$ because the laser phase noise does not couple directly as a full optical path fluctuation. Instead, it is suppressed by the small amount of carrier light leaking through the interferometer contrast defect. The approximate amplitude coupling factor is

$$\sqrt{\frac{P_{\text{cd}}}{P_{\text{arm}}}} = \sqrt{\frac{0.1\text{ W}}{5000\text{ W}}} \simeq 4.5\times10^{-3}. \tag{9.4}$$

Applying this factor to the raw phase-to-displacement conversion gives

$$2.5\times10^{-14}\times4.5\times10^{-3} \simeq 1.1\times10^{-16}\frac{\text{m}}{\sqrt{\text{Hz}}}, \tag{9.5}$$

which explains why the no-attenuation laser phase-noise trace sits a little above 10^{-16} m/$\sqrt{\text{Hz}}$ rather than near 10^{-14} m/$\sqrt{\text{Hz}}$. The LFC parameters, including the cavity length and mirror transmissions, determine the filtered laser phase-noise traces, but they do not change the base phase-to-displacement conversion in Eq. 9.2 except through whatever optical coupling is explicitly applied. In this estimate, the arm power and contrast-defect carrier power are the main parameters setting the suppression of the unfiltered displacement-equivalent laser noise.

9.1 Design Requirements

The LFC must pass the carrier field while rejecting laser noise at MHz offsets. It acts as a low-pass filter for laser frequency and amplitude noise. The cavity must remain resonant for the carrier, but its linewidth must be narrow enough that noise near $\epsilon_r \simeq 17.6$ MHz is strongly attenuated.

The reference GQuEST design uses a LFC with a round-trip length near 4.45 m and input/output coupler transmissivities with a nominal value of 883 ppm [1]. The mechanical design for the LFC's spacer and end mirrors sets $L_{\text{LFC}} = 4.45$ m, giving an FSR of

$$\text{FSR}_{\text{LFC}} = \frac{c}{L_{\text{LFC}}} \simeq 67.37\text{ MHz}, \tag{9.6}$$

and an optics-derived finesse of order

$$\mathcal{F}_{\text{LFC}} \simeq 3550, \tag{9.7}$$

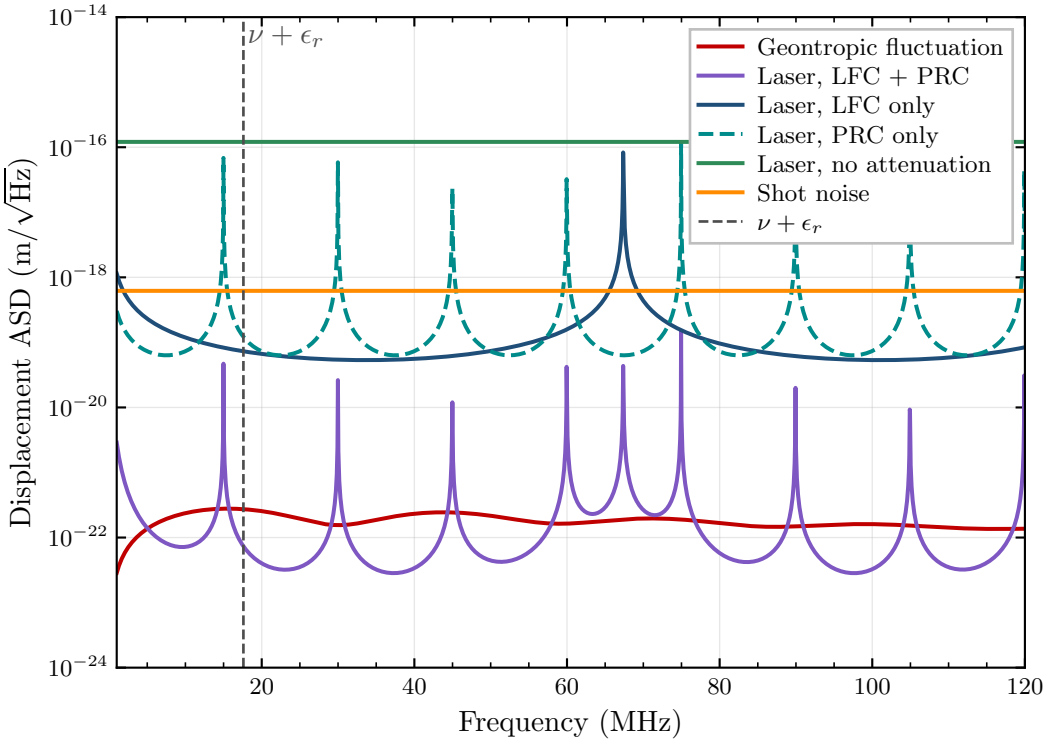


Figure 9.1: Laser-noise displacement budget illustrating the role of the LFC and PRC. The green trace gives the laser phase-noise contribution without cavity attenuation, the blue and dashed cyan traces give the individual suppression from the LFC and PRC, and the purple trace gives the combined attenuation. In the GQuEST design, the corresponding output-filtered laser-noise rates are about 2×10^{8} Hz with no filtering, 2×10^{2} Hz with the LFC only, 2×10^{2} Hz with the PRC only, and 1×10^{-4} Hz with both stages. The dashed vertical line marks the GQuEST operating frequency $\nu+\epsilon_r$, where the combined filtering brings the laser-noise contribution below the geontropic target curve.

from the nominal coating transmissions before internal scatter and absorption are included. The value used in the current GQuEST laser-noise budget in Fig. 9.1 is $\mathcal{F}_{\mathrm{LFC}} \simeq 3550$ [1]. The corresponding design linewidth is

$$\Delta f_{\mathrm{FWHM,LFC}} = \frac{\mathrm{FSR}_{\mathrm{LFC}}}{\mathcal{F}_{\mathrm{LFC}}} \tag{9.8}$$

with a pole at

$$f_{\mathrm{pole,LFC}} = \frac{\Delta f_{\mathrm{FWHM,LFC}}}{2}. \tag{9.9}$$

For $\mathcal{F}_{\mathrm{LFC}} \simeq 3550$, these are approximately

$$\Delta f_{\mathrm{FWHM,LFC}} \simeq 19.0\ \mathrm{kHz} \tag{9.10}$$

and

$$f_{\mathrm{pole,LFC}} \simeq 9.48\ \mathrm{kHz}. \tag{9.11}$$

Quantity	Symbol	Reference value
Laser wavelength	λ	1550 nm
Round-trip optical path length	L_{LFC}	4.45 m
Free spectral range	$\mathrm{FSR}_{\mathrm{LFC}}$	67.37 MHz
Input/output coupler transmissivity (M1/M2)	T_{1550}	883 ppm nominal
Optics-derived finesse	$\mathcal{F}_{\mathrm{LFC}}$	3550
Design FWHM	$\Delta f_{\mathrm{FWHM,LFC}}$	19.0 kHz
Design pole	$f_{\mathrm{pole,LFC}}$	9.48 kHz
LFC-only suppression		$\sim$ 60 dB
PRC-only suppression		$\sim$ 60 dB
Combined suppression		$\sim$ 125 dB
No-filter laser-noise photon flux	$\dot{N}_{\mathrm{pass}}^{\mathrm{LP}}$	2×10^{8} Hz
LFC-only laser-noise photon flux	$\dot{N}_{\mathrm{pass}}^{\mathrm{LP}}$	2×10^{2} Hz
PRC-only laser-noise photon flux	$\dot{N}_{\mathrm{pass}}^{\mathrm{LP}}$	2×10^{2} Hz
LFC+PRC laser-noise photon flux	$\dot{N}_{\mathrm{pass}}^{\mathrm{LP}}$	1×10^{-4} Hz
Demonstrator input power	P_{in}	1 W
Final input power	$P_{\mathrm{in\,final}}$	15 W

Table 9.1: Reference design parameters for the LFC. The measured ringdown finesse is summarized in Table 9.5.

9.2 Filter Response

For frequencies well below the FSR, the LFC acts as a single-pole optical low-pass filter. The power transfer function for laser noise sidebands can be approximated as

$$|H_{\mathrm{LFC}}(f)|^2 = \frac{1}{1 + \left(\frac{f}{f_{\mathrm{pole,LFC}}}\right)^2}. \tag{9.12}$$

At $f = \epsilon_r$, with $f_{\mathrm{pole,LFC}} \sim 9.48$ kHz, the suppression is

$$-10 \log_{10} |H_{\mathrm{LFC}}(\epsilon_r)|^2 \simeq 20 \log_{10} \left(\frac{\epsilon_r}{f_{\mathrm{pole,LFC}}}\right), \tag{9.13}$$

which is order 60 dB. The PRC provides an additional resonant filtering stage, but as seen in Fig. 9.1 it does not have the same pole as the LFC. Although the PRC is designed to enhance the signal sideband, it still suppresses laser-noise sidebands. The effective PRC path length is $L_{\mathrm{PRC}} = 10$ m, corresponding to the 7 m arm path

plus the 3 m distance from the power-recycling mirror (PRM) to the beam splitter. With $\mathcal{F}_{\mathrm{PRC}} \simeq 3000$, the corresponding isolated-cavity pole is

$$f_{\mathrm{pole,PRC}} = \frac{c}{4L_{\mathrm{PRC}}\mathcal{F}_{\mathrm{PRC}}} \simeq 2.50\ \mathrm{kHz}. \tag{9.14}$$

At the GQuEST operating frequency, the full periodic cavity response gives 60 dB of additional power suppression. The filtered phase-noise spectrum incident on the interferometer can be written as

$$S^{\mathrm{LP}}_{\phi,\mathrm{filt}}(f) = |H_{\mathrm{LFC}}(f)|^2 |H_{\mathrm{PRC}}(f)|^2 S^{\mathrm{LP}}_{\phi,\mathrm{in}}(f). \tag{9.15}$$

This spectrum is propagated through the interferometer when estimating the residual laser-noise contribution. In Fig. 9.1, the result is presented as an equivalent displacement ASD before and after the LFC and PRC filtering. The filtered laser-noise sidebands incident on the output-filter chain can be written as a dimensionless fractional-power PSD, $S^{\mathrm{LP}}_{\mathrm{rel}}(\epsilon)$. The corresponding photon-flux PSD is obtained by multiplying by the output optical power and dividing by the photon energy:

$$S^{\mathrm{LP}}_{\dot{N}}(\epsilon) = S^{\mathrm{LP}}_{\mathrm{rel}}(\epsilon)\frac{P_{\mathrm{out}}\lambda}{hc}, \tag{9.16}$$

where P_{out} is the optical power incident on the OFCs for the field carrying the laser-noise sidebands. This conversion applies to a PSD; an ASD spectrum must be squared before using Eq. 9.16.

The laser phase-noise photon flux in the output-filter pass band is

$$\begin{aligned} \dot{N}^{\mathrm{LP}}_{\mathrm{pass}} &= \int_{-\infty}^{\infty} S^{\mathrm{LP}}_{\dot{N}}(\epsilon) P_{\mathrm{out,OFC}}(\epsilon - \epsilon_r) d\epsilon \\ &= \frac{P_{\mathrm{out}}\lambda}{hc} \int_{-\infty}^{\infty} S^{\mathrm{LP}}_{\mathrm{rel}}(\epsilon) P_{\mathrm{out,OFC}}(\epsilon - \epsilon_r) d\epsilon, \end{aligned} \tag{9.17}$$

where $P_{\mathrm{out,OFC}}$ is the power transmission of the OFC chain. The LFC requirement can be stated directly in photon-counting language: the contribution in Eq. 9.17 must be below the signal photon flux and below the dominant classical-noise photon flux so as not to seriously limit GQuEST's sensitivity.

9.3 Optical Layout

The LFC is placed between the laser source and the interferometer and is the first system that the laser sees. This is intentional because the LFC is the main stage for suppressing laser noise before it reaches the interferometer. The LFC must be designed to handle the full input power, which is 1 W for the initial test system

(the "demonstrator") and 15 W for the final system. Losses in the non-cavity LFC system, i.e., input optics before the cavity, do not directly reduce the SNR and can be compensated for by increasing the input laser power. However, losses in the LFC will reduce the suppression of the broadband laser noise and directly contribute to increased laser-noise photon flux at the output. The LFC must have low internal losses to maintain the required noise suppression. The LFC is an actively controlled resonant cavity for the 1550 nm carrier. It tracks the laser by changing the cavity length with a piezoelectric actuator on the M4 cavity mirror, which allows the cavity to remain resonant and compensate for slow drifts in the laser frequency or cavity length.

9.3.1 Input Optics

The input optics for the LFC, in Fig. 9.2, serve to allow correction for laser amplitude fluctuations, to provide a phase modulation path for PDH control, to mode match the input beam to the cavity eigenmode, and to prevent unwanted back reflections back into the laser amplifier. The laser is fiber coupled to the laser amplifier and then is delivered to the LFC through a free-space path. Since the laser is at 1 W it won't be coupled into a fiber after the amplifier, so the free-space path starts immediately after the amplifier output. The input optics are designed to handle the full 15 W input power for the final system, so they should be able to handle the 1 W input power for the demonstrator without issue. The first optical element is a Faraday isolator, which prevents back reflections from the LFC from reaching the laser amplifier. The next stage is a power control stage, which will allow for the adjustment of the input power by using a PBS and $\lambda/2$ waveplate. This allows the input power to be reduced for testing and alignment without changing the laser output power or risking damage to the optics. The power control stage also includes an EOAM to correct for non-DC amplitude fluctuations. This allows the LFC to suppress amplitude-noise of the laser. The phase modulation path uses a non-resonant EOPM to apply phase modulation sidebands at a frequency near 28 MHz, which is used for PDH sensing of the cavity resonance condition. The non-resonant condition allows for a tuneable tank circuit to be attached to the EOPM and for the modulation frequency to be adjusted as needed for the PDH control.

The mode-matching optics are designed to match the input beam to the cavity eigenmode, which is determined by the cavity geometry and mirror curvatures. Proper mode matching is central for maximizing the power coupled into the cavity and for minimizing coupling to higher-order modes, which can reduce the effective fi-

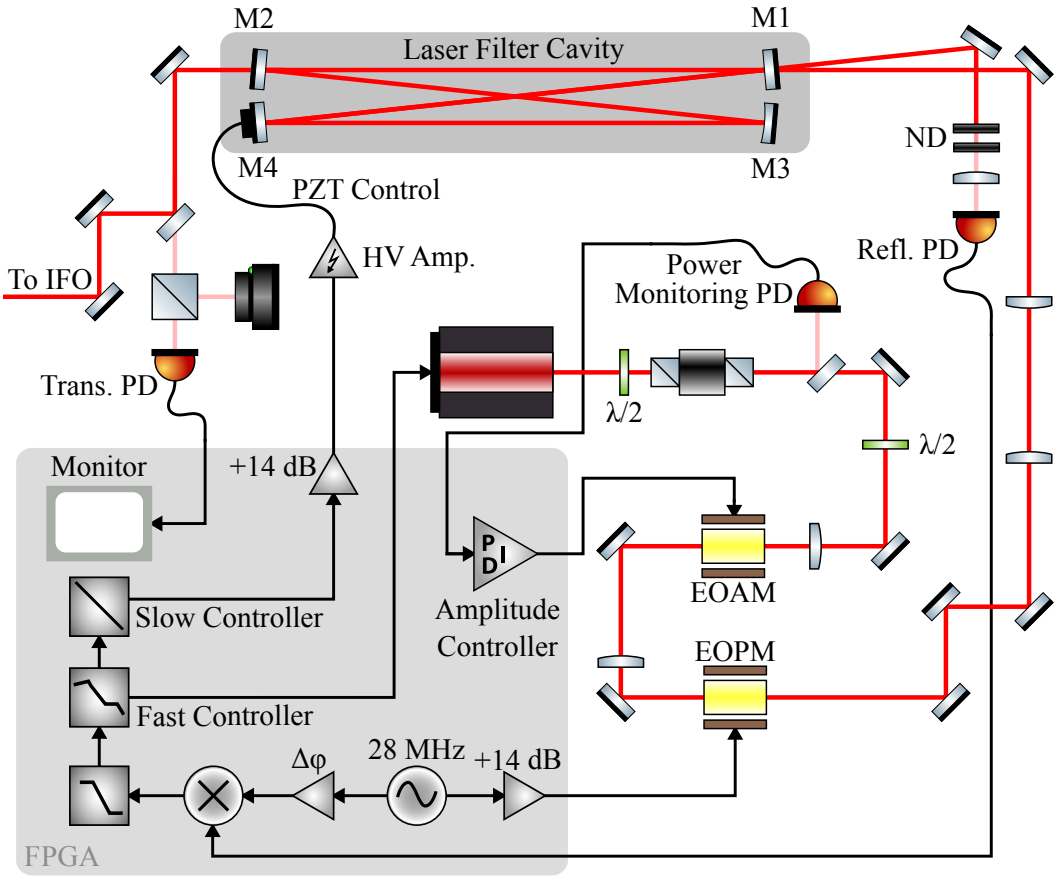


Figure 9.2: Optical layout and control scheme for the LFC. The red path is the 1550 nm laser field. After the main laser passes through the Faraday isolator and power-control optics, the beam passes through the EOAM intensity actuator and the EOPM phase modulator before passing through the mode-matching telescope and finally entering the four-mirror bow-tie cavity. The EOPM is driven at 28 MHz to generate the PDH sidebands. The reflected beam is detected on the reflection PD and demodulated in the FPGA; the demodulation phase delay enhances the error signal. The fast controller feeds back to the laser-frequency actuator, while the slow controller drives the M4 PZT through the high-voltage amplifier to remove long-term cavity-length drift and keep the laser actuator centered. The EOAM is controlled by a separate amplitude PID controller using the input-power monitoring photodiode, so intensity control is kept separate from the PDH frequency lock. The transmitted beam is monitored on the transmission photodiode for alignment, lock acquisition, and out-of-loop power checks before being sent toward the interferometer. The transmission camera is used to check the transmitted mode shape and to help with alignment.

nesse and noise suppression of the cavity. The mode-matching optics are chosen to produce a waist of about 800 μm on the face of M1, with a beam radius near 1.6 mm at M3, as described in Section 9.3.2. The mode-matching telescope was designed using custom mode matching software Mode-matching Optimization THing (MOTH) [151]. The MOTH software allows for the optimization of mode matching in a multi-lens system using a browser-based interface.

9.3.2 Cavity Design

The optical layout of the cavity was planned as a compact four-mirror bow-tie cavity. An approximate 1.5° incidence angle on the cavity mirrors gives a 3° separation between the incident and reflected beams. For the 4.45 m round-trip length, the two straight sections are about 1.112 m long, the two diagonal sections are 1.113 m long, and the transverse separation between the two straight arms is around 5.9 cm. This separation is large enough to route the input, reflected, and transmitted beams without requiring large incidence angles on the cavity mirrors, while also allowing for the cavity to fit into a narrow vacuum tube.

The cavity was designed as a bow-tie rather than a linear cavity for three reasons. First, the reflected control field can be separated from the input beam without relying only on polarization optics. Second, the transmitted beam exits at a convenient location for routing to the interferometer input path. Third, the traveling wave geometry avoids direct back reflection into the LFC's input optics and possibly the laser amplifier.

To design the cavity, the Finesse package was used [145].[1] The Finesse package is a frequency-domain optical modeling software package that can be used to simulate the behavior of complex optical systems, including cavities, interferometers, and mode-matching optics. It allows for the calculation of cavity eigenmodes, beam propagation, and transfer functions, which are essential for designing and optimizing the LFC.

The Finesse model was used to check the Gaussian beam size and transverse-mode spacing. With a curved M3 mirror radius of curvature of -3 m, the modeled waist radius is 805 μm and the Rayleigh range is 1.31 m. The beam reaches its largest radius, about 1.60 mm, on M3 and returns to the waist at M1 after one round trip, as plotted in Fig. 9.3. Using the round-trip Gouy-phase relation in Eq. 3.29, the

[1] The Finesse package is confusingly named. The Finesse software will appear capitalized as "Finesse" in the text or referred to as software or a package, while the cavity finesse is denoted $\mathcal{F}$ and is always lowercase.

accumulated round-trip Gouy phase is 118.9°, as plotted in Fig. 9.4. This value is close to one third of a full cycle, so the third-, sixth-, and ninth-order transverse-mode families repeat near successive longitudinal resonances, as summarized in Fig. 9.5. This spacing should be checked during alignment because near-degenerate higher-order modes can make mode matching and lock acquisition less forgiving.

