

LASER INTERFEROMETER GRAVITATIONAL WAVE OBSERVATORY  
- LIGO -  
CALIFORNIA INSTITUTE OF TECHNOLOGY  
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

|                                                  |                  |            |
|--------------------------------------------------|------------------|------------|
| Technical Note                                   | LIGO-T2600158-v6 | 2026/09/26 |
| <b>Validating LIGO Interferometer<br/>Models</b> |                  |            |
| Abhiram Sharma                                   |                  |            |

California Institute of Technology  
LIGO Project, MS 18-34  
Pasadena, CA 91125  
Phone (626) 395-2129  
Fax (626) 304-9834  
E-mail: [info@ligo.caltech.edu](mailto:info@ligo.caltech.edu)

Massachusetts Institute of Technology  
LIGO Project, Room NW22-295  
Cambridge, MA 02139  
Phone (617) 253-4824  
Fax (617) 253-7014  
E-mail: [info@ligo.mit.edu](mailto:info@ligo.mit.edu)

LIGO Hanford Observatory  
Route 10, Mile Marker 2  
Richland, WA 99352  
Phone (509) 372-8106  
Fax (509) 372-8137  
E-mail: [info@ligo.caltech.edu](mailto:info@ligo.caltech.edu)

LIGO Livingston Observatory  
19100 LIGO Lane  
Livingston, LA 70754  
Phone (225) 686-3100  
Fax (225) 686-7189  
E-mail: [info@ligo.caltech.edu](mailto:info@ligo.caltech.edu)

<http://www.ligo.caltech.edu/>

## Abstract

While previous studies have been done on the theoretical modeling of mirror thermal distortions within the Laser Interferometer Gravitational-Wave Observatory (LIGO), there is still a mismatch between simulation predictions and experimental measurements. This is believed to originate from the nonstationary evolution of thermal distortions in the optics due to incomplete knowledge of their initial thermal state and incomplete modeling of their thermal dynamics, which can introduce time-dependent systematic errors into the calibration and optical modes. This project aims to construct a more accurate model of the LIGO Livingston Observatory (LLO) during the O4 run by including more accurate thermal effects as well as using Markov chain Monte Carlo (MCMC) sampling as a diagnostic tool to predict discrepant model parameters. The predictions of this project's optical model can be validated against real LIGO data, specifically from stretches of data where the interferometer remains locked and the mirrors are allowed to reach a thermal steady-state. This approach allows for a significant reduction in the complexity of the parameter space compared to previous efforts, as we do not have to account for the transient dynamics of the interferometer immediately after it acquires lock. Data that satisfies these criteria is not trivial to obtain, as lock losses due to known and unknown causes are common. Ultimately, these optical simulations will improve our understanding of thermal effects in LIGO and may enable the development of more robust active thermal compensation techniques, mitigating the impact of thermal noise and allowing for even higher strain sensitivity needed for future gravitational wave detectors.

## 1 Introduction

Gravitational waves (GWs) are distortions in the fabric of space-time, first predicted by Albert Einstein in 1915 as a result of general relativity. These waves propagate from energetic physical processes (colliding black holes, supernovae, colliding neutron stars) at the speed of light, carrying information about the events that created them. GWs are difficult to measure due to the minimal spatial distortion (strain) that they impose on Earth's surface, on the order of  $10^{-21}$ . LIGO was the first experiment to observe a GW in 2015 [1].

To reach a strain sensitivity capable of observing gravitational waves, LIGO must overcome many sources of noise including but not limited to: seismic noise, optical vacuum fluctuations, cosmic rays, and thermal noise [2]. This project will focus on simulating how mirror thermal distortions in particular affect observed optical signals, in order to reproduce experimental data.

While previous studies have been done on the theoretical modeling of mirror thermal fluctuations [3, 4], there is still a mismatch between simulation predictions and experimental measurements. This is believed to originate from the nonstationary evolution of thermal distortions in the optics, due to incomplete knowledge of their initial thermal state and incomplete modeling of their thermal dynamics, which can introduce time-dependent systematic errors into the calibration and optical modes [5].

This project aims to construct accurate models of the interferometer that capture these thermal effects, and also utilize MCMC algorithms to analyze parameter sensitivity within the interferometer. The predictions of this project’