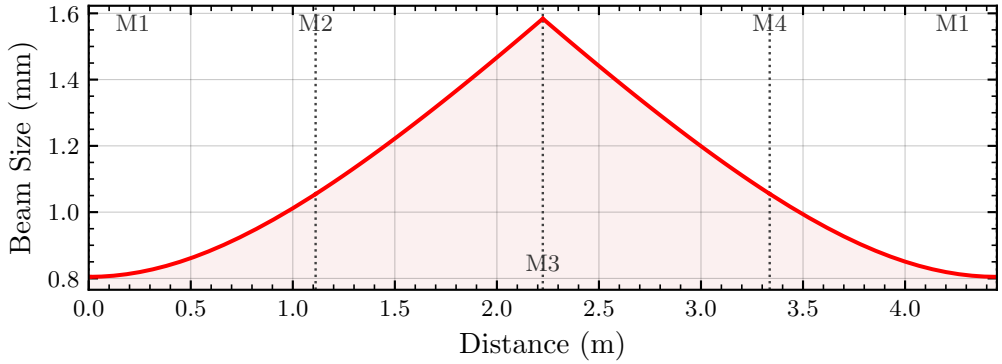


Figure 9.3: Modeled Gaussian beam radius around the unfolded LFC round trip. The cavity eigenmode has a waist of radius 805 μm at M1 and reaches a maximum radius of 1.60 mm at the curved M3 mirror before returning to M1.

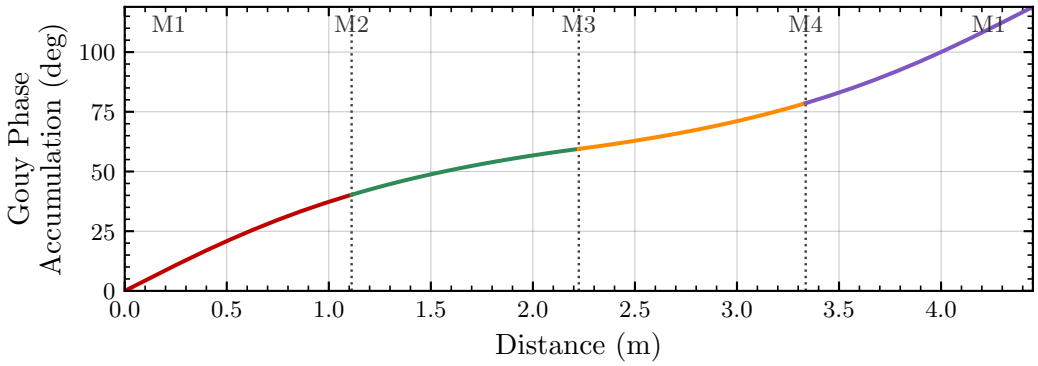


Figure 9.4: Accumulated Gouy phase around the unfolded LFC round trip. For the planned geometry and the −3 m M3 radius of curvature, the round-trip Gouy phase is 118.9°, close to one third of a full transverse-mode cycle.

The target eigenmode also sets the external mode-matching requirement. The input beam should be matched to a waist of 805 μm at M1, with the curved M3 mirror seeing a beam radius near 1.60 mm. The mode-matching telescope does not need to make a very small waist, but it does need to control the waist position over a meter-scale propagation path. Poor mode matching would couple light into the near-degenerate transverse-mode families in Fig. 9.5, reducing transmitted carrier power and adding structure to the reflected PDH signal during lock acquisition.

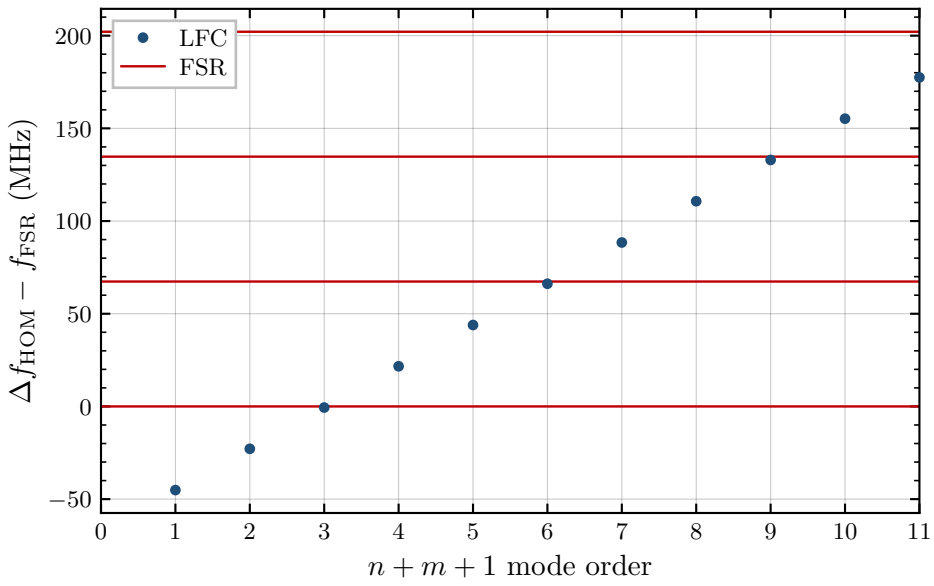


Figure 9.5: Higher-order transverse-mode spacing from the LFC optical model. The markers give the signed offset $\Delta f_{\mathrm{HOM}} - f_{\mathrm{FSR}}$ for mode order $n+m+1$. The horizontal lines mark zero offset and one, two, and three FSRs. Because the round-trip Gouy phase is close to 118°, the third-, sixth-, and ninth-order families fall near successive longitudinal resonances.

The practical alignment sequence follows the single-wavelength limit of the OFC procedure in Section 8.3.5. Since the LFC only uses the 1550 nm resonant path, the 775 nm alignment steps are omitted. In place of the OFC alignment tools, the beam is checked directly against the centers of the cavity mirrors. For initial alignment, the mirrors in the M1 and M2 mounts can be temporarily replaced with 1 in. spacers with holes through their centers. If the beam passes through both holes, the input position and angle are centered on the first two cavity mirrors. The remaining steps use the same logic as the OFC procedure: define the cavity axis with the fixed mirrors, then use the input steering mirrors and remaining adjustable cavity mirrors to walk the transmitted pattern toward the lowest-order mode.

The mirror coating requirements follow from the filter requirement in Section 9.1. M1 and M2 are the input and output couplers, respectively, with nominal reflectivity $R_{1550} = 99.9117\%$, corresponding to $T_{1550} = 883$ ppm when loss is neglected. The remaining mirrors use the same 1550 nm high-reflector coating as the OFC M3/M4 mirrors in Table 8.3, with $R_{1550} > 99.9999\%$, or $T_{1550} < 1$ ppm, also ignoring loss. This yields a nominal round-trip transmission loss of about 1767 ppm.

The measured ringdown finesse gives

$$A_{\mathrm{rt}} \simeq \frac{2\pi}{\mathcal{F}_{\mathrm{LFC}}} = 1.56 \times 10^{-3}, \tag{9.18}$$

where A_{rt} is the approximate total round-trip power loss, including transmission and internal loss. This corresponds to 1561 ppm of total round-trip loss inferred from the ringdown finesse. The inferred value is close to, but lower than, the nominal loss calculated from the coating specifications in Table 8.3. This is acceptable because the table values are nominal coating specifications rather than the final measured transmission of the installed optics. The cavity optics have undergone thorough cleaning using First Contact, a polymer-based surface cleaner that can be peeled off after use to remove contaminants. It is likely that this cleaning process as well as normal variation in optical properties of the mirror coatings could yield a lower inferred round-trip loss [152, 153]. The high-reflector transmission on M3 and M4 remains negligible in either case compared with the input/output couplers. Coating-witness measurements or an independent linewidth scan are needed to separate the exact coupler transmission from scatter, absorption, and other internal losses.

Quantity	Symbol	Value
Round-trip length	L_{LFC}	4.45 m
Incidence angle on mirrors		1.5°
Input/reflected beam separation		3°
Long straight sections		1.112 m
Diagonal sections		1.113 m
Straight-arm separation		5.9 cm
Curved mirror radius of curvature	R_{M3}	−3 m
Waist radius at M1	w_0	805 μm
Rayleigh range	z_R	1.31 m
Maximum beam radius near M3	w_{max}	1.60 mm
Round-trip Gouy phase	ψ_{rt}	118.9°
M1/M2 transmissivity at 1550 nm	T_{1550}	883 ppm nominal
Nominal coating-limited round-trip loss	$A_{\mathrm{rt,opt}}$	1767 ppm
Total round-trip loss from ringdown	A_{rt}	1561 ppm
M3/M4 HR transmission at 1550 nm	$T_{1550,\mathrm{HR}}$	< 1 ppm
Mode-matching efficiency target		$> 90\%$

Table 9.2: Optical design parameters for the LFC. The round-trip length, beam geometry, and Gouy phase come from modeling. The coating values use the same 1550 nm specifications as the OFCs in Table 8.3, while the LFC geometry uses its own curved-mirror radius. The round-trip loss is inferred from the measured ringdown finesse.

Polarization control is also needed because the cavity is operated at nonzero incidence angle which can cause polarization-dependent loss and reduce the effective finesse and noise suppression of the cavity. The input optics include a polarization control stage consisting of a half-wave plate, to ensure that the input beam is properly aligned with the cavity eigenpolarization.

9.3.3 Output Optics

The LFC output optics serve to route the transmitted beam toward the interferometer input path. A beam sampler is used to monitor the transmitted beam, pulling 1% of the beam's power and then splitting that power between the LFC's transmission monitor and a camera sensitive at 1550 nm. The transmitted beam is sent through a mode-matching telescope to the interferometer input path. A possible future addition is a Faraday isolator on the transmitted beam to prevent back reflections from the interferometer from reaching the LFC and an AOM for fast control of amplitude of the transmitted beam if needed to stabilize the interferometer input power. The transmitted beam is also sent to the OFC for filtering of the laser noise sidebands, as described in Chapter 8.

9.4 Vacuum System

The LFC is operated in vacuum, which reduces in-air acoustic coupling, air-index fluctuations, in-air loss, and contamination of the cavity optics. Vacuum operation also improves long-term stability because air currents and pressure changes no longer move the cavity resonance through refractive-index changes.

A computer-aided design (CAD) rendering of the main parts of the vacuum system appears in Fig. 9.6. The rendering depicts two end chambers connected by the long cavity tube, with a central pumping station attached between them. The end chambers provide optical access to the mirror assemblies and allow the input, reflected, and transmitted beams to enter or leave the vacuum system, while also serving as a location to connect the bases of the LFC spacer, which is discussed in Sections 9.6 and 9.7. The central pumping station provides reasonable conductance to the full cavity volume without forcing the pumps to sit directly over one of the mirror assemblies.

The vacuum system uses a staged pumping scheme. A roughing pump first brings the chamber pressure low enough for the turbomolecular pump to operate safely. The turbo pump then pumps the chamber down to the high-vacuum regime. Once the chamber reaches sufficiently low pressure, an ion pump maintains the vacuum

during sensitive operation. This lets the roughing pump and turbo pump be valved off and turned off while the LFC is running at the design sensitivity, eliminating mechanical pump vibrations that could otherwise couple into the cavity length.

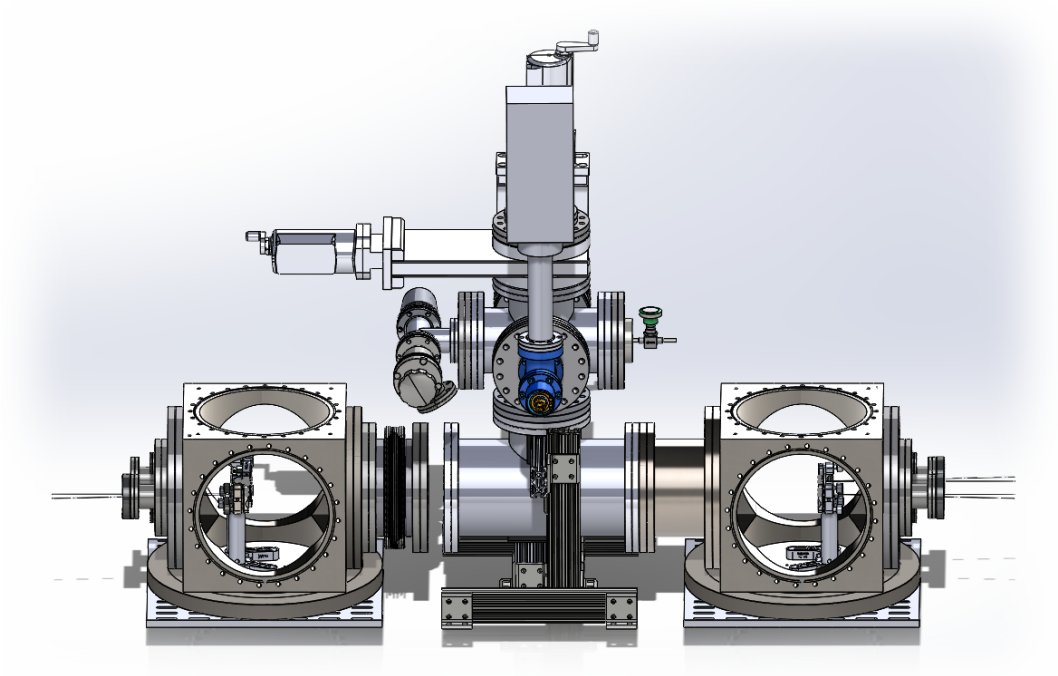


Figure 9.6: Vacuum system for the LFC. The two end chambers house the mirror stations, while the long central tube encloses the cavity spacer and beam path. The central pumping station allows the roughing and turbo pumps to bring the system to high vacuum, after which an ion pump can maintain the pressure during quiet operation. This figure depicts the LFC without the cavity spacer.

The target vacuum level is set by the GQuEST requirement that air-index and acoustic effects stay well below the laser-noise and thermal-noise budgets. The design pressure of the GQuEST interferometer is 10 μPa, equivalent to 7.5×10^{-8} Torr. This value is a natural target for the LFC because the transmitted field is sent directly to the interferometer input. The LFC can be operated at atmospheric pressure, including locking and alignment, but the noise suppression will be reduced by air-index fluctuations and acoustic coupling.

The pump-down sequence is staged so that mechanical pumps are not needed during quiet operation. A typical sequence is to rough the chamber to the crossover pressure for the turbo, start the turbo pump, and allow the pressure to fall into the high-vacuum regime. Once the pressure is in the high-vacuum regime, let the vacuum system stabilize and pump out any water vapor. Next the ion pump is started. The ion pump will cause a slight increase in pressure when turned on because of the outgassing

from the ion pump itself, but this should be small compared with the overall pressure and should stabilize after a short time. Once the ion pump is running and the pressure has returned to the previous level, the turbo pump can be valved off and turned off for quiet operation. The roughing-pump crossover pressure is expected to be 10^{-1}-10^{-2} Torr. The turbo-pump handoff to ion-pump operation is expected to occur near 8×10^{-8} Torr.

The viewports are placed so that alignment can be performed without opening the chamber. At minimum, optical access is required for the input beam, reflected PDH beam, and the transmitted beam. The viewport anti-reflective (AR) coatings are designed for 1550 nm, and any large incidence angle is avoided to avoid polarization-dependent loss.

In-vacuum cable routing is limited to the PZT and goes directly from the M4 mirror assembly to the vacuum feedthroughs. The feedthroughs are chosen to be compatible with the vacuum level and to have low outgassing rates. The feedthroughs have free hanging wires on the vacuum side, so as not to couple vibrations from the feedthroughs into the cavity spacer or mirror assemblies.

9.5 Cleaning and Vacuum Bakeout

The components that went into the vacuum chamber were all cleaned and baked to prevent contamination of the chamber and reduce the outgassing rate to achieve the target vacuum level. The components followed different cleaning procedures depending on their material.

The vacuum chamber itself was cleaned using a solvent-based degreaser to remove any oils or contaminants. It was not baked because it has been previously used in vacuum and has a low outgassing rate.

The cavity spacer was contaminated by the oil manufacturing process and was cleaned using four ultrasonic baths. The first two were done with a 2% bath of Liquinox detergent in deionized (DI) water, which is a common cleaning solution for vacuum components. The third and fourth baths were done with a 2% bath of Citranox acidic cleaner in DI water, which is effective at removing metal oxides and other residues. After the ultrasonic baths, the spacer was rinsed with deionized water and dried with filtered nitrogen gas. The spacer was then baked in an in-air oven at 120°C for 48 hours to further reduce the outgassing rate caused by excess water.

The optics mounts from Thorlabs polaris line were only baked out in the vacuum

oven at 120°C for 48 hours.

The optics themselves were cleaned using First Contact, a polymer-based surface cleaner. The optics were not baked because the coatings may be sensitive to high temperatures, and the outgassing from the coatings is expected to be low compared with the spacer and mounts.

For the Viton damping elements (the Viton balls), the cleaning procedure involved walkerization, which is a process of cleaning and conditioning the Viton material to reduce outgassing and improve its performance in vacuum [154]. The procedure involved scrubbing the Viton balls with a nylon brush and DI water before cleaning them in a Liquinox solution before rinsing them with DI water. Next, the Viton balls were pressure cooked at 15 psi for two hours; then the water was changed and the pressure-cooking process was repeated four times. Finally, the Viton balls were dried in an air bake oven for 5 hours at 182°C before being vacuum baked at 182°C for 48 hours. This process helps to reduce the outgassing rate of the Viton and other fluoroelastomers, especially those from molded parts, which can have high outgassing rates due to ingestion of mold release agents and other contaminants during the manufacturing process.

9.6 Spacer and Mechanical Design

High stability optical reference cavities often use monolithic spacers, where the mirrors are directly bonded to a single piece of low-expansion materials like ULE glass/Zerodur. This design minimizes relative motion between the mirrors and can achieve very high passive stability [155–157]. The same approach was applied to the LFC, though for slightly different reasons. The LFC is not designed to be a reference cavity, so the requirements on the mirror stability are less stringent than for optical cavities used in applications such as atomic clocks. The fact that the LFC has a PZT actuator on one of the mirrors also allows for active compensation of slow drifts in the cavity length, but also allows for another potential source of instability if the mirror is not rigidly attached to the spacer.

Even though the LFC is not a reference cavity, it still needs to maintain a stable length even if that length is not perfectly fixed. A monolithic spacer design can help achieve this by reducing the number of mechanical interfaces and potential sources of relative motion between the mirrors as well as limiting the coupling of vibrations from the table and vacuum chamber into the cavity length. The monolithic design also provides a rigid reference for the mirror mounts and helps set the cavity

geometry, which is needed for maintaining the desired beam path and mode structure within the cavity. The rigid structure also allows the spacer to act as a floating optical table for the cavity, which can be supported on vibration isolation elements to further reduce coupling of external vibrations into the cavity length, while also maintaining the relative mirror locations. The monolithic design also simplifies the assembly and alignment of the cavity, as the mirrors are all colinear and parallel.

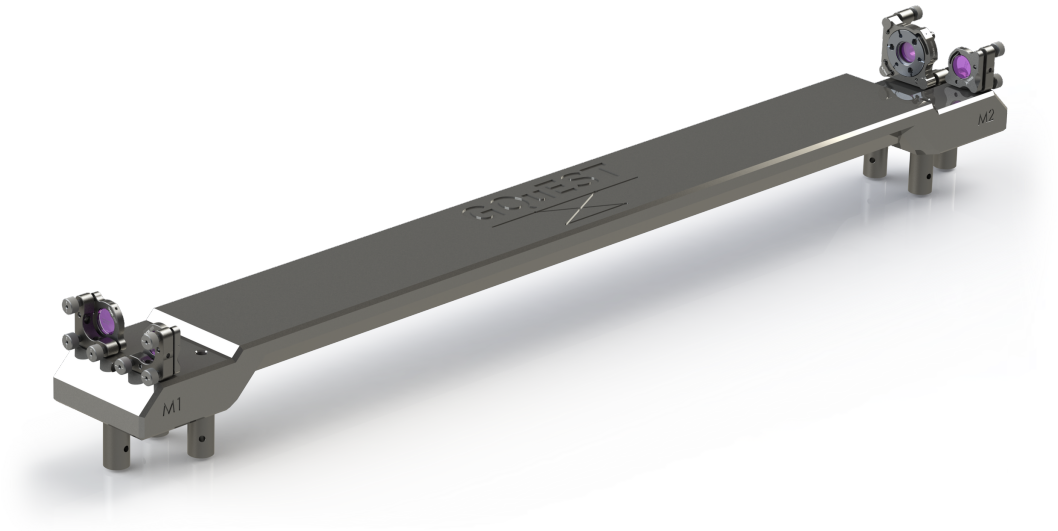


Figure 9.7: CAD rendering of the LFC spacer assembly. The spacer body defines the cavity geometry and holds the mirrors at a fixed distance. The support posts carry the spacer inside the vacuum chamber, and the damping elements under the posts reduce table vibration coupled into the cavity length.

9.6.1 Spacer Material and Geometry

The LFC spacer in Fig. 9.7 is machined from 6082 aluminum.[2] This alloy is a structural aluminum alloy with high stiffness-to-weight ratio, good machinability, and good thermal conductivity [158, 159]. These properties make it convenient for a large spacer that must be machined accurately and then installed inside a vacuum system. The main drawback is its large coefficient of thermal expansion compared with low-expansion optical-spacer materials such as fused silica or ULE glass/Zerodur. For the LFC, this is acceptable because the cavity is actively controlled, but it means that temperature drift must be handled by the PDH loop and the slow actuator range. The actuator's range is limited by the PZT stroke length.

The spacer was fabricated with a bead-blasted finish. This helped during alignment because it makes reflections off the spacer more obvious. It is less ideal for vacuum

[2] Also called EN AW-6082, DIN 3.2315, AA 64430, and AlSi1MgMn.

than a smooth machined or polished finish because textured surfaces are harder to clean, have much more surface area, and can retain more water or particles [160, 161]. Fine bead blasting is common in vacuum hardware, but high-vacuum optical assemblies are usually treated more conservatively. For this reason, the spacer had to be cleaned carefully, as discussed in Section 9.5.

Quantity	Symbol	Value
Alloy		6082 aluminum
Density	ρ	2.70 g/cm^3
Young's modulus	Y	70 GPa
Shear modulus	G	26 GPa
Poisson ratio	ν_{P}	0.33
Thermal expansion	α_{Al}	24×10^{-6} K^{-1}
Thermal conductivity	κ	180 W/m K
Specific heat capacity	c_p	897 J/kg K
Thermal diffusivity	D_{th}	7.4×10^{-5} m^2/s
Proof stress	$\sigma_{0.2}$	255 MPa
Tensile strength	σ_{UTS}	300 MPa
Hardness		91 HB
Spacer volume	V_{sp}	4.546×10^{-3} m^3
Spacer surface area	A_{sp}	0.3718 m^2
Surface-area-to-volume ratio	$A_{\mathrm{sp}}/V_{\mathrm{sp}}$	81.8 m^{-1}
Spacer mass	m_{sp}	12.3 kg (27.12 lbs)
Spacer heat capacity	C_{sp}	11.0 kJ/K
Surface finish		bead blasted

Table 9.3: Material and geometric properties of the LFC spacer. The mass and heat capacity are calculated from the measured CAD volume and representative room-temperature 6082 aluminum material properties.

The spacer is designed around the optical round-trip length rather than around a single linear mirror separation. The round-trip optical path fixes the FSR through Eq. 9.6. Since the cavity is a bowtie, changing the total round-trip path length by one wavelength moves the resonance by one FSR. For the carrier, one FSR of length tuning corresponds to $\Delta L_{\mathrm{FSR}} = \lambda = 1550$ nm. This is the minimum length scale that must be covered during lock acquisition, if the laser frequency is fixed, to guarantee that a resonance can be found.

The length-to-frequency conversion follows from the cavity resonance condition:

$$\frac{\delta \nu}{\nu} = -\frac{\delta L}{L_{\mathrm{LFC}}}, \tag{9.19}$$

and

$$\frac{\nu}{L_{\mathrm{LFC}}} = 4.35 \times 10^{13}\ \mathrm{Hz/m}. \tag{9.20}$$

The same coefficient converts round-trip length noise into frequency noise. For the linewidth inferred from the ringdown finesse in Table 9.5, 1 pm of round-trip length motion shifts the cavity resonance by 43.5 Hz, and one FWHM corresponds to 385 pm. These values set the mechanical stability scale for the spacer, suspension, and slow length-control loop.

The mirror mounts are referenced to the spacer rather than to independent optical posts. This keeps the cavity geometry tied to one mechanical object, which reduces relative motion between mirrors. The end stations hold the mirror assemblies for the two compact ends of the bow-tie. The long central body provides the separation needed for the steady round-trip path.

9.6.2 Thermal Expansion and Drift

The thermal expansion of the spacer gives the largest passive length drift. Aluminum has a significant thermal expansion coefficient and is not a replacement for ULE glass. The LFC must be actively compensated for length changes. For a uniform temperature change ΔT, the round-trip length change is

$$\Delta L_{\mathrm{th}} = \alpha_{\mathrm{Al}} L_{\mathrm{LFC}} \Delta T. \tag{9.21}$$

The corresponding resonance-frequency shift is

$$\Delta \nu_{\mathrm{th}} = -\nu \alpha_{\mathrm{Al}} \Delta T. \tag{9.22}$$

For 6082 aluminum and the LFC's round-trip length,

$$\frac{dL_{\mathrm{LFC}}}{dT} = 1.07 \times 10^{-4}\ \mathrm{m/K} \tag{9.23}$$

and

$$\frac{d\nu}{dT} = -4.64\ \mathrm{GHz/K}. \tag{9.24}$$

A lab temperature stability of 0.5°C is reasonable. If the spacer follows that temperature change uniformly, which isn't a given because it is insulated by a vacuum and sitting on Viton balls which have reasonably low thermal conductivity, the round-trip length changes by $\Delta L_{0.5^\circ\mathrm{C}} = 53.4\ \mu\mathrm{m}$ and the cavity resonance shifts by $\Delta \nu_{0.5^\circ\mathrm{C}} = -2.32$ GHz. This is equivalent to about 34.5 FSRs, or about 1.39×10^5

cavity linewidths inferred from the ringdown finesse. The passive aluminum spacer cannot be treated as a frequency reference on laboratory temperature timescales. Its role is to keep the cavity geometry rigid and mechanically quiet while the control system tracks slow drift. The fact that a slight change in temperature can move the cavity length that many FSRs means that the PZT slow actuator must have a range of as many FSRs to stay continuously locked during temperature changes. If the PZT range is limited, then the control system must be able to handle lock losses and reacquisition on a different resonance.

The thermal response of the spacer is set by its heat capacity, its surface area, and the thermal conductance to the chamber and supports. The spacer heat capacity is about 11.0 kJ/K, as listed in Table 9.3, so a heat input of 1 W would change the average spacer temperature at a rate of

$$\frac{dT}{dt} = \frac{P}{C_{\mathrm{sp}}} = 9.1 \times 10^{-5}\ \mathrm{K/s} \tag{9.25}$$

before any heat is removed by the supports or radiation. In vacuum, radiative exchange with the chamber can also matter. Using the spacer surface area[3] and a room-temperature linearized radiation conductance,

$$G_{\mathrm{rad}} \simeq 4\epsilon k_{\mathrm{B}} T^3 A_{\mathrm{sp}}, \tag{9.26}$$

where ϵ is the effective emissivity of the spacer and k_{B} is Boltzmann's constant. With an estimated $\epsilon = 0.2$, the radiative conductance is about

$$G_{\mathrm{rad}} \simeq 0.42\ \mathrm{W/K}, \tag{9.27}$$

which gives a radiative thermal time constant, or the time for the spacer to change temperature by $1/e$ of its initial value after a step change in heat input, of about

$$\tau_{\mathrm{rad}} = \frac{C_{\mathrm{sp}}}{G_{\mathrm{rad}}} \simeq 7.2\ \mathrm{hr}. \tag{9.28}$$

The real thermal time constant will also include conduction through the support posts, Viton contacts, and any in-vacuum wiring as well as the emissivity of the mirrors and their mounts. The main point is that room-temperature changes on hour timescales can move the passive aluminum cavity by many linewidths, so slow feedback range is required even if the lab temperature is held well within 0.5°C.

[3] The spacer surface geometry sets the radiative heat transfer. However, since none of these calculations will change the conclusions, this analysis makes some order-of-magnitude estimates.

9.6.3 Spacer Mode Shapes

The spacer stiffness was modeled with a COMSOL Multiphysics® eigenmode model [162]. The five lowest mode-shape views exported from that model are displayed in Fig. 9.8. The frequencies in the panel labels are the eigenfrequencies of the modes. The color scale and exaggerated deformation indicate the relative shape of each mode, not the measured displacement amplitude.

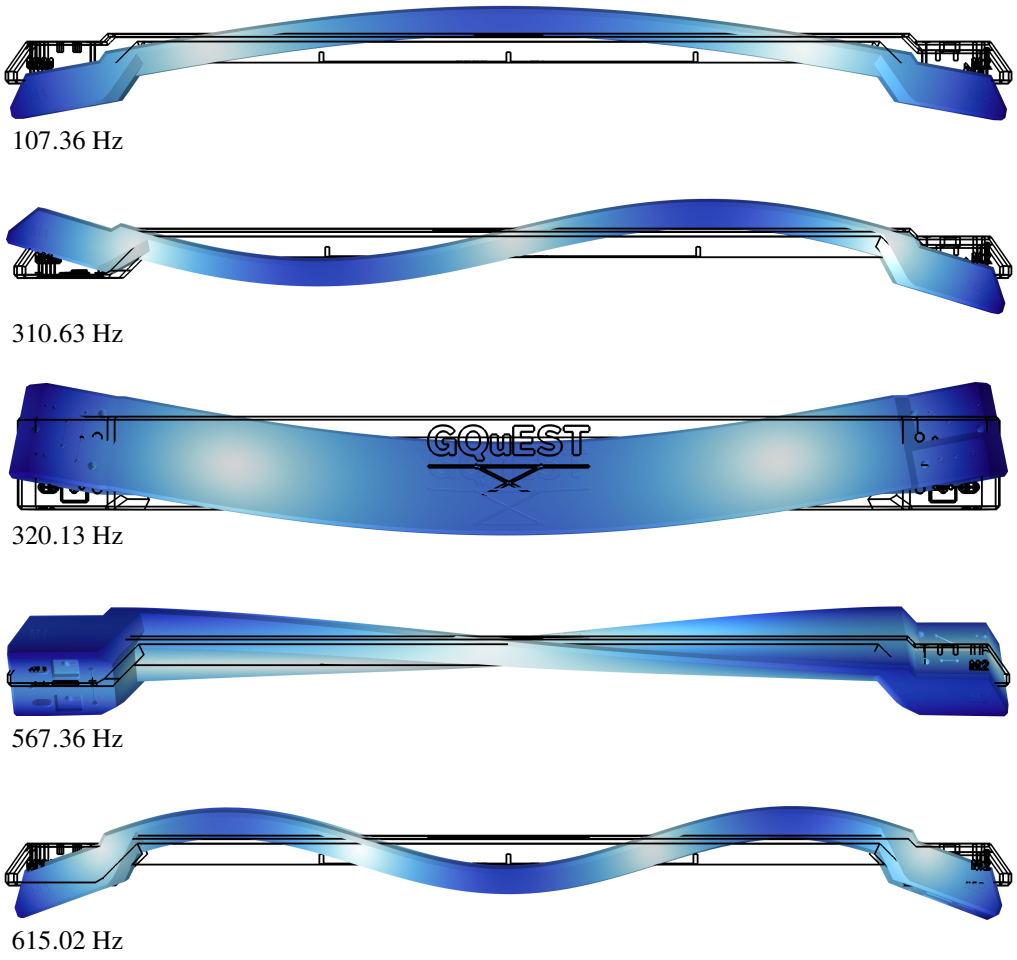


Figure 9.8: Low frequency mode shapes for the LFC spacer. The overlaid labels give the eigenfrequency for each mode. The deformation is plotted with an exaggerated scale so that the spatial character of each mode can be seen. The top two and bottom two modes are viewed from the side where the deformation is most pronounced and the middle mode is viewed from the top where its deformation is most visible.

The lowest spacer modes are bending modes with both vertical and horizontal deformation visible across the mode set. The spacer modes fall into the acoustic range and are higher than the bounce modes of the spacer on top of its Viton isolation

elements. The lowest spacer mode is at 107 Hz, which is discussed in Section 9.7. The eigenfrequencies are in the hundred-hertz range, well above the slow thermal drift that sets the large dynamic range of the PZT, but still low enough to matter for vibration coupling through the spacer supports.

9.7 Vibration-Isolation and Damping

The LFC spacer is supported on posts, and Viton balls are placed under the posts for passive damping and seismic isolation, as seen in Fig. 9.9. The Viton contacts act as lossy mechanical elements. They reduce the quality factor of support resonances and lower the amount of table vibration transmitted to the cavity spacer. This design is simple, vacuum compatible after proper cleaning, and easy to adjust.

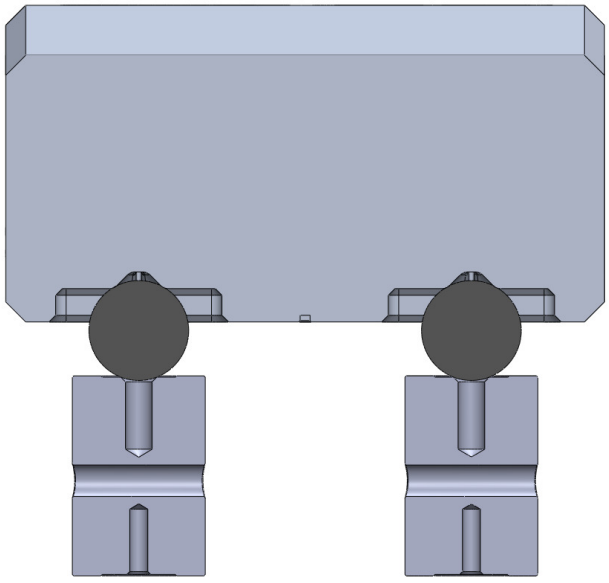


Figure 9.9: Cross-section of the LFC support-post interface. Each support post carries the spacer through a Viton ball contact, so the spacer is mechanically supported while the lossy elastomer contact damps motion transmitted from the table and vacuum hardware.

The support system can be modeled as a base-excited mass on a lossy spring. This is the standard single-degree-of-freedom vibration-isolation model [163]. Let $x_{\mathrm{tab}}(t)$ be the table displacement and $x_{\mathrm{sp}}(t)$ be the spacer displacement. The spring and damper respond to the relative displacement and velocity between the spacer and the table, so the equation of motion is

$$m_{\mathrm{sp}}\ddot{x}_{\mathrm{sp}} + c_{\mathrm{supp}}\left(\dot{x}_{\mathrm{sp}} - \dot{x}_{\mathrm{tab}}\right) + k_{\mathrm{supp}}\left(x_{\mathrm{sp}} - x_{\mathrm{tab}}\right) = 0, \tag{9.29}$$

where k_{supp} is the effective support stiffness and c_{supp} is the effective damping coefficient of the support.

For harmonic table motion,

$$x_{\text{tab}}(t) = X_{\text{tab}} e^{i\omega t}, \tag{9.30}$$

$$x_{\text{sp}}(t) = X_{\text{sp}} e^{i\omega t}, \tag{9.31}$$

so Eq. 9.29 becomes

$$\left(k_{\text{supp}} - m_{\text{sp}}\omega^2 + ic_{\text{supp}}\omega\right) X_{\text{sp}} = \left(k_{\text{supp}} + ic_{\text{supp}}\omega\right) X_{\text{tab}}. \tag{9.32}$$

The displacement transmissibility from table motion to spacer motion is thus

$$T_{\text{vib}}(\omega) \equiv \frac{X_{\text{sp}}}{X_{\text{tab}}} = \frac{k_{\text{supp}} + ic_{\text{supp}}\omega}{k_{\text{supp}} - m_{\text{sp}}\omega^2 + ic_{\text{supp}}\omega}. \tag{9.33}$$

It is convenient to write this in terms of the support resonance frequency and damping ratio,

$$\omega_0 = \sqrt{\frac{k_{\text{supp}}}{m_{\text{sp}}}}, \qquad \zeta_{\text{sus}} = \frac{c_{\text{supp}}}{2m_{\text{sp}}\omega_0}, \tag{9.34}$$

where ω_0 is the natural frequency of the support resonance and ζ_{sus} is the damping ratio. The magnitude of Eq. 9.33 is

$$|T_{\text{vib}}(f)|^2 = \frac{1 + (2\zeta_{\text{sus}}\omega/\omega_0)^2}{\left(1 - (\omega/\omega_0)^2\right)^2 + (2\zeta_{\text{sus}}\omega/\omega_0)^2}. \tag{9.35}$$

This is the expression used for the approximate table-to-spacer displacement coupling.

The behavior of this transfer function is straightforward. Below the support resonance, $\omega/\omega_0 \ll 1$, the spacer follows the table and $|T_{\text{vib}}| \simeq 1$. Near $\omega = \omega_0$, the response is resonantly enhanced, but the peak is reduced by the damping in the Viton contacts. Above resonance, the transmitted motion is suppressed. For a lightly damped support, the spring term gives an approximate f^{-2} rolloff over the isolation band of interest. At very high frequency in the ideal viscous-damping model, the damper provides a direct velocity-coupled path and the asymptotic rolloff becomes closer to f^{-1} [163]. The real Viton contact is viscoelastic rather than an ideal viscous damper.

The support resonance can be calculated from the spacer mass and the Viton-ball contact stiffness. The spacer mass, given in Table 9.3, is supported by four 3/4 in diameter Viton balls [164], as seen in Fig. 9.9. Each ball carries

$$F_{\text{ball}} = \frac{1}{4} m_{\text{sp}} g = 30.2 \text{ N}. \tag{9.36}$$

The elastic modulus of the Viton is obtained from the 75A Shore hardness using the Gent hardness-modulus relation [165, 166]. For Shore-A hardness of 75, the relation gives an elastic modulus of $E_{\text{Viton}} = 7.08$ MPa. Taking $\nu_{\text{Viton}} = 0.49$, the effective Hertz modulus for contact against a much stiffer metal surface is

$$E^* = \frac{E_{\text{Viton}}}{1 - \nu_{\text{Viton}}^2} = 9.32 \text{ MPa}. \tag{9.37}$$

The support geometry in Fig. 9.9 places each Viton ball between the spacer and a support post, giving two compressed contacts per ball. For one spherical contact, the Hertz relation gives [167, 168]

$$F_{\text{ball}} = \frac{4}{3} E^* \sqrt{R_{\text{ball}}} \delta_c^{3/2}, \tag{9.38}$$

where R_{ball} is the ball radius and δ_c is the indentation of the ball into the support post and spacer. Solving for the static indentation gives $\delta_c = 0.852$ mm. The small-signal stiffness of one contact is

$$k_c = \frac{dF}{d\delta_c} = \frac{3F_{\text{ball}}}{2\delta_c} = 5.31 \times 10^4 \text{ N/m}. \tag{9.39}$$

The two contacts on a single ball act in series, while the four balls act in parallel:

$$k_{\text{ball}} = \frac{k_c}{2} = 2.66 \times 10^4 \text{ N/m}, \qquad k_{\text{supp}} = 4k_{\text{ball}} = 1.06 \times 10^5 \text{ N/m}. \tag{9.40}$$

The vertical support resonance is

$$f_0 = \frac{1}{2\pi} \sqrt{\frac{k_{\text{supp}}}{m_{\text{sp}}}} = 14.8 \text{ Hz}. \tag{9.41}$$

The remaining piece of the support model is the damping. For the Viton balls, the damping comes from viscoelastic loss in the compressed elastomer contacts. This loss is usually described by a material loss factor, $\eta_{\text{Viton}} = \tan\delta$, which depends on temperature, frequency, preload, and strain amplitude [163, 169]. Since the exact loss factor for the installed Viton is not easily determined, except through

Quantity	Symbol	Value
Spacer mass	m_{sp}	12.3 kg
Number of Viton balls	N_{ball}	4
Load per ball	F_{ball}	30.2 N
Ball diameter		3/4 in
Ball radius	R_{ball}	9.525 mm
Durometer		75A
Viton modulus from Gent relation	E_{Viton}	7.08 MPa
Poisson ratio	ν_{Viton}	0.49
Effective contact modulus	E^*	9.32 MPa
Indentation per contact	δ_c	0.852 mm
Stiffness per contact	k_c	5.31×10^4 N/m
Stiffness per ball	k_{ball}	2.66×10^4 N/m
Total vertical stiffness	k_{supp}	1.06×10^5 N/m
Support resonance	f_0	14.8 Hz
Damping ratio	ζ_{sus}	set by η_{Viton}

Table 9.4: Calculated vertical support parameters for the LFC spacer on four 3/4 in 75A Viton balls. The stiffness calculation uses the Gent hardness-modulus relation and a Hertz contact model with two contacts in series per ball and four balls in parallel.

experimental validation, the damping is only treated as the mechanism that broadens the support resonance and reduces the resonant amplification.

The vibration requirement can be translated into cavity-frequency units using Eq. 9.19. A table-to-spacer motion of 1 pm/$\sqrt{\mathrm{Hz}}$ produces 43.5 Hz/$\sqrt{\mathrm{Hz}}$ of cavity frequency noise. A motion of 1 nm moves the cavity by 43.5 kHz, more than two linewidths. Small mechanical motion of the spacer can produce a large frequency shift compared with the cavity linewidth.

The support geometry in Fig. 9.9 leaves the spacer kinematically supported by the posts while inserting a lossy Viton contact in the mechanical path. The calculation above gives a vertical support resonance of 14.8 Hz. Below this frequency, the spacer follows the table motion. Near this frequency, the Viton loss broadens the support resonance and reduces the resonant amplification. Above this frequency, the support begins to isolate the spacer from table motion. This places the main support resonance well below the LFC's cavity pole and far below the MHz signal band. While the Viton isolation system does help improve the robustness of the lock, its main purpose is to improve noise performance at high frequency.

9.7.1 Mirror Mount Vibration Damping

When attempting to lock the cavity using the slow PZT actuator, whose loop is discussed in Section 9.10, the loop bandwidth was limited to about 200 Hz by a resonance of the M4 mirror-mount assembly. The M4 mirror is the one actuated by the slow PZT loop, so its mechanical response and UGF directly limit the loop bandwidth. The open-loop transfer function of the M4 actuation path appears in Fig. 9.10. The magnitude response has a clear resonance at 211.4 Hz, and the phase response has the corresponding rapid phase rotation. This resonance severely limited the gain of the slow loop before the damping was added.

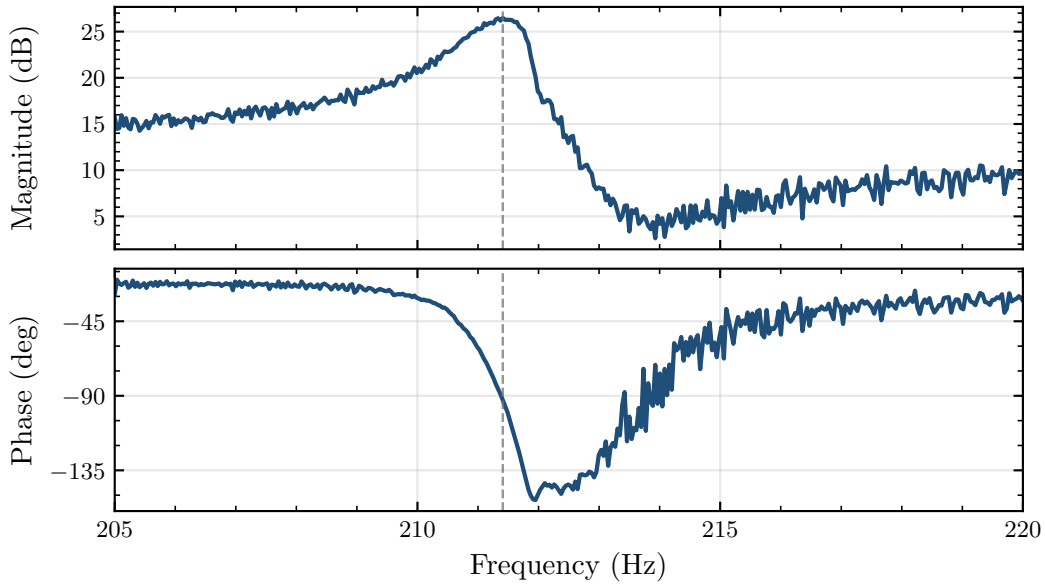


Figure 9.10: Measured open-loop transfer function of the M4 mirror actuation path before adding mirror-mount damping. The magnitude response has a clear resonance near 211 Hz, close to the bandwidth needed for the slow PZT loop, and the phase response has the corresponding rapid phase rotation.

The mirror mount is a Thorlabs POLARIS-K2 mirror mount, which is a side optic retention mirror mount for 2 inch optics. The mount holds a custom 2 inch diameter aluminum assembly which holds the 1 inch mirror and piezo stack. This assembly can be seen in the CAD model in Fig. 9.7. The mount is designed for high stiffness and low mass, but its low contact area and large mass allow for a low-frequency mechanical resonance that can be driven by the PZT feedback.

To damp this resonance, the mirror mount was braced by a piece of Viton and Teflon on a 1 inch diameter post. The Viton on the post was compressed against the face of the mirror mount in front of the top adjustment screw which allows the Viton to

damp the resonance in the cavity axis direction without clipping the incoming and outgoing beams.

The measured transfer function in Fig. 9.10 demonstrates why the M4 mirror mount needed to be treated as part of the control plant rather than as an ideal length actuator. The response peaks near 211 Hz, so increasing the slow-loop bandwidth toward this frequency would drive the mirror-mount resonance and reduce the phase margin of the lock. This resonance set the practical limit on the PZT feedback bandwidth before damping.

The same M4 actuator and mirror-mount dynamics also appeared in the noise measurements. Before the mirror damper was added, the locked displacement spectrum showed a narrow feature around 40 kHz, visible in the noise spectra discussed in Fig. 9.18. After the damper was added, this feature was significantly reduced, which suggests that the mirror mount resonance was affecting the whole cavity noise performance, not just the slow loop bandwidth.

After damping, the measured PZT loop in Section 9.11.1 reached a unity-gain frequency of 1.92 kHz. The damped mount therefore allowed the length actuator to operate well above the original 211 Hz mechanical limit while still keeping the active control bandwidth below the measured cavity pole.

9.8 Finesse Measurements

The LFC finesse was measured from transmitted-power ringdowns. For each ringdown, the cavity was first held on resonance with the laser frequency control loop, discussed in Section 9.10, while the cavity piezo loop remained unlocked. A ringdown was produced by opening the laser frequency control loop, which let the laser move away from resonance without blocking or deflecting the input beam. The transmitted photodiode signal was recorded and fit to extract the cavity power lifetime.

A linewidth scan provides a diagnostic cross-check, but it depends on the calibration and linearity of the laser-frequency sweep. The ringdown measurement is cleaner because the finesse is inferred from a decay time rather than from a calibrated frequency axis.

The finesse from a linewidth measurement is

$$\mathcal{F}_{\mathrm{LFC}} = \frac{\mathrm{FSR}_{\mathrm{LFC}}}{\Delta f_{\mathrm{FWHM,LFC}}}. \tag{9.42}$$

For a ringdown measurement, the relevant convention is whether the fitted exponential describes the optical field or the optical power. If the fitted time constant is the field decay time,

$$\mathcal{F}_{\mathrm{LFC}} = \pi \mathrm{FSR}_{\mathrm{LFC}} \tau_{\mathrm{field}} \tag{9.43}$$

and if the decay is in optical power, then

$$\mathcal{F}_{\mathrm{LFC}} = 2\pi \mathrm{FSR}_{\mathrm{LFC}} \tau_{\mathrm{power}}. \tag{9.44}$$

The transmitted photodiode measures optical power, so the power-decay convention in Eq. 9.44 is used.

Each ringdown was fit with the model

$$V_{\mathrm{tr},i}(t) = V_{0,i} e^{-(t-t_{0,i})/\tau_i} + V_{\mathrm{base},i}, \tag{9.45}$$

where i labels the ringdown, $V_{0,i}$ is the initial transmitted signal, $t_{0,i}$ is the fit start time, τ_i is the fitted power decay time, and $V_{\mathrm{base},i}$ is the fitted baseline. Over a span of 52 seconds, 49 ringdown events were detected. Each candidate ringdown was fit with the model in Eq. 9.45, and the fit quality was assessed using the reduced chi-squared statistic,

$$\chi^2_{\nu,i} = \frac{1}{N_i - p_{\mathrm{fit}}} \sum_{j=1}^{N_i} \left[\frac{V_{\mathrm{tr},i}(t_j) - V_{\mathrm{fit},i}(t_j)}{\sigma_{V,i}} \right]^2, \tag{9.46}$$

where N_i is the number of samples in the fit window, p_{fit} is the number of fit parameters, and $\sigma_{V,i}$ is the voltage uncertainty estimated from the pre-ringdown baseline noise. Ringdowns with a reduced chi-squared statistic of $\chi^2_\nu > 15$ were rejected as outliers, leaving 39 accepted ringdowns, N_{acc}. The rejected ringdowns had reduced chi-squared values well over 15 and were visibly poor fits. The accepted ringdowns had reduced chi-squared values between 0.5 and 12, with a mean of 3.0, which is slightly larger than expected.[4]

The averaged power decay time was computed from the fits,

$$\tau_{\mathrm{power}} = \frac{1}{N_{\mathrm{acc}}} \sum_{i=1}^{N_{\mathrm{acc}}} \tau_i, \tag{9.47}$$

[4]The fact that the mean is above 1 suggests that the voltage uncertainty may be underestimated, but the fit quality is still good enough to give a precise measurement of the power decay time.

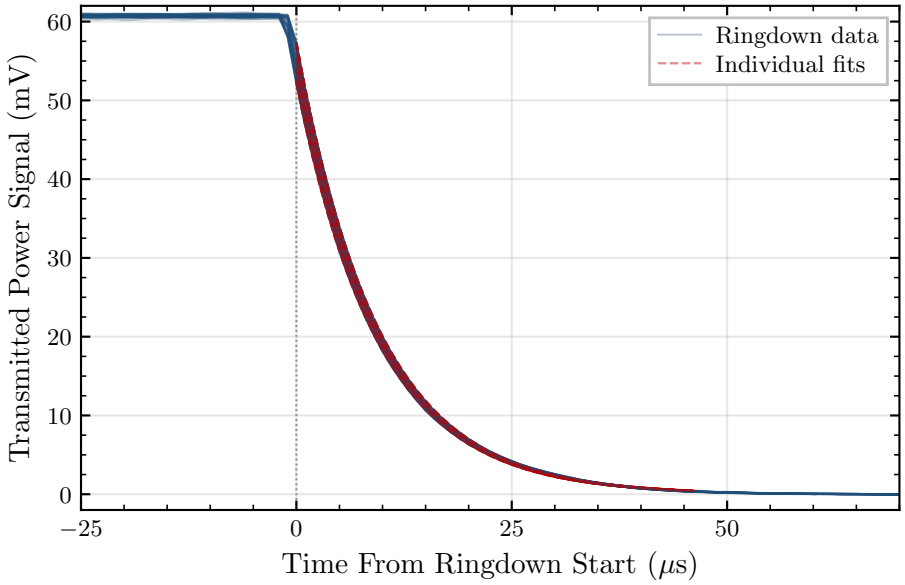


Figure 9.11: LFC transmitted-power ringdown measurements. The blue traces are the ringdowns and the dashed curves are individual single-exponential fits. The measurements give a decay constant of $\tau_{\text{power}} = 9.51 \pm 0.02$ μs and $\mathcal{F}_{\text{LFC}} = 4024.5 \pm 8.6$ when using the design FSR in Eq. 9.6.

with one-standard-deviation distribution width

$$\delta_\tau = \sqrt{\frac{1}{N_{\text{acc}} - 1} \sum_{i=1}^{N_{\text{acc}}} \left(\tau_i - \tau_{\text{power}}\right)^2}. \tag{9.48}$$

The ringdowns give a mean and distribution width of

$$\tau_{\text{power}} = (9.51 \pm 0.02)\ \mu\text{s}. \tag{9.49}$$

Using FSR_{LFC} in Eq. 9.6 and the power-decay convention in Eq. 9.44, the measured finesse is

$$\mathcal{F}_{\text{LFC}} = 4024.5 \pm 8.6. \tag{9.50}$$

From this finesse, the linewidth can be calculated using Eq. 9.42

$$\Delta f_{\text{FWHM,LFC}} = (16.740 \pm 0.036)\ \text{kHz} \tag{9.51}$$

and the cavity pole can be calculated as

$$f_{\text{pole,LFC}} = (8.370 \pm 0.018)\ \text{kHz}. \tag{9.52}$$

The quoted uncertainties are one standard deviation of the ringdown distribution plotted in Fig. 9.12, not the standard error of the mean. They describe the spread

of the individual accepted fits. They do not include systematic uncertainty from the design FSR, the timing calibration, the finite turn-off time of the input field, or the photodetector response, since those effects are much smaller than the spread of the ringdown fits.

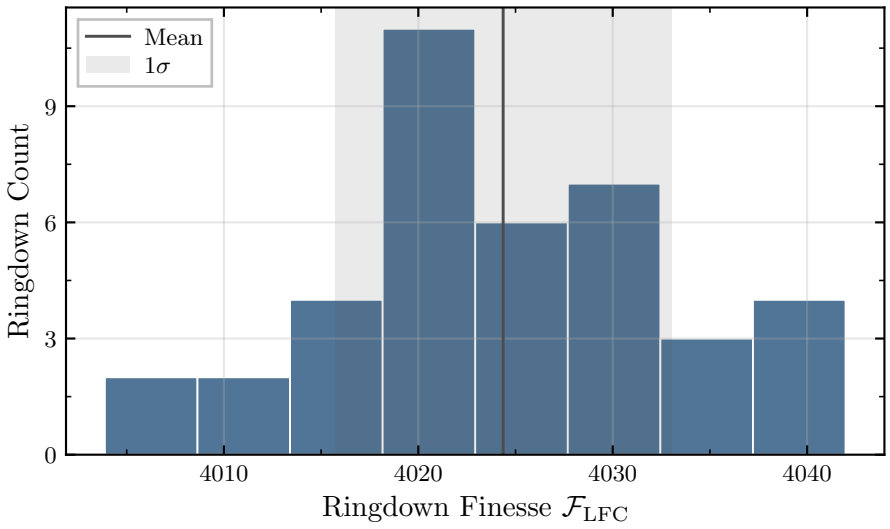


Figure 9.12: Distribution of the accepted LFC ringdown finesse values. The vertical line marks the mean finesse and the shaded band marks the one-standard-deviation width of the distribution.

The measured power lifetime also gives a physical scale for the stored light. It corresponds to

$$N_{\mathrm{rt}} = \frac{\tau_{\mathrm{power}}}{L_{\mathrm{LFC}}/c} = 6.40 \times 10^{2} \tag{9.53}$$

round trips before the stored power decays by a factor of e.

In the high-finesse limit, the total round-trip power loss is approximately

$$A_{\mathrm{rt}} \simeq \frac{2\pi}{\mathcal{F}_{\mathrm{LFC}}} = (1561 \pm 3)\ \mathrm{ppm}. \tag{9.54}$$

This loss includes all power removed from the resonant mode per round trip, including mirror transmission, scatter, absorption, and mode mismatch into nonresonant modes.

The ringdown result is higher than the optics-derived value in Eq. 9.7 because it corresponds to a lower inferred round-trip loss than the nominal mirror-coating estimate. Table 9.2 gives $T_{1550} = 883$ ppm for each of the M1 and M2 couplers and less than 1 ppm for each high reflector, giving a nominal coating-limited round-trip loss of about 1767 ppm. In contrast, Eq. 9.54 gives (1561 ± 3) ppm. The difference

Measurement	Result	Notes
Accepted ringdowns	39	used in average
Ringdown time	9.51 ± 0.02 μs	transmitted-power fit
Finesse from ringdown	4024.5 ± 8.6	power-decay convention
Linewidth	16.740 ± 0.036 kHz	inferred from ringdown
Cavity pole	8.370 ± 0.018 kHz	from ringdown-inferred linewidth
Total round-trip loss	1561 ± 3 ppm	from $2\pi/\mathcal{F}_{\mathrm{LFC}}$
Stored round trips	6.40×10^2	power $1/e$ lifetime

Table 9.5: Measured LFC finesse and quantities calculated from the transmitted-power ringdowns. The measured finesse is higher than the optics-derived value from the nominal mirror coatings.

is about 200 ppm per round trip, or roughly 12% of the nominal loss budget. This is reasonable because the coating values are nominal specifications rather than measurements of the installed optics, and because the installed mirrors were cleaned with First Contact before use, which has been found to improve finesse [152, 153]. This comparison does not by itself identify whether the lower loss comes from slightly lower M1/M2 transmission, lower scatter and absorption, or some combination of those effects. A coating-witness measurement or an independent calibrated linewidth measurement would be needed to separate those contributions. The high-reflector transmission remains negligible in either case compared with the M1/M2 coupler transmission.

9.9 Cavity Scan Measurements

A cavity scan is a diagnostic of the cavity response away from the locked operating point. By sweeping the laser frequency through successive resonances, the full structure of transverse electromagnetic mode (TEM) modes is revealed. The large TEM_{00} peaks give the longitudinal-mode spacing, smaller features reveal residual coupling to higher-order spatial modes, and the peak widths provide an independent check of the cavity linewidth. These scans are used first to calibrate the scan voltage to equivalent round-trip length, then to display the higher-order-mode leakage, and finally to estimate the apparent linewidth of the cavity.

9.9.1 Cavity Scan Calibration

For this measurement, the laser-frequency actuator was driven with a repeated scan and the transmitted and reflected photodiode signals were recorded. The scan was split into linearly ramping triangle-wave half-sweeps so that the voltage axis could

be treated consistently. Each accepted half-sweep contains two adjacent TEM_{00} resonances with weaker higher-order-mode features between them. The endpoint TEM_{00} resonances fall into two voltage families, one with $V_{peak} < 0$, at one end of the sweep, and one with $V_{peak} > 0$, at the other end. This sign convention refers only to the measured scan voltage and is used below when comparing the two apparent linewidth groups. The scans first calibrate the scan voltage to equivalent round-trip length, then display the higher-order-mode leakage, and finally estimate the apparent linewidth of the cavity.

Since the cavity length was held fixed during the scans by not changing the PZT voltage, a change in laser frequency can be written as an equivalent change in cavity round-trip length. Adjacent longitudinal resonances are separated by one FSR, which corresponds to one laser wavelength of round-trip length change for the LFC. The FPGA recorded the scan voltage directly, without extra voltage gain, so the first calibration is simply the voltage separation between the two endpoint TEM_{00} peaks in each accepted half-sweep. Over the 199 half-sweeps discussed in Section 9.9.2, the median separation between the adjacent TEM_{00} peaks is

$$\Delta V_{FSR} = 1.3289\ \text{V}, \tag{9.55}$$

which would give a single linear calibration of 1.166 μm/V.

That single number is not quite enough for these scans. When the data are plotted with this single linear calibration, the two endpoint TEM_{00} families do not behave the same way: the $V_{peak} > 0$ peaks appear broader than the $V_{peak} < 0$ peaks. Also, the higher-order-mode features between the endpoint peaks also appear shifted toward negative voltage instead of landing at the same fractional position in each sweep. This indicates that the laser-frequency actuator is not a perfectly linear proxy for optical detuning over the whole sweep. The observed higher-order-mode structure is used to build a piecewise calibration for each sweep.

The ideal mode structure of the LFC is described in Section 9.3.2, which gives an accumulated round-trip Gouy phase of about 118° per round trip. This should produce three transverse-mode families separated by roughly one third of an FSR. Visually, this is similar to the HOM modeling of the OFCs in Fig. 8.11. For each monotonic half-sweep, the two adjacent TEM_{00} peaks were used as endpoints and assigned to 0 and 1550 nm; the two strongest repeated higher-order-mode features between them were then used as intermediate markers. The assigned marker lengths

were

$$L_i = \left\{0, \frac{\lambda}{3}, \frac{2\lambda}{3}, \lambda\right\}, \tag{9.56}$$

with

$$\left.\frac{dL}{dV}\right|_i = \frac{L_{i+1} - L_i}{V_{i+1} - V_i}. \tag{9.57}$$

This is not a precision measurement of the Gouy phase. It is a calibration that asks how much the voltage axis has to be warped to make the dominant higher-order-mode families line up with the expected one-third-FSR spacing.

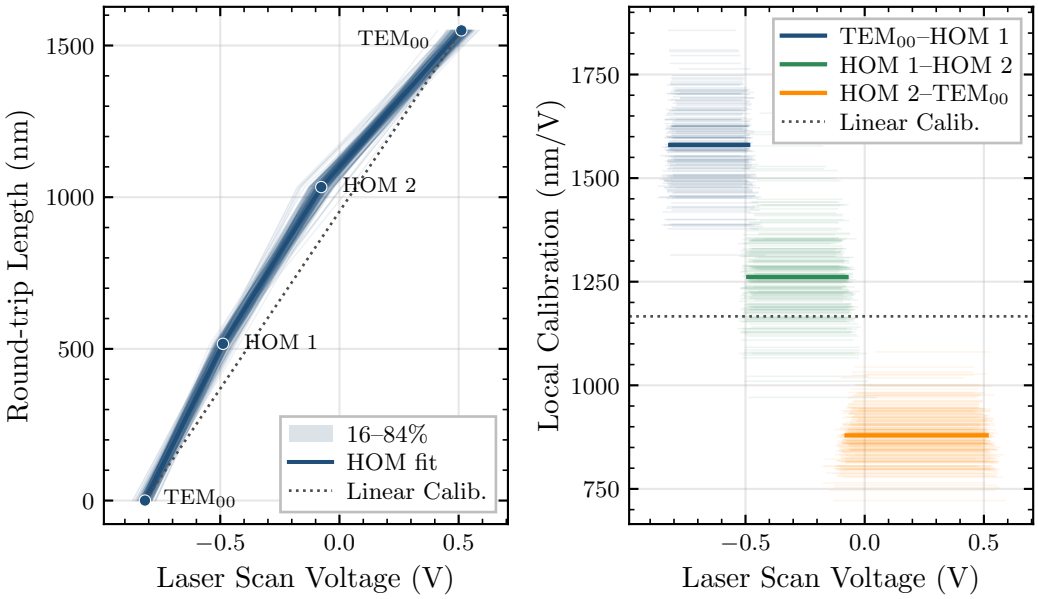


Figure 9.13: Nonlinear voltage calibration for the LFC scans. The left panel gives the piecewise mapping from scan voltage to assigned round-trip length for each accepted half-sweep. The adjacent TEM_{00} peaks set the endpoints at 0 and 1550 nm, while the two dominant higher-order-mode features are assigned to $\lambda/3$ and $2\lambda/3$. The dark curve is the median calibration, the shaded band gives the 16–84% spread, and the dotted line is the single-FSR linear calibration. The right panel gives the local voltage-to-length conversion for each interval in the piecewise calibration. Each faint horizontal segment is the slope for one half-sweep over the voltage range where that slope is applied, the thick colored segments are the median slopes for the three intervals, and the dotted horizontal line is the single-FSR calibration for comparison.

The two higher-order-mode markers appear at median positions of 381 nm and 865 nm when the simple linear FSR calibration is used. After forcing those same features to the expected one-third-FSR spacing, the median local calibrations are

1.580 μm/V from the negative-voltage TEM_{00} peak to the first higher-order-mode marker, 1.261 μm/V between the two higher-order-mode markers, and 0.879 μm/V from the second marker to the positive-voltage TEM_{00} peak. The variation across the sweep confirms that the scan voltage should not be treated as a globally linear frequency axis when scanning over large voltages.

9.9.2 Cavity Scan for Higher-Order Modes

With the nonlinear calibration applied, Fig. 9.14 contains the same half-sweeps after aligning each pair of endpoint TEM_{00} resonances across one FSR. The two large edge features are the $V_{\mathrm{peak}} < 0$ and $V_{\mathrm{peak}} > 0$ TEM_{00} resonances described above, and the smaller features between them are the higher-order modes used as calibration checks. The transmitted signal is plotted on a logarithmic scale so that the weak higher-order-mode transmission can be seen at the same time as the large TEM_{00} peaks. The reflected signal is plotted as an upside-down logarithmic deficit because the same resonances appear as dips in reflection.

The nonlinear calibration makes the dominant higher-order-mode families overlap more cleanly, but it does not remove all of the structure. In particular, the higher-order-mode family near $2\lambda/3$ appears as several nearby peaks rather than a single feature. This is consistent with the bow-tie cavity being very slightly astigmatic because the mirrors are used at nonzero incidence angle. Small alignment errors and mirror-figure errors can also split and mix the ideal transverse-mode families. The peaks in the scan should be interpreted as observed higher-order-mode leakage and scan-linearity diagnostics rather than as a precise measurement of the Gouy phase.

The scans indicate that the LFC has some residual coupling to higher-order modes, which could be improved through updated mode matching and alignment. The largest higher-order-mode peaks are about $1/100^{\mathrm{th}}$ relative to the TEM_{00} peaks, which is reasonably small given the design constraints. The presence of these features confirms that some light is coupled into non-TEM_{00} modes, but they are still small enough that they should not significantly affect the performance of the LFC.

9.9.3 Linewidth Measurements

The calibrated scans can also be used to estimate a cavity linewidth. Since the finesse value of the 1550 nm resonance is reasonably high, fast scans may not allow enough time to mitigate cavity storage time effects. The linewidth was taken from a slower

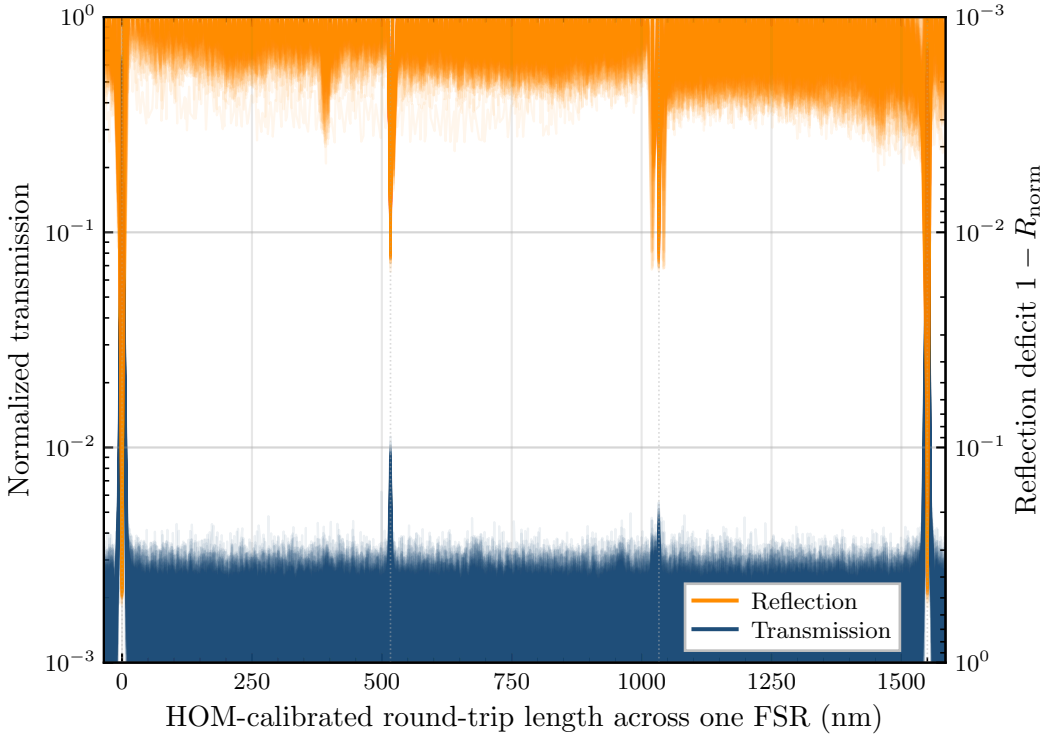


Figure 9.14: Layered LFC scans over one FSR using the nonlinear calibration in Fig. 9.13. The transmission traces are normalized to the TEM_{00} peak and plotted on the left axis. The reflection traces are plotted as the reflection deficit, $1 - R_{\mathrm{norm}}$, on an inverted logarithmic right axis so that reflection dips appear in the same visual direction as transmission peaks. R_{norm} is the normalized reflection coefficient. The two large edge features are the adjacent TEM_{00} resonances, and the smaller repeated features between them are residual higher-order-mode coupling.

scan in which the TEM_{00} peaks were cleanly resolved. Eliminating non-Lorentzian outliers, 70 TEM_{00} peaks were compiled from cavity scans and calibrated using the same nonlinear calibration procedure found in Section 9.9.1. Each peak was centered on its maximum and measured by direct interpolation at half height to give the apparent FWHM. The filtered scan gives a median apparent FWHM of

$$\Delta L_{\mathrm{FWHM,scan}} = 0.499\ \mathrm{nm}. \tag{9.58}$$

Using the design FSR, this corresponds to an apparent frequency linewidth of

$$\Delta f_{\mathrm{FWHM,scan}} = \frac{\Delta L_{\mathrm{FWHM,scan}}}{\lambda}\mathrm{FSR}_{\mathrm{LFC}} = 21.7\ \mathrm{kHz}, \tag{9.59}$$

or an apparent finesse of about 3.11×10^{3}. The centered peaks and linewidth distribution appear in Fig. 9.15.

The scan contained 5.19×10^{6} samples over 10.390 s. The scan-derived linewidth is broader than the 16.740 kHz linewidth inferred from the ringdown measurement

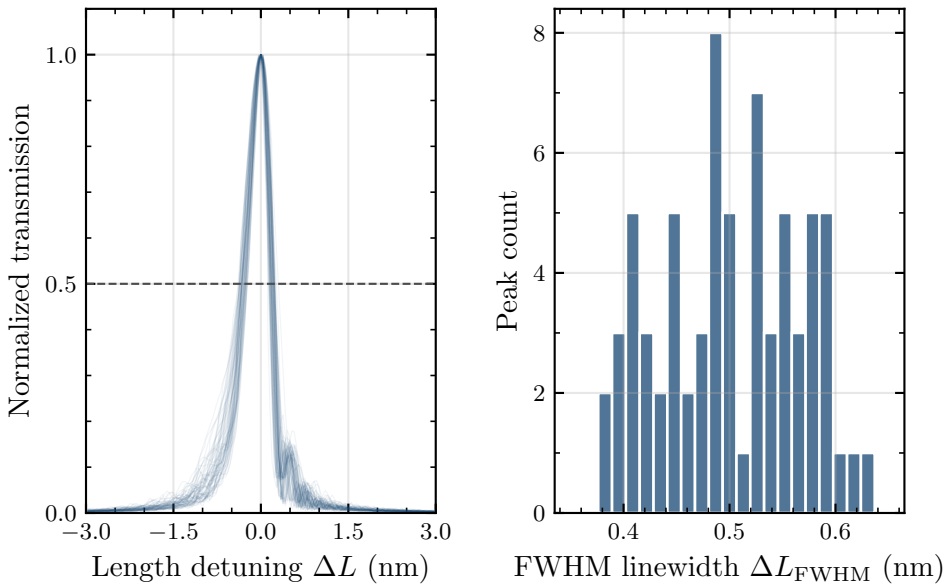


Figure 9.15: Linewidth scan of a single set of LFC transmission peaks. The left panel overlays the detected TEM_{00} peaks after centering them on resonance. The right panel gives the measured FWHM distribution after removing non-Lorentzian outliers.

in Table 9.5, but this factor is not concerning for a high-finesse resonance measured during a frequency sweep. During a scan, the resonance is not held at a fixed detuning, so finite sweep speed, cavity storage time, sampling, and residual voltage-calibration errors can all make the peak appear broader than the intrinsic linewidth. Because the ringdown measurement does not rely on the laser scan voltage, it is the cleaner measurement of the present cavity finesse. The linewidth and cavity scan above still verify the expected FSR structure, give a direct voltage-to-length calibration for lock acquisition, and provide an independent check that the scanned TEM_{00} resonance width is close to the ringdown-inferred value.

9.10 Control Scheme

The purpose of the LFC is to remove those broadband laser-noise sidebands, discussed in Section 3.3.2, before the light reaches the interferometer. The control system does not provide that MHz suppression directly. Instead, it keeps the carrier resonant with the LFC at high frequency and keeps the LFC resonant with the carrier at low frequency so that the passive optical response in Eq. 9.12 filters the noise sidebands at the correct operating point.

This gives the LFC control scheme five goals. First, the LFC must be kept on resonance with the laser carrier, to correct for slow drifts of the cavity length and

not inject low-frequency noise into the transmitted carrier. Second, the carrier must remain locked to the cavity's TEM_{00} resonance so that the transmitted field sent to the interferometer is stable. Third, the feedback bandwidth must remain far below the signal-sideband frequency, ϵ_r, and far below the first FSR; otherwise the loop would become part of the sideband response rather than leaving the sideband rejection to the cavity. Fourth, the system must make acquisition and commissioning practical by providing a scan, a signed detuning signal, transmitted-power monitoring, and enough actuator range to find a resonance. Fifth, frequency control and intensity control must remain separate: the PDH loop controls laser-cavity detuning, while the EOAM provides an independent actuator for laser-intensity noise and input-power stabilization.

The resulting optical and control architecture appears in Fig. 9.2. The diagram combines the 1550 nm optical path, the cavity monitors, and the feedback paths used for control and stabilization. The input optics contain two electro-optic actuators with different roles. The EOAM is used for low-frequency intensity stabilization or calibrated amplitude-noise injection, while the EOPM applies the phase modulation used for PDH sensing. After the mode-matching telescope, the beam enters the four-mirror bow-tie cavity. The reflected beam is detected on the reflection PD and used for the frequency lock, while pick-offs of the transmitted and input beams provide power and alignment monitors. The transmission camera shows the TEM_{00} mode to verify that the cavity is aligned and that the transmitted light is in the correct spatial mode.

The PDH frequency-control loop is implemented with a Moku:Pro manufactured by Liquid Instruments, which is an FPGA with a bandwidth up to 600 MHz.[5] The FPGA provides the local oscillator for PDH, drives the EOPM, digitizes the AC coupled reflected photodiode signal, demodulates it at the modulation frequency, and filters the result to produce an error signal proportional to the carrier detuning from resonance, as discussed in Section 3.2.3. A scan or acquisition ramp can be summed into the feedback path to sweep the laser or PZT through the cavity modes. Once the correct resonance is found, the scan is removed. The primary fast feedback path is sent to the laser to control its frequency. A slow control signal is amplified by an Analog Devices EVAL-ADHV4702-1 high-voltage amplifier board with a gain of 21 V/V before being sent to the PZT pressing on the back of M4. That slow

[5]While the Moku:Pro has a bandwidth up to 600 MHz, it can only operate at that bandwidth with limited computational resources. The LFC does not require nor utilize such high bandwidth.

control signal can be used for acquisition, long-term drift removal, or centering of the laser-frequency actuator.

The transmitted photodiode is not the primary in-loop frequency sensor. During alignment and acquisition, the transmission PD identifies the TEM_{00} resonance peaks and reveals residual higher-order-mode coupling, while the reflection PD supplies the PDH detuning signal needed for feedback. This separation follows the scan measurements in Section 9.9.1: adjacent longitudinal resonances are separated by one FSR, corresponding to one laser wavelength or 1550 nm of round-trip length change. The measured scan calibration helps acquisition, but the ringdown measurement gives the cleaner linewidth and pole used for the control targets.

The intensity-control loop uses a separate PD for sensing intensity. The preferred sensor is an input-power pick-off after the EOAM and before the cavity, so that the intensity loop measures the power delivered toward the LFC without confusing cavity detuning, mode matching, or lock transients with laser-intensity noise. The transmitted photodiode can be used as a slow out-of-loop power monitor, but it should not be the primary sensor for a fast intensity actuator. The EOAM loop is also be checked for residual amplitude modulation, RF feedthrough, and control sidebands that could fall in or near the OFC pass band.

It is possible to improve the intensity stabilization by using the transmitted photodiode as the sensor for a second intensity-control loop that feeds back to a second EOAM that is in the transmitted laser path after the LFC. This would give the option to stabilize the intensity of the light sent to the interferometer, rather than just the light sent to the LFC. However, this would add complexity and is not strictly necessary for the LFC to perform its function of filtering laser noise before the interferometer. This second optional EOAM loop would have to only be closed when the cavity is locked, and would have to be automatically engaged and disengaged when the cavity is unlocked. A possible second intensity stabilization loop would have to be designed carefully to avoid coupling cavity detuning into the intensity noise.

The hierarchy described above preserves the optical filtering requirement. The LFC removes laser-noise sidebands near ϵ_r through its passive cavity response, while the active loops maintain the carrier operating point and stabilize the input power. Using the cavity pole measured from ringdown, of about 8 kHz, the passive low-pass response gives about 60 dB of power suppression at ϵ_r. The fast PDH path suppresses laser-cavity detuning noise by feeding back to the laser frequency

actuator. The slow PZT path removes long-term cavity-length drift and provides the large-range actuator for acquisition. The EOAM path stabilizes laser intensity independently of the frequency lock.

The measured PZT loop had a unity-gain frequency of 1.92 kHz as discussed in Section 9.11.1. This bandwidth gave enough gain for cavity-length drift control while remaining below the measured cavity pole, the signal-sideband frequency, and the first FSR.

The PDH modulation frequency is $f_{\mathrm{PDH}} = 28$ MHz. This placed the modulation sidebands above the measured cavity pole and below the first FSR, so they acted as a phase reference without resonating in the cavity or overlapping neighboring longitudinal modes. The selected RF frequency was checked against the signal-sideband region near ϵ_r, harmonics and mixing products from the OFCs. The demodulation phase was set to maximize the detuning quadrature and maximize the error signal slope.

The reflected photodiode was chosen to have sufficient bandwidth for the PDH modulation frequency and sufficient linear range for the reflected power during acquisition. It is a Thorlabs PDA05CF2 with a 150 MHz bandwidth. Its demodulated output formed the in-loop frequency discriminator. Because the reflected signal sent to the Moku:Pro was AC coupled, acquisition used the transmission PD, scan-calibrated resonance peaks, and any available DC reflection monitor to identify the desired cavity mode before the PDH loop was engaged. During alignment, the transmitted signal measured the coupled TEM_{00} power and higher-order-mode content, while the reflected PDH signal gave the asymmetric detuning near resonance.

After lock acquisition, the in-loop PDH error signal can be calibrated to frequency using the ringdown-inferred linewidth. This converts the Moku:Pro error signal from volts to hertz of laser-cavity detuning. While the transmitted photodiode functioned as a lock monitor, mode-matching diagnostic, and out-of-loop power check, it did not determine frequency stability because it also depended on alignment, mode matching, and input intensity noise.

The EOAM intensity path has not been calibrated as its main job is control rather than measurement. The input-power pick-off before the LFC measured the power delivered toward the cavity without using the cavity resonance as part of the intensity error signal. The EOAM response, intensity-photodiode response, and RF feedthrough from the EOPM and PDH electronics set the usable intensity-loop band-

width. This loop suppressed low-frequency input-power fluctuations and provided a calibrated amplitude-noise actuator. It has not been implemented above 1 Hz because it has not been needed for the LFC commissioning.

A second EOAM after the LFC is a possible upgrade for stabilizing the transmitted power delivered to the interferometer. That loop uses the transmitted photodiode as its sensor and is valid only while the cavity is locked. It could be calibrated to power and act as a power monitor at the output of the LFC; even without the second EOAM.

The final commissioning measurements were the open-loop transfer function of the PZT path, the inferred closed-loop suppression response, and the locked displacement spectra; these are discussed in Section 9.11. The PZT open-loop transfer function gives the length-actuator unity-gain frequency, phase margin, gain margin, inferred actuator response, and remaining resonances in the controlled plant.

The PDH error-signal calibration converted the in-loop error signal into frequency noise, giving the residual carrier detuning and testing whether the lock injected excess low-frequency noise into the transmitted carrier.

9.11 Commissioning Measurements

The commissioning measurements connect the optical characterization of the cavity to its behavior as a locked system. A stable lock requires both a control loop with adequate margin and a mechanical configuration quiet enough that the loop is not constantly correcting large environmental motion. The transfer functions measure the actuator and feedback dynamics, while the spectra measure the residual displacement noise once the cavity is operating in its final configuration and being controlled by the feedback loops.

9.11.1 Transfer Functions

The PZT loop was characterized with a Moku:Pro frequency-response measurement. The measured open-loop transfer function is

$$G_{\mathrm{PZT}}(f) = P_{\mathrm{amp+PZT}}(f) K_{\mathrm{eff}}(f), \tag{9.60}$$

where $P_{\mathrm{amp+PZT}}(f)$ is the combined high-voltage-amplifier and cavity-length plant, and $K_{\mathrm{eff}}(f)$ is the effective controller present during the measurement. This controller is the product of the slow PZT controller, the fast laser-frequency controller, and an additional +24 dB PDH-signal gain applied in the FPGA as a digital gain. For

the measurement in Fig. 9.16, the slow controller had proportional gain −10 dB and an integrator unity-gain frequency of 369.9 Hz. The fast controller had proportional gain −48 dB and an integrator crossover at 36.99 kHz. The controller used to infer the plant was

$$K_{\mathrm{eff}}(f) = 10^{24/20} K_{\mathrm{slow}}(f) K_{\mathrm{fast}}(f), \tag{9.61}$$

$$K_{\mathrm{slow}}(f) = 10^{-10/20} - \mathrm{i}\frac{369.9\ \mathrm{Hz}}{f}, \tag{9.62}$$

$$K_{\mathrm{fast}}(f) = 10^{-48/20}\left(1 - \mathrm{i}\frac{36.99\ \mathrm{kHz}}{f}\right). \tag{9.63}$$

The sign of the feedback loop is kept in the measured open-loop transfer function, so Eq. 9.63 represents the controller shape and digital gain that were divided out of the measured loop. Following the control-theory convention in footnote 3, instability occurs at $G = 1$ rather than $G = -1$. The phase margin is therefore the phase of G at the UGF, measured from 0°.

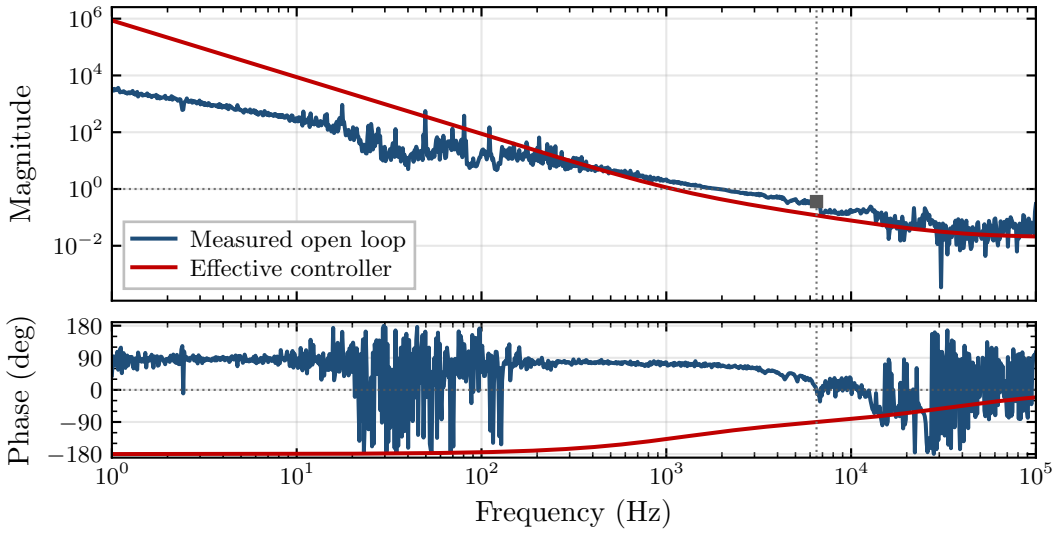


Figure 9.16: Measured open-loop transfer function of the LFC PZT length-control path, with the effective controller from Eq. 9.63 overlaid. The controller includes the slow and fast Moku controllers and the additional +24 dB PDH-signal digital gain. The unity-gain frequency is 1.92 kHz, and the phase margin is 70.2° using the convention that instability occurs at $G = 1$. The gain margin is conservatively taken from the smallest zero-phase crossing above the UGF, giving 8.9 dB at 6.50 kHz.

Dividing the measured open-loop transfer function by $K_{\mathrm{eff}}(f)$, both shown in Fig. 9.16, gives the inferred amplifier-plus-PZT plant. The plotted PZT-only response is found by also dividing out the high-voltage amplifier gain as

$$P_{\mathrm{PZT}}(f) = \frac{G_{\mathrm{PZT}}(f)}{K_{\mathrm{eff}}(f) K_{\mathrm{HV}}}, \tag{9.64}$$

where $K_{\mathrm{HV}} = 21$ V/V for the EVAL-ADHV4702-1 amplifier board. The closed-loop suppression follows the same denominator as Eq. 4.14. For the trace plotted here, the disturbance enters at the sensed cavity-length degree of freedom rather than before the PZT plant, so the plotted response is $1/(1 - G_{\mathrm{PZT}})$ rather than $P_{\mathrm{PZT}}/(1 - G_{\mathrm{PZT}})$. This makes the low-frequency response small when the loop gain is large, because the curve represents residual length motion after feedback rather than command following by the actuator. The PZT-only response and this suppression response are plotted in Fig. 9.17. The suppression response has a maximum peaking of 4.1 dB near 6.42 kHz.

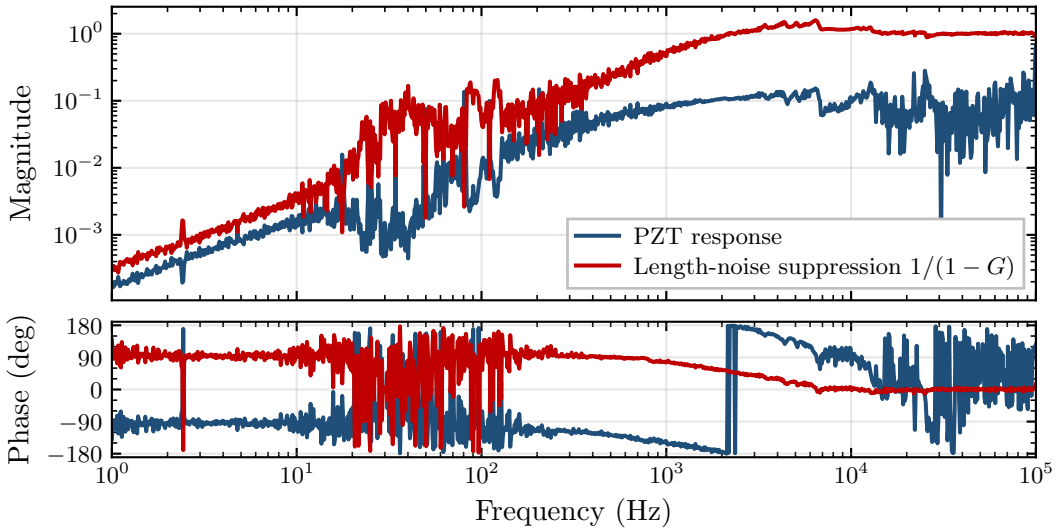


Figure 9.17: Inferred PZT-only response and closed-loop suppression response. The intermediate amplifier-plus-PZT plant is obtained by dividing the measured open-loop transfer function by the effective controller in Eq. 9.63; the plotted PZT-only response is then obtained by dividing out the amplifier gain. The closed-loop trace is $1/(1 - G_{\mathrm{PZT}})$, which gives the suppression of disturbances entering at the sensed cavity-length DOF.

Quantity	Value	Notes
PZT loop UGF	1.917 kHz	first unity-gain crossing
Phase margin	70.2°	phase at UGF
Conservative gain margin	8.9 dB	at 6.50 kHz
Additional FPGA PDH gain	+24 dB	included in K_{eff}
High-voltage amplifier gain	21 V/V	EVAL-ADHV4702-1 board

Table 9.6: Measured PZT loop parameters from the open-loop transfer function in Fig. 9.16. The gain margin is the smallest positive gain margin above the UGF.

The measured loop is below the measured cavity pole of 8.37 kHz. It is fast

enough to control drift of the aluminum spacer and cavity with the PZT-mounted M4 mirror and suppress some low-frequency noise on the order of a few kHz. The transfer-function measurement shows that the length actuator can keep the carrier on resonance without making the control loop part of the signal-sideband response. It also shows that the loop has enough influence to control the cavity at low frequency, avoiding injecting its length noise into the carrier laser through the laser frequency control loop.

9.11.2 Noise Spectra

The noise spectra in Fig. 9.18 were measured with the LFC locked to the laser at low frequency and the laser locked to the cavity at high frequency. The spectra give the displacement noise after the laser frequency controller. The spectra were calibrated as equivalent cavity-length motion using the linear version of the calibration described in Section 9.9.1 and then divided by two to report the equivalent motion of one cavity mirror. The first trace was taken in air before the spacer from Section 9.6 and the suspension were installed and before the PZT mirror damper was added, as discussed in Section 9.7.1. The second trace was taken after the spacer was mounted in the vacuum system, the PZT damper was added, and the vacuum was pumped down.

The large reduction below a few hundred hertz is the main commissioning result of this comparison. In the in-air configuration, acoustic pickup and table motion drive broad displacement noise and many narrow resonances. After the spacer was mounted in the vacuum system, that broadband motion was strongly reduced over much of the low-frequency band. The in-vacuum spacer spectrum has a 6.96×10^{-11} m RMS of equivalent single-mirror motion, or 0.139 nm RMS of round-trip length motion. Using the LFC FSR in Eq. 9.6, this is equivalent to 6.05 kHz RMS of cavity-resonance frequency motion. The remaining narrow lines are consistent with mechanical resonances, electronics pickup, or residual control-loop features rather than a single smooth cavity-noise process. These spectra are locked, in-loop displacement measurements: within the laser-frequency loop bandwidth, the laser follows the cavity resonance, so the calibrated post-controller signal gives the residual cavity motion seen by the lock.

The post-controller signal is the appropriate quantity for this comparison because it is the calibrated laser-frequency actuation required to keep the laser on the cavity resonance. When the laser-frequency loop has high bandwidth, the laser frequency

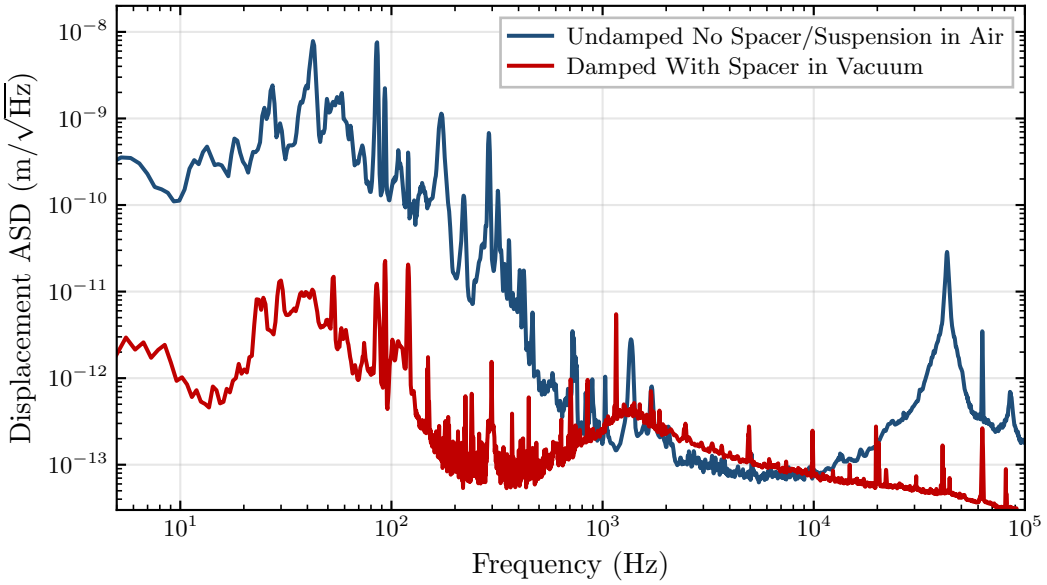


Figure 9.18: Measured LFC displacement ASDs with the cavity locked. The blue trace gives the earlier in-air configuration without the spacer/suspension damping or the mirror mount damping. The red trace gives the damped spacer configuration in vacuum. The low-frequency and acoustic-band motion are strongly reduced by the mechanical and vacuum configuration, while the two traces approach a similar high-frequency floor.

tracks the cavity resonance, so the actuator monitor gives a direct in-loop estimate of the cavity motion over that band. This interpretation relies on the independent scan calibration that converts laser-frequency actuation to equivalent round-trip length, and it is most reliable below the loop bandwidth where the actuator is actively following the cavity. At frequencies where the loop gain is small, i.e. at high frequencies, the post-controller signal should instead be read as the applied correction, not as a complete measurement of free-running cavity motion.

9.12 Future Laser Phase-Noise Measurements

The commissioning measurements in Section 9.11 demonstrate that the LFC can be locked and that the cavity displacement noise is reduced by the final mechanical configuration. A remaining measurement to fully validate the LFC's performance is a direct verification of the MHz laser phase-noise suppression that enters the GQuEST photon-counting noise budget. This measurement should compare the laser phase-noise spectrum before the LFC to the spectrum transmitted by the locked LFC. The same measurement should then be converted to the photon-counting equivalent using the output-filter pass-band integral in Eqs. 9.16 and 9.17. This gives both an optical verification of the cavity transfer function and the quantity

that matters directly for the GQuEST readout.

A suitable future setup for this measurement is the fiber delay-line interferometer in Fig. 9.19. The laser field is split by a 50:50 fiber beam splitter (BS). One copy travels through a short reference arm, while the other travels through a fiber delay line before the two fields are recombined. The two output ports are detected on PD_1 and PD_2, and the balanced difference signal is used as the phase-noise readout. Because common intensity noise appears nearly equally on both output ports, the balanced signal suppresses first-order laser intensity noise and is mainly sensitive to the phase difference between the present laser field and the delayed laser field.

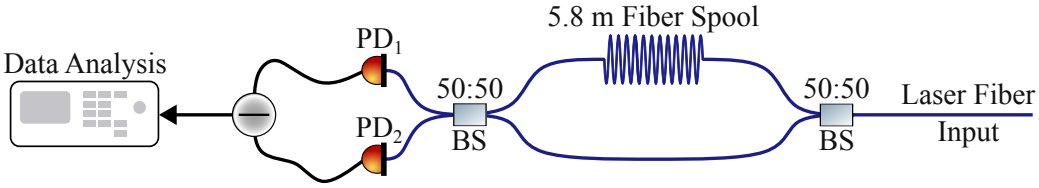


Figure 9.19: Proposed fiber delay-line interferometer for a future direct measurement of the LFC laser phase-noise suppression. The laser input is split by a 50:50 fiber BS. One arm is delayed by a 10 m fiber spool and then recombined with the undelayed arm on a second 50:50 BS. The two output ports are detected on the two PDs, and the differential signal is sent to the data-analysis chain. The static interferometer phase should be held at quadrature, corresponding to an optical path bias of $\lambda/4$ modulo $\lambda/2$, so that small laser phase fluctuations are converted linearly into differential power fluctuations.

The measurement principle can be seen by writing the laser field as

$$E(t) = E_0 e^{i[\omega_0 t + \phi(t)]}, \tag{9.65}$$

where $\phi(t)$ is the laser phase noise. If the delay-line arm has a group delay τ, the interferometer compares $\phi(t)$ with $\phi(t-\tau)$.

The delay-line interferometer measures laser phase noise by interfering the field with a delayed copy of itself. If the delay-line arm has group delay $\tau = n_g L_d/c$, then the interferometer senses

$$\delta\phi_\tau(t) = \phi(t) - \phi(t-\tau). \tag{9.66}$$

The balanced output converts this delayed phase difference into a differential power signal. For small phase fluctuations, the signal is largest when the static interferometer delay is at quadrature. Since the bias phase is $\Phi_0 = 2\pi\Delta L_{\text{bias}}/\lambda$, the optimal

bias condition is

$$\Phi_0 = \frac{\pi}{2} + m\pi \qquad \Rightarrow \qquad \Delta L_{\text{bias}} = \frac{(2m+1)\lambda}{4}. \tag{9.67}$$

Thus the interferometer should be held a quarter wavelength away from a bright or dark fringe. At this point the output is linear in small phase fluctuations; at a fringe, the first-order phase response vanishes.

The quarter-wave condition sets the static interferometer bias, while the fiber length sets the frequency response of the delay-line measurement. For a delay $\tau = n_g L_d/c$, delayed self-interference gives a phase-noise response proportional to

$$|G(f)|^2 = 4\sin^2(\pi f \tau). \tag{9.68}$$

The response is largest when $f\tau = (2k+1)/2$ and zero when $f\tau$ is an integer [170]. A delay optimized for a target frequency f_0 has a length of

$$L_{d,\text{opt}} = \frac{(2k+1)c}{2n_g f_0}, \qquad k = 0, 1, 2, \ldots. \tag{9.69}$$

For $\epsilon_r = 17.6$ MHz and $n_g \simeq 1.47$, the first optimum is

$$L_{d,\text{opt}} \simeq 5.8 \text{ m}, \tag{9.70}$$

with the next maximum near 17.4 m. This measurement should use a delay line close to one of these lengths, while the static interferometer bias is held separately at quadrature.

The measured before-and-after spectra should be compared with the expected filtered phase noise in Eq. 9.15. The final result would yield both the optical phase-noise spectrum and the photon-counting equivalent computed with the same output-filter integral used in Eq. 9.17. This would be a direct verification of the LFC performance in the frequency band relevant for GQuEST's photon counting readout and would confirm that the LFC is delivering the expected laser-noise suppression before the interferometer.

9.13 Integration with GQuEST

The LFC sits upstream of the interferometer and will be integrated with the rest of the laser input chain. The transmitted field must have enough power, the correct spatial mode, and low enough residual phase noise. The LFC control loop must

also be compatible with the interferometer control loops. Control noise introduced below the LFC or by the LFC locking noise must remain well below the MHz signal band, where the photon-counting readout is performed.

The power requirement has two stages. The demonstrator input power is 1 W, while the final input power is 15 W, as listed in Table 9.1. The LFC must work first as a commissioning cavity at moderate power and later as part of the full high-power laser chain. If the resonant transmission is > 90%, the final transmitted power would be > 13.5 W. The input power could be boosted to reach the 15 W target going into the interferometer.

The spatial-mode requirement is set by the PRC and interferometer input mode. The LFC transmitted beam should be mode matched into the interferometer with minimal higher-order-mode content. The design mode inside the LFC has a waist radius of 805 μm on the M1 mirror. The external mode-matching telescope should be straightforward to design because only light that fits the internal resonant mode can be transmitted through the cavity. A distorted or multi-peaked resonance scan would indicate misalignment and coupling to the higher-order modes in Fig. 9.5.

The phase-noise requirement is the main system-level requirement. The unfiltered laser phase-noise photon flux would be 2×10^8 Hz in the output-filter pass band. The LFC alone reduces this to 2×10^2 Hz, and the LFC plus PRC reduce it to 1×10^{-4} Hz. This is below the expected signal count rate for α of order unity and below the dominant classical background used in the photon-counting sensitivity estimate [1].

It is difficult to measure the LFC performance directly since the phase noise will be suppressed so much by the LFC. To measure the LFC performance, the laser phase-noise spectrum should be measured before and after the cavity with a delay-line interferometer setup seen in Fig. 9.19 or an independent reference cavity. The photon-counting equivalent can also be measured by comparing the count rate in the OFC pass band with and without the LFC in operation.

9.14 Summary

The LFC is a narrow-linewidth optical cavity that transmits the laser carrier while passively suppressing MHz laser-noise sidebands before they enter the GQuEST photon-counting readout. The carrier is held on resonance with a PDH control loop, while the laser-noise rejection is set by the cavity pole.

The four-mirror bow-tie cavity has a 4.45 m round-trip length and an FSR of

67.37 MHz. Transmitted-power ringdowns give $\mathcal{F}_{\mathrm{LFC}} = 4024.5 \pm 8.6$ and a cavity pole of $f_{\mathrm{pole,LFC}} = (8.370 \pm 0.018)$ kHz. This measured pole implies about 60 dB of LFC-only power suppression at the nominal readout offset. In the photon-flux model, the LFC reduces the output-filtered laser-noise photon flux from 2×10^{8} Hz to about 2×10^{2} Hz, and the LFC together with the PRC reduces it to about 1×10^{-4} Hz.

The mechanical and control design provides stable cavity operation in vacuum. The aluminum spacer defines the optical geometry and provides rigid damping for the cavity, and the slow PZT path removes long-term drift while the fast laser-frequency actuator maintains the PDH lock. The remaining system-level verification is a direct before-and-after laser phase-noise measurement using a delay-line interferometer.

BIBLIOGRAPHY

[1] S. M. Vermeulen, T. Cullen, D. Grass, I. A. O. MacMillan, A. J. Ramirez, J. Wack, B. Korzh, V. S. H. Lee, K. M. Zurek, C. Stoughton, et al., "Photon-counting interferometry to detect geontropic space-time fluctuations with GQuEST", Phys. Rev. X **15**, 011034 (2025).

[2] B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, et al. (LIGO Scientific Collaboration and Virgo Collaboration), "Observation of gravitational waves from a binary black hole merger", Phys. Rev. Lett. **116**, 061102 (2016).

[3] A. A. Michelson, *Détermination expérimentale de la valeur du mètre en longueurs d'ondes lumineuses* (Gauthier-Villars et fils, Paris, 1894).

[4] H. Kamerlingh Onnes, ""the liquefaction of helium."", in *Through measurement to knowledge: the selected papers of heike kamerlingh onnes 1853–1926*, edited by K. Gavroglu and Y. Goudaroulis (Springer Netherlands, Dordrecht, 1991), pp. 164–187.

[5] H. K. Onnes, "Further experiments with liquid helium. g. on the electrical resistance of pure metals, etc. vi. on the sudden change in the rate at which the resistance of mercury disappears.", in *Through measurement to knowledge: the selected papers of heike kamerlingh onnes 1853–1926*, edited by K. Gavroglu and Y. Goudaroulis (Springer Netherlands, Dordrecht, 1991), pp. 267–272.

[6] R. A. Millikan, "The isolation of an ion, a precision measurement of its charge, and the correction of stokes's law", Phys. Rev. (Series I) **32**, 349–397 (1911).

[7] R. A. Millikan., "On the elementary electrical charge and the avogadro constant", Phys. Rev. **2**, 109–143 (1913).

[8] C. T. R. Wilson, "On a method of making visible the paths of ionising particles through a gas", Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character **85**, 285–288 (1911).

[9] The Nobel Committee for Physics, *The nobel prize in physics 2017*, Awarded to R. Weiss, B. C. Barish, and K. S. Thorne "for decisive contributions to the LIGO detector and the observation of gravitational waves", Oct. 2017.

[10] L. McCuller, "Single-photon signal sideband detection for high-power michelson interferometers", (2022).

[11] A. Einstein, "Die grundlage der allgemeinen relativitätstheorie", Annalen der Physik **354**, 769–823 (1916).

[12] A. Einstein, "Zur elektrodynamik bewegter körper", Annalen der Physik **17**, 891–921 (1905).

[13] S. M. Carroll, *Spacetime and geometry: an introduction to general relativity* (Cambridge University Press, 2019).

[14] K. Schwarzschild, "Über das gravitationsfeld eines massenpunktes nach der einsteinschen theorie", Sitzungsberichte der königlich preussischen Akademie der Wissenschaften, 189–196 (1916).

[15] R. P. Kerr, "Gravitational field of a spinning mass as an example of algebraically special metrics", Phys. Rev. Lett. **11**, 237–238 (1963).

[16] A. Friedman, "Über die krümmung des raumes", Zeitschrift für Physik **10**, 377–386 (1922).

[17] G. Lemaître, "Un univers homogène de masse constante et de rayon croissant rendant compte de la vitesse radiale des nébuleuses extra-galactiques", Annales de la Société Scientifique de Bruxelles **A47**, 49–59 (1927).

[18] H. P. Robertson, "Kinematics and world-structure", Astrophysical Journal **82**, 284 (1935).

[19] A. G. Walker, "On milne's theory of world-structure", Proceedings of the London Mathematical Society **s2-42**, 90–127 (1937).

[20] F. W. Dyson, A. S. Eddington, and C. Davidson, "A determination of the deflection of light by the sun's gravitational field, from observations made at the total eclipse of may 29, 1919", Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character **220**, 291–333 (1920).

[21] R. V. Pound and G. A. Rebka Jr, "Apparent weight of photons", Physical review letters **4**, 337 (1960).

[22] C. M. Will, "The confrontation between general relativity and experiment", Living reviews in relativity **17**, 1–117 (2014).

[23] A. Einstein, "Näherungsweise integration der feldgleichungen der gravitation", Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften, 688–696 (1916).

[24] A. Einstein, "Über gravitationswellen", Sitzungsber. Preuss. Akad. Wiss. **1**, 154–167 (1918).

[25] A. Einstein and N. Rosen, "On gravitational waves", Journal of the Franklin Institute **223**, 43–54 (1937).

[26] D. H. Reitze, P. R. Saulson, and H. Grote, eds., *Advanced interferometric gravitational-wave detectors*, Vol. 5, 100 Years of General Relativity (World Scientific Publishing Co., 2019).

[27] P. C. Peters and J. Mathews, "Gravitational radiation from point masses in a keplerian orbit", Physical Review **131**, 435 (1963).

[28] C. Cutler and E. E. Flanagan, "Gravitational waves from merging compact binaries: how accurately can one extract the binary's parameters from the inspiral waveform?", Physical Review D **49**, 2658 (1994).

[29] M. B. Green, J. H. Schwarz, and E. Witten, *Superstring theory, Volume 1: introduction and volume 2: loop amplitudes, anomalies and phenomenology* (Cambridge University Press, Cambridge, UK, 1987).

[30] J. H. Schwarz, "Superstring theory", Physics Reports **89**, 223–322 (1982).

[31] C. Rovelli, "Loop quantum gravity", Living reviews in relativity **11**, 5 (2008).

[32] C. M. Caves, "Quantum-mechanical noise in an interferometer", Phys. Rev. D **23**, 1693–1708 (1981).

[33] E. P. Wigner, "Relativistic invariance and quantum phenomena", Rev. Mod. Phys. **29**, 255–268 (1957).

[34] H. Salecker and E. P. Wigner, "Quantum limitations of the measurement of space-time distances", Phys. Rev. **109**, 571–577 (1958).

[35] Y. J. Ng and H. van Dam, "Limit to space-time measurement", Modern Physics Letters A **9**, 335–340 (1994).

[36] D. Li, V. S. H. Lee, Y. Chen, and K. M. Zurek, "Interferometer response to geontropic fluctuations", Phys. Rev. D **107**, 024002 (2023).

[37] B. Sharmila, S. M. Vermeulen, and A. Datta, "Signatures of correlation of spacetime fluctuations in laser interferometers", Nature Communications **17**, 701 (2025).

[38] G. Amelino-Camelia, "Gravity-wave interferometers as quantum-gravity detectors", Nature **398**, 216–218 (1999).

[39] Y. J. Ng, "Selected topics in Planck-scale physics", Modern Physics Letters A **18**, 1073–1097 (2003).

[40] F. Karolyhazy, "Gravitation and quantum mechanics of macroscopic objects", Il Nuovo Cimento A (1965-1970) **42**, 390–402 (1966).

[41] E. P. Verlinde and K. M. Zurek, "Observational signatures of quantum gravity in interferometers", Physics Letters B **822**, 136663 (2021).

[42] K. M. Zurek, "Snowmass 2021 white paper: observational signatures of quantum gravity", arXiv preprint arXiv:2205.01799 (2022).

[43] L. J. Garay, "Quantum gravity and minimum length", International Journal of Modern Physics A **10**, 145–165 (1995).

[44] S. Hossenfelder, "Minimal length scale scenarios for quantum gravity", Living Reviews in Relativity **16**, 2 (2013).

[45] G. Amelino-Camelia, "Quantum-spacetime phenomenology", Living Reviews in Relativity **16**, 5 (2013).

[46] S. M. Vermeulen, L. Aiello, A. Ejlli, W. L. Griffiths, A. L. James, K. L. Dooley, and H. Grote, "An experiment for observing quantum gravity phenomena using twin table-top 3d interferometers", Classical and Quantum Gravity **38**, 085008 (2021).

[47] K. M. Zurek, "On vacuum fluctuations in quantum gravity and interferometer arm fluctuations", Physics Letters B **826**, 136910 (2022).

[48] O. Kwon, "Phenomenology of holography via quantum coherence on causal horizons", Foundations of Physics **55**, 1–42 (2025).

[49] E. Verlinde and K. M. Zurek, "Modular fluctuations from shockwave geometries", Physical Review D **106**, 106011 (2022).

[50] M. W. Bub, Y. Chen, Y. Du, D. Li, Y. Zhang, and K. M. Zurek, "Quantum gravity background in next-generation gravitational wave detectors", Phys. Rev. D **108**, 064038 (2023).

[51] D. Reitze, R. X. Adhikari, S. Ballmer, B. Barish, L. Barsotti, G. Billingsley, D. A. Brown, Y. Chen, D. Coyne, R. Eisenstein, et al., "Cosmic explorer: the u.s. contribution to gravitational-wave astronomy beyond ligo", (2019).

[52] A. Abac, R. Abramo, S. Albanesi, A. Albertini, A. Agapito, M. Agathos, C. Albertus, N. Andersson, T. Andrade, I. Andreoni, et al., "The science of the einstein telescope", Journal of Cosmology and Astroparticle Physics **2026**, 081 (2026).

[53] R. A. Hulse and J. Taylor, "A high-sensitivity pulsar survey", Astrophysical Journal, vol. 191, p. L59 **191**, L59 (1974).

[54] R. A. Hulse and J. H. Taylor, "Discovery of a pulsar in a binary system", Astrophysical Journal, vol. 195, Jan. 15, 1975, pt. 2, p. L51-L53. **195**, L51–L53 (1975).

[55] J. H. Taylor, L. A. Fowler, and P. M. McCulloch, "Measurements of general relativistic effects in the binary pulsar PSR1913+ 16", Nature **277**, 437–440 (1979).

[56] J. Weber, "Gravitational radiation", Phys. Rev. Lett. **18**, 498–501 (1967).

[57] J. Weber, "Gravitational-wave-detector events", Phys. Rev. Lett. **20**, 1307–1308 (1968).

[58] J. Weber, "Evidence for discovery of gravitational radiation", Phys. Rev. Lett. **22**, 1320–1324 (1969).

[59] M. C. Miller and N. Yunes, "The new frontier of gravitational waves", Nature **568**, 469–476 (2019).

[60] R. Weiss, *Electromagnetically coupled broadband gravitational antenna*, Research Laboratory of Electronics Quarterly Progress Report 105, Reprinted in *General Relativity and Gravitation* **54**, 153 (2022) (Massachusetts Institute of Technology, 1972), pp. 54–76.

[61] R. W. P. Drever, F. J. Raab, K. S. Thorne, R. Vogt, and R. Weiss, *A laser interferometer gravitational-wave observatory (LIGO)*, Proposal to the National Science Foundation LIGO-M890001, Original proposal for the Laser Interferometer Gravitational-Wave Observatory (California Institute of Technology and Massachusetts Institute of Technology, 1989).

[62] D. Ganapathy, W. Jia, M. Nakano, V. Xu, N. Aritomi, T. Cullen, N. Kijbunchoo, S. E. Dwyer, A. Mullavey, L. McCuller, et al. (LIGO O4 Detector Collaboration), "Broadband quantum enhancement of the ligo detectors with frequency-dependent squeezing", Phys. Rev. X **13**, 041021 (2023).

[63] K. Arai, *On the accumulated round-trip Gouy phase shift for a general optical cavity*, Technical Note LIGO-T1300189-v1 (LIGO Scientific Collaboration, Nov. 2013).

[64] S. Gigan, L. Lopez, N. Treps, A. Maître, and C. Fabre, "Image transmission through a stable paraxial cavity", Phys. Rev. A **72**, 023804 (2005).

[65] Y. Guo, "Optical strategies for gravitational-wave detector enhancement", PhD thesis (University of Amsterdam, Amsterdam, Netherlands, Mar. 2025).

[66] R. W. Drever, J. L. Hall, F. V. Kowalski, J. Hough, G. Ford, A. Munley, and H. Ward, "Laser phase and frequency stabilization using an optical resonator", Applied Physics B **31**, 97–105 (1983).

[67] E. D. Black, "An introduction to pound-drever-hall laser frequency stabilization", American Journal of Physics **69**, 79–87 (2001).

[68] D. Grass, S. M. Vermeulen, I. A. O. MacMillan, and L. McCuller, "High-frequency thermal noise in michelson interferometers", arXiv preprint, 10.48550/arXiv.2605.16710 (2026).

[69] F. Meylahn and B. Willke, "Characterization of laser systems at 1550 nm wavelength for future gravitational wave detectors", Instruments **6**, 10.3390/instruments6010015 (2022).

[70] H. B. Callen and T. A. Welton, "Irreversibility and generalized noise", Phys. Rev. **83**, 34–40 (1951).

[71] Y. Levin, "Internal thermal noise in the ligo test masses: a direct approach", Phys. Rev. D **57**, 659–663 (1998).

[72] A. Chou, H. Glass, H. R. Gustafson, C. Hogan, B. L. Kamai, O. Kwon, R. Lanza, L. McCuller, S. S. Meyer, J. Richardson, et al., "The holometer: an instrument to probe planckian quantum geometry", Classical and Quantum Gravity **34**, 065005 (2017).

[73] G. M. Harry, A. M. Gretarsson, P. R. Saulson, S. E. Kittelberger, S. D. Penn, W. J. Startin, S. Rowan, M. M. Fejer, D. R. M. Crooks, G. Cagnoli, et al., "Thermal noise in interferometric gravitational wave detectors due to dielectric optical coatings", Classical and Quantum Gravity **19**, 897 (2002).

[74] G. M. Harry, H. Armandula, E. Black, D. R. M. Crooks, G. Cagnoli, J. Hough, P. Murray, S. Reid, S. Rowan, P. Sneddon, et al., "Thermal noise from optical coatings in gravitational wave detectors", Appl. Opt. **45**, 1569–1574 (2006).

[75] M. M. Fejer, S. Rowan, G. Cagnoli, D. R. M. Crooks, A. Gretarsson, G. M. Harry, J. Hough, S. D. Penn, P. H. Sneddon, and S. P. Vyatchanin, "Thermoelastic dissipation in inhomogeneous media: loss measurements and displacement noise in coated test masses for interferometric gravitational wave detectors", Phys. Rev. D **70**, 082003 (2004).

[76] M. Evans, S. Ballmer, M. Fejer, P. Fritschel, G. Harry, and G. Ogin, "Thermo-optic noise in coated mirrors for high-precision optical measurements", Phys. Rev. D **78**, 102003 (2008).

[77] J. W. Gardner, T. Gefen, S. A. Haine, J. J. Hope, J. Preskill, Y. Chen, and L. McCuller, "Stochastic waveform estimation at the fundamental quantum limit", PRX Quantum **6**, 030311 (2025).

[78] E. Payne, L. McCuller, and K. Chatziioannou, "Photon temporal-mode readout for inference of neutron star merger remnant gravitational waves", Phys. Rev. D **113**, 084065 (2026).

[79] E. Capote, W. Jia, N. Aritomi, M. Nakano, V. Xu, R. Abbott, I. Abouelfettouh, R. X. Adhikari, A. Ananyeva, S. Appert, et al., "Advanced LIGO detector performance in the fourth observing run", Physical Review D **111**, 062002 (2025).

[80] LIGO Scientific Collaboration, "Advanced LIGO", Classical and Quantum Gravity **32**, 074001 (2015).

[81] F. Matichard, B. Lantz, K. Mason, R. Mittleman, B. Abbott, S. Abbott, E. Allwine, S. Barnum, J. Birch, S. Biscans, et al., "Advanced ligo two-stage twelve-axis vibration isolation and positioning platform. part 1: design and production overview", Precision Engineering **40**, 273–286 (2015).

[82] J. Harms, "Terrestrial gravity fluctuations", Living Reviews in Relativity **18**, 3 (2015).

[83] S. L. Danilishin and F. Y. Khalili, "Quantum measurement theory in gravitational-wave detectors", Living Reviews in Relativity **15**, 5 (2012).

[84] D. V. Martynov, "Lock acquisition and sensitivity analysis of advanced ligo interferometers", PhD thesis (California Institute of Technology, 2015).

[85] A. Buonanno and Y. Chen, "Quantum noise in second generation, signal-recycled laser interferometric gravitational-wave detectors", Physical Review D **64**, 042006 (2001).

[86] P. Nguyen, R. M. S. Schofield, A. Effler, C. Austin, V. Adya, M. Ball, S. Banagiri, K. Banowetz, C. Billman, C. D. Blair, et al., "Environmental noise in advanced ligo detectors", Classical and Quantum Gravity **38**, 145001 (2021).

[87] J. C. Maxwell, "On governors", Proceedings of the Royal Society of London **16**, 270–283 (1868).

[88] O. Mayr, "Victorian physicists and speed regulation: an encounter between science and technology", Notes and records of the Royal Society of London **26**, 205–228 (1971).

[89] L. Rayleigh, *The theory of sound*, 1st ed., Vol. 1 (MacMillan and Co., London, 1877).

[90] T. Shinbrot, C. Grebogi, J. Wisdom, and J. A. Yorke, "Chaos in a double pendulum", American Journal of Physics **60**, 491–499 (1992).

[91] T. Lenart, "Design of reconfigurable hardware architectures for real-time applications", PhD thesis (Department of Electrical and Information Technology, Lund University, 2008).

[92] IEEE, *Ieee standard for floating-point arithmetic*, July 2019.

[93] M. Ortmanns, A. Buhmann, and Y. Manoli, "Interface circuits", in *Comprehensive microsystems*, Vol. 1, edited by Y. B. Gianchandani, O. Tabata, and H. P. Zappe (Elsevier, 2007) Chap. 1.16, pp. 465–493.

[94] M. Evans, *Digital filter noise*, LIGO-G0900928-v1 (Oct. 2009).

[95] T. Chang, "A unified analysis of roundoff noise reduction in digital filters", in Icassp '81. ieee international conference on acoustics, speech, and signal processing, Vol. 6 (1981), pp. 1209–1212.

[96] K. Yao, "Minimum round-off noise second-order digital filter with practical complexity constraints", in International telemetering conference proceedings, Vol. 14 (Nov. 1978).

[97] P. Benner, V. Mehrmann, V. Sima, S. Van Huffel, and A. Varga, "SLICOT: A Subroutine Library in Systems and Control Theory", in *Applied and Computational Control, Signals, and Circuits: Volume 1*, edited by B. N. Datta (Birkhäuser, Boston, MA, 1999), pp. 499–539.

[98] A. Varga and K. H. Fasol, "A New Square-root Balancing-Free Stochastic Truncation Model Reduction Algorithm", IFAC Proceedings Volumes, 12th Triennal Wold Congress of the International Federation of Automatic Control. Volume 1 Theory, Sydney, Australia, 18-23 July **26**, 469–472 (1993).

[99] R. L. Ward, R. Adhikari, B. Abbott, R. Abbott, D. Barron, R. Bork, T. Fricke, V. Frolov, J. Heefner, A. Ivanov, et al., "Dc readout experiment at the caltech 40m prototype interferometer", Classical and Quantum Gravity **25**, 114030 (2008).

[100] I. A. O. MacMillan and L. P. McCuller, "Robust bilinear-noise-optimal control for gravitational-wave detectors: a mixed lqg and $\mathcal{H}_\infty$ approach", Phys. Rev. D **113**, 082001 (2026).

[101] J. Doyle, K. Glover, P. Khargonekar, and B. Francis, "State-space solutions to standard H_2 and H_∞ control problems", IEEE Trans. Autom. Control **34**, 831–847 (1989).

[102] V. Ionescu and M. Weiss, "The l^2-control problem for time-varying discrete systems", Systems & control letters **18**, 371–381 (1992).

[103] A. Varga, "On Computing 2-Norms of Transfer Function Matrices", in 1992 Am. Control Conf. (June 1992), pp. 1054–1056.

[104] A. C. Antoulas, "6. Sylvester and Lyapunov Equations", in *Approximation of Large-Scale Dynamical Systems*, Advances in Design and Control (Society for Industrial and Applied Mathematics, Jan. 2005), pp. 173–204.

[105] S. J. HAMMARLING, "Numerical Solution of the Stable, Non-negative Definite Lyapunov Equation Lyapunov Equation", IMA Journal of Numerical Analysis **2**, 303–323 (1982).

[106] M. Athans and E. Tse, "A direct derivation of the optimal linear filter using the maximum principle", IEEE Trans. Autom. Control **12**, 690–698 (1967).

[107] A. Laub, "A Schur method for solving algebraic Riccati equations", IEEE Trans. Autom. Control **24**, 913–921 (1979).

[108] A. J. Laub, "Invariant Subspace Methods for the Numerical Solution of Riccati Equations", in *The Riccati Equation*, edited by S. Bittanti, A. J. Laub, and J. C. Willems, Communications and Control Engineering Series (Springer, Berlin, Heidelberg, 1991), pp. 163–196.

[109] P. Van Dooren, "A Generalized Eigenvalue Approach for Solving Riccati Equations", SIAM J. Sci. and Stat. Comput. **2**, 121–135 (1981).

[110] R. James, J. Langou, and B. R. Lowery, *On matrix balancing and eigenvector computation*, Jan. 2014.

[111] B. N. Parlett and C. Reinsch, "Balancing a matrix for calculation of eigenvalues and eigenvectors", Numer. Math. **13**, 293–304 (1969).

[112] M. Şuşcă, V. Mihaly, M. Stănese, and P. Dobra, "Iterative Refinement Procedure for Solutions to Algebraic Riccati Equations", in 2020 IEEE Int. Conf. Autom. Qual. Test. Robot. AQTR (May 2020), pp. 1–5.

[113] A. Varga, "Computation of irreducible generalized state-space realizations", Kybernetika **26**, 89–106 (1990).

[114] M. G. Safonov, D. J. N. Limebeer, and R. Y. Chiang, "Simplifying the H_∞ theory via loop-shifting, matrix-pencil and descriptor concepts", Int. J. Control **50**, 2467–2488 (1989).

[115] K. Zhou, J. C. Doyle, and K. Glover, *Robust and optimal control* (Prentice Hall, 1996).

[116] D. Lemonnier and P. Van Dooren, "Balancing Regular Matrix Pencils", SIAM J. Matrix Anal. & Appl. **28**, 253–263 (2006).

[117] R. C. Ward, "Balancing the Generalized Eigenvalue Problem", SIAM J. Sci. and Stat. Comput. **2**, 141–152 (1981).

[118] Y. Liu and B. Anderson, "Singular perturbation approximation of balanced systems", in Proc. 28th IEEE Conf. Decis. Control (Dec. 1989), 1355–1360 vol.2.

[119] N. A. Bruinsma and M. Steinbuch, "A fast algorithm to compute the H_∞-norm of a transfer function matrix", Systems & Control Letters **14**, 287–293 (1990).

[120] K. Glover and J. C. Doyle, "State-space formulae for all stabilizing controllers that satisfy an H_∞-norm bound and relations to relations to risk sensitivity", Systems & Control Letters **11**, 167–172 (1988).

[121] J.-L. Hong and C.-C. Teng, "A derivation of the glover-doyle algorithms for general H_∞ control problems", Automatica **32**, 581–589 (1996).

[122] P. M. Gahinet, "Troubleshooting for state-space H_∞ design", in 1992 Am. Control Conf. (June 1992), pp. 634–638.

[123] D. Bernstein and W. Haddad, "LQG control with an H_∞ performance bound: a Riccati equation approach", IEEE Trans. Autom. Control **34**, 293–305 (1989).

[124] W. Haddad and D. Bernstein, "Generalized Riccati equations for the full- and reduced-order mixed-norm H_2/H_∞ standard problem", in Proc. 28th IEEE Conf. Decis. Control (Dec. 1989), 397–402 vol.1.

[125] J. Doyle, K. Zhou, and B. Bodenheimer, "Optimal control with mixed H_2 and H_∞ performance objectives", in 1989 Am. Control Conf. (June 1989), pp. 2065–2070.

[126] H.-H. Yeh, S. Banda, and B.-C. Chang, "Necessary and sufficient conditions for mixed H_2 and H_∞ optimal control", IEEE Trans. Autom. Control **37**, 355–358 (1992).

[127] P. Khargonekar and M. Rotea, "Mixed H_2/H_∞ control: a convex optimization approach", IEEE Trans. Autom. Control **36**, 824–837 (1991).

[128] M. Rotea and T. Iwasaki, "Generalized H_2/H_∞ control: general formulas", in Proc. 1995 34th IEEE Conf. Decis. Control, Vol. 3 (Dec. 1995), 3158–3163 vol.3.

[129] M. A. Rotea and P. P. Khargonekar, "Generalized H_2/H_∞ Control", in Robust Control Theory, edited by B. A. Francis and P. P. Khargonekar, The IMA Volumes in Mathematics and Its Applications (1995), pp. 81–103.

[130] V. Kapila and W. M. Haddad, "H_2 and mixed H_2/H_∞ stable stabilization", in Proceedings of the 34th conference on decision and control, Vol. 2 (1995), pp. 1911–1916.

[131] Y. W. Wang, W. M. Haddad, and D. S. Bernstein, "Stable stabilization with H_2 and H_∞ performance constraints", Journal of Mathematical Systems, Estimation, and Control **6**, 181–194 (1996).

[132] M. Athans, "The matrix minimum principle", Information and Control **11**, 592–606 (1967).

[133] A. Buikema, C. Cahillane, G. L. Mansell, C. D. Blair, R. Abbott, C. Adams, R. X. Adhikari, A. Ananyeva, S. Appert, K. Arai, et al., "Sensitivity and performance of the advanced ligo detectors in the third observing run", Phys. Rev. D **102**, 062003 (2020).

[134] R. Abbott, T. Abbott, S. Abraham, F. Acernese, K. Ackley, A. Adams, C. Adams, R. Adhikari, V. Adya, C. Affeldt, et al., "GWTC-2: compact binary coalescences observed by LIGO and Virgo during the first half of the third observing run", Physical Review X **11**, 021053 (2021).

[135] R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, N. Adhikari, R. X. Adhikari, V. B. Adya, C. Affeldt, D. Agarwal, et al. (LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration), "GWTC-3: compact binary coalescences observed by LIGO and Virgo during the second part of the third observing run", Physical Review X **13**, 041039 (2023).

[136] R. Abbott, T. Abbott, S. Abraham, F. Acernese, K. Ackley, A. Adams, C. Adams, R. Adhikari, V. Adya, C. Affeldt, et al., "Tests of general relativity with binary black holes from the second LIGO-Virgo gravitational-wave transient catalog", Physical Review D **103**, 122002 (2021).

[137] R. Abbott, H. Abe, F. Acernese, K. Ackley, N. Adhikari, R. X. Adhikari, V. K. Adkins, V. B. Adya, C. Affeldt, D. Agarwal, et al. (The LIGO Scientific Collaboration, the Virgo Collaboration, and the KAGRA Collaboration), "Tests of general relativity with GWTC-3", Physical Review D **112**, 084080 (2025).

[138] E. D. Hall, K. Kuns, J. R. Smith, Y. Bai, C. Wipf, S. Biscans, R. X. Adhikari, K. Arai, S. Ballmer, L. Barsotti, et al., "Gravitational-wave physics with cosmic explorer: limits to low-frequency sensitivity", Phys. Rev. D **103**, 122004 (2021).

[139] L. S. Finn and D. F. Chernoff, "Observing binary inspiral in gravitational radiation: One interferometer", Physical Review D **47**, 2198–2219 (1993).

[140] L. McCuller, *Wield iirrational*, `https://github.com/wieldphysics/wield-iirrational`, 2023.

[141] Y. Nakatsukasa, O. Sète, and L. N. Trefethen, "The AAA Algorithm for Rational Approximation", SIAM J. Sci. Comput. **40**, A1494–A1522 (2018).

[142] J. Doyle, "Guaranteed margins for LQG regulators", IEEE Trans. Autom. Control **23**, 756–757 (1978).

[143] J. Buchli, B. Tracey, T. Andric, C. Wipf, Y. H. J. Chiu, M. Lochbrunner, C. Donner, R. X. Adhikari, J. Harms, I. Barr, et al., "Improving cosmological reach of a gravitational wave observatory using deep loop shaping", Science **389**, 1012–1015 (2025).

[144] I. Galinskiy, Y. Tsaturyan, M. Parniak, and E. S. Polzik, "Phonon counting thermometry of an ultracoherent membrane resonator near its motional ground state", Optica **7**, 718–725 (2020).

[145] D. D. Brown, A. Freise, H. T. Cao, A. Ciobanu, J. Gobeil, A. Green, P. Hapke, P. Jones, M. van der Kolk, K. Kuns, et al., *Finesse*, version 3.0a32, Mar. 2025.

[146] V. Srivastava, G. Mansell, C. Makarem, M. Noh, R. Abbott, S. Ballmer, G. Billingsley, A. Brooks, H. T. Cao, P. Fritschel, et al., "Piezo-deformable mirrors for active mode matching in advanced ligo", Opt. Express **30**, 10491–10501 (2022).

[147] J. R. Lawall, "High resolution determination of radii of curvature using fabry-perot interferometry", Measurement Science and Technology **20**, 045302 (2009).

[148] J. Kim, M. Lee, D. Kim, W. Seo, H.-G. Hong, Y. Song, and K. An, "Prescribed nondegenerate high-order modes in an axial-asymmetric high-finesse fabry–perot microcavity", Opt. Lett. **37**, 1457–1459 (2012).

[149] I. Nardecchia, Y. Minenkov, M. Lorenzini, L. Aiello, E. Cesarini, D. Lumaca, V. Malvezzi, F. Paoletti, A. Rocchi, and V. Fafone, "Optimized radius of curvature tuning for the virgo core optics", Classical and Quantum Gravity **40**, 055004 (2023).

[150] I. MacMillan and L. McCuller, *Buzz*, `https://git.ligo.org/ian.macmillan/buzz`, 2023.

[151] I. A. O. MacMillan, *Moth: mode-matching optimization thing*, version 2.0.0, Apr. 2026.

[152] Y. Drori, J. Eichholz, T. Edo, H. Yamamoto, Y. Enomoto, G. Venugopalan, K. Arai, and R. X. Adhikari, "Scattering loss in precision metrology due to mirror roughness", J. Opt. Soc. Am. A **39**, 969–978 (2022).

[153] O. Wipfli, H. F. Passagem, C. Fischer, M. Grau, and J. P. Home, "Integration of a high finesse cryogenic build-up cavity with an ion trap", Review of Scientific Instruments **94**, 083204 (2023).

[154] J. Vanosky and G. Moreno, *General class a fluoroelastomers processing procedure*, Engineering Note LIGO-E2100086 (LIGO Laboratory, 2021).

[155] Y. Jiang, A. Ludlow, N. D. Lemke, R. W. Fox, J. A. Sherman, L.-S. Ma, and C. W. Oates, "Making optical atomic clocks more stable with 10- 16-level laser stabilization", Nature Photonics **5**, 158–161 (2011).

[156] T. Legero, T. Kessler, and U. Sterr, "Tuning the thermal expansion properties of optical reference cavities with fused silica mirrors", J. Opt. Soc. Am. B **27**, 914–919 (2010).

[157] M. Notcutt, L.-S. Ma, A. D. Ludlow, S. M. Foreman, J. Ye, and J. L. Hall, "Contribution of thermal noise to frequency stability of rigid optical cavity via hertz-linewidth lasers", Phys. Rev. A **73**, 031804(R) (2006).

[158] Aalco Metals Limited, *Aluminium Alloy – Commercial Alloy – 6082 – T6~T651 Plate*, Material datasheet for EN AW-6082 / DIN 3.2315 / ISO AlSi1MgMn, n.d.

[159] Hydro, *Technical Datasheet – Extruded Products: Alloy EN AW-6082 [AlSi1MgMn]*, Hydro 6082 Rev. 4, 03/2023; material datasheet for EN AW-6082, 2023.

[160] Pfeiffer Vacuum, *Vacuum Chamber Design and Function*, Vacuum technology knowledge base section 3.4 on vacuum chambers and surface treatment, n.d.

[161] M. Sullivan, *Vacuum Optics*, Presentation slides, December 2, 2006; hosted by University of Arizona Optomechanics, 2006.

[162] COMSOL, *COMSOL Multiphysics®*, version 6.4, Stockholm, Sweden, Nov. 2025.

[163] D. J. Inman, *Engineering vibration*, 4th ed. (Pearson, 2014).

[164] McMaster-Carr, *Chemical-Resistant Viton Fluoroelastomer Rubber Balls, 3/4 in Diameter, 75A Durometer, 3645K7*, 2026.

[165] A. N. Gent, "On the relation between indentation hardness and young's modulus", Rubber Chemistry and Technology **31**, 896–906 (1958).

[166] I. M. Meththananda, S. Parker, M. P. Patel, and M. Braden, "The relationship between shore hardness of elastomeric dental materials and young's modulus", Dental Materials **25**, 956–959 (2009).

[167] K. L. Johnson, *Contact mechanics* (Cambridge University Press, 1985).

[168] V. L. Popov, M. Heß, and E. Willert, *Handbook of contact mechanics: exact solutions of axisymmetric contact problems* (Springer, 2019).

[169] J. C. Snowdon, *Vibration isolation: use and characterization*, NBS Handbook 128 (National Bureau of Standards, 1979).

[170] M. Parniak, I. Galinskiy, T. Zwettler, and E. S. Polzik, "High-frequency broadband laser phase noise cancellation using a delay line", Opt. Express **29**, 6935–6946 (2021).

INDEX

A

AB09ND, 101, 103, 106, 107, 117

AB13BD, 113

AB13DD, 121

Acousto-optic modulators (AOMs), 182, 198, 202

  double-pass geometry, 182, 198

  frequency calibration, 202–203

Algebraic Riccati equations, 111–123, 129–134, 161

Alignment sensing and control (ASC), vii, 138–172

  DHARD yaw, 138–172

  hard and soft modes, 145

Alignment-to-length coupling, 65, 146, 149, 152

B

Bilinear noise, 134, 135, 138–156

  spectral convolution, 154–156

Binary black holes (BBHs), 4, 12

Binary neutron stars (BNS), vii, 12, 55, 140, 149, 151

  detection range, vii, 140–172

C

Caltech 40m prototype interferometer, 107

Cavity length noise, 178–184, 213, 233–244, 260–261

Christoffel symbols, 9, 10, 12

COMSOL Multiphysics®, 238

Contrast defect, 27, 176

Control theory, 4, 67–86, 109, 138, 198–205, 253–260

Controllability, 72, 113

Controller implementation, 88–108, 116–120, 133, 160–170, 198–205, 253–260

  dynamic implementation, 160

Controls noise, vii, 4, 110–137, 141–172, 263

Convex optimization, 130, 169–172

Counterpropagating beams, 44, 184

D

Deep loop shaping (DLS), 167–172
Differential arm length (DARM), 25–27, 141–172
Digital filters, vii, 4, 88–108
    architecture selection, vii, 105–108
    Direct Form II (DFII), 92–94, 105
    filter realizations, vii, 88
    Low-Noise Biquad Form I, 94–97, 105
    Low-Noise Biquad Form II, 97–99, 105
    state-space realizations, 99–101, 105
Digital signal processing (DSP), 4, 88

E
Einstein field equation, 10, 12, 13

F
Fabry-Pérot cavities, 40
Feedback control, 4, 67, 73–86, 109–172, 198–205, 253–261
    closed-loop transfer function, 76–85, 146–172, 257–260
    disturbance rejection, 4, 74–77
    in-loop measurements, 260
    lock-loss threshold, 143
    loop gain, 77, 118
    negative feedback, 73
    open-loop transfer function, 77–85, 118, 146–172, 257–260
    PID controllers, 85–86
Figures of merit (FOMs), 110–137, 139–172
    broadband RMS, 149–154
Finesse, 32–33, 177–178, 209–210, 220, 244–248
Floating-point arithmetic, 4, 88, 101, 102
    IEEE 754, 88
    precision, 4, 88, 101, 102
    rounding, 89
Fluctuation-dissipation theorem (FDT), 51–52
Foton, 105
Free spectral range (FSR), 31–34, 178–184, 209–210, 220–223, 248–253
Frequency-noise calibration, 202, 213, 248, 260
Friedmann-Lemaître-Robertson-Walker metric, 10

G
Geometric units, 9
GQuEST, vii, 1, 22, 42–61, 263
  experimental design, 44–48
  photon-counting readout, vii, 1, 43–61, 65, 174–218, 263
    correlated readout, 59–61
    count statistics, 43, 56, 58
    dark counts and optical backgrounds, 58, 215, 216
  readout band, 42–61, 175
  signal and null channels, 45, 59–61
  twin-interferometer correlation, 45, 59–61
Gravitational waves (GWs), 11–17
  linearized gravity, 12–15
  polarizations, 14–15
    cross, 14–15
    plus, 14–15, 27
  strain, 15–17, 27
Gravitational-wave detection, 1, 4, 11–17, 28–42, 138–172
Gravity
  Loop quantum gravity, 17, 21, 22
  Quantum gravity, 17, 18, 21, 22

H
$\mathcal{H}_\infty$ control, vii, 120–123, 129–137, 160–172
  gamma bound, 120–123, 129–137, 161–167
$\mathcal{H}_2$ control, vii, 113–120, 129–137, 146–172
Heisenberg uncertainty principle, 19
Higher-order modes (HOMs), viii, 34, 184–186, 205–209, 211, 216, 226, 251
  laser filter cavity, 251
Homodyne readout, 43, 55

I
Interconnected systems, 73–74, 111
Interferometers (IFOs), vii, 3–4, 18–22, 25–66, 138–172
  dark-fringe operation, 25–27
  Michelson interferometer, 25–27

K

Kerr solution, 10

L

Laplace transform, 70
Laser filter cavity (LFC), vii, 3, 44, 219–265
    cavity scans, 248–253
    control system, 253–261
    design requirements, 220–222
    filter response, 51, 222–223
    finesse and ringdown, vii, 244–248
    GQuEST integration, vii, 263–264
    laser phase-noise measurement, 261–263
    locked noise spectra, 257, 260–261
    mechanical resonances, 238–239, 243–244
    optical layout, 223–230
    spacer design, 233–239
    thermal drift, 236–237
    vacuum and bakeout, 230–233
Laser Interferometer Gravitational-Wave Observatory (LIGO), vii, 1, 4, 28–42, 138–172
Laser noise, vii, 3, 49, 219, 261
    amplitude noise, 219
    phase noise, vii, 49–51, 219, 261
Laser-frequency control, 29, 198, 253
Linear time-invariant (LTI) systems, 70–73, 99–101, 111–137
Linear-quadratic-Gaussian (LQG) control, 114–137, 146–172
Lyapunov equation, 113, 123–129
    covariance and RMS, 123–129, 157

M

Metric tensor, 7–11
Mirror coatings, 54, 186–188, 228
Mixed-sensitivity control, vii, 111–137, 139–172
    actuator-centered formulation, 135
    holonomy solution, 132
    measurement-only formulation, 137
    Riccati formulation, 131–134

strong stability, 133
total-noise formulation, 136
Model reduction, 101, 117, 120, 133
singular perturbation, 101, 117, 120

N
Newtonian noise, 63
Noise budgets, 2, 48–66, 140–172, 174–184, 205–214, 219–223, 263–265
coupling functions, 61, 149–154, 222
GQuEST, 48–55, 174, 220, 261
Numerical conditioning, vii, 101, 116, 118, 120, 151
dynamic range, 101, 116, 151
Numerical noise, 4, 88–108

O
Observability, 73, 113
Optical cavities, 3, 28–42, 174–265
ABCD matrices, 33
cavity pole, 32, 177, 220
Fabry–Pérot geometry, 28
Gaussian eigenmode, 34, 184, 226
geometric stability, 33, 184
Gouy phase, 33, 184, 206, 226
linewidth, 33, 177, 220, 251
resonance condition, 31, 233
ringdown, 209, 244
storage time, 32, 244
traveling-wave and bow-tie geometries, 35, 184, 226
Optical filtering, 3, 44–61, 174–265
Optimal control, vii, 4, 109–172
noise-shaping filters, 110, 118, 147
weighting filters, 110, 149–154
well-posed problems, 116–120, 146
Output filter cavities (OFCs), viii, 3, 44, 46, 174–218
alignment, 192–198
auxiliary 775 nm lock, 44, 182, 198–205
background and leakage checks, 216–217

carrier rejection, viii, 46, 176–178, 180
cavity scans, 211–213
construction, 189
control system, 44, 198–205
finesse and ringdown, 209–210
four-cavity chain, viii, 44, 174–217
higher-order-mode modeling, 205–209, 211
length offsets, 183–184
locked noise spectra, 213–214
optical and mechanical design, 184–198
pass-band alignment, 46, 178–184
photon-counting integration, 174, 214–217
radius-of-curvature study, 184, 206–209
signal-sideband transmission, viii, 44, 175, 180

P

Pareto front, vii, 152–154, 157, 162
Piezoelectric actuators (PZTs), 188–205, 213, 243–260
Planck length, 21
Pound-Drever-Hall (PDH) locking, 36–39, 200–205, 224, 253–261
error signal, 36–39, 200, 253
modulation and demodulation, 36–39, 200, 224, 253
Power recycling, 40–41, 44, 219–223, 263–265
Power spectral density (PSD), 21–22, 48–66, 89–108, 110–114, 140–172, 205–214, 260–265

Q

Quantization noise, vii, 88–108
additive white-noise model, 89–92
arithmetic roundoff, vii, 89–94
coefficient quantization, 94, 104
input dependence, 103–105, 107
precision-subtraction measurement, 101, 103
Quantizer, 89
uniform quantization, 90
Quantum
Field theory, 17

gravity, 17–22
loop quantum gravity, 17, 21
string theory, 17, 21
Mechanics, 17, 18, 20
noise, 18–19, 63–65
radiation-pressure noise, 19, 64
shot noise, 19, 64
Spacetime fluctuations, 18, 20–22
spacetime fluctuations (QST), vii, 1, 3–4, 21–22, 42
experimental constraints, 22
geontropic model, 42
length-scaling exponent, 21, 42
models, 22

R
Relativity
General relativity, 7–11
Special relativity, 8
Resonant-mass detectors (Weber bars), 25
Ricci scalar, 10, 12
Ricci tensor, 10, 12
Riemann curvature tensor, 10, 12
Robust control, vii, 110, 160, 162, 165

S
Salecker-Wigner thought experiment, 18–21
SB08CD, 120
Scattered light, 65, 216
Schwarzschild radius, 20, 21
Schwarzschild solution, 10, 20
Second-harmonic generation (SHG), 198
Seismic noise, 4, 61–63
Sensor noise, 4, 110–120, 141–145, 198–205, 213, 253–261
Signal recycling, 41
Signal sidebands, 3, 41, 175
Signal-to-noise ratio (SNR), 46, 89, 101–105, 140
Spacetime, 7–12, 14, 15, 17, 18, 20, 21

Squeezed light, 29, 64
Stability, 68, 79–85
    margins, 83–85, 160, 257
        gain margin, 85, 160, 257
        phase margin, 83–85, 120–123, 160–167, 257
        unity-gain frequency, 83–84, 160–167, 257
    Nyquist criterion, 82
    Routh–Hurwitz criterion, 68, 80
Standard Quantum Limit (SQL), 18–19, 63
    distance measurement, 18–19
    imprecision-back-action tradeoff, 19, 63
State-space, 71–73, 99, 100, 111, 118, 147
    AB09ND balancing, 101, 117
    Schur form, 100, 115
    similarity transformations, 100
    system augmentation, 111, 118, 147
Superconducting nanowire single-photon detectors (SNSPDs), 48, 215
System norms
    $\mathcal{H}_\infty$ norm, 110, 161
    $\mathcal{H}_2$ norm, 110, 146

T
Technical noise, 65
Tensor, 8, 14
Thermal noise, 51–55
    acoustic resonances, 51–55
    Brownian noise, 51–54
    thermo-optic noise, 54–55
Transfer function, 70–85
    Bode plots, 77–79
    frequency response, 70–79
    laser filter cavity measurements, 257
    poles and zeros, 71–85, 101, 118–120, 133
    zero-pole-gain (ZPK) representation, 71, 119

V
Vibration isolation, 61–63, 188, 239

laser filter cavity, 232, 239–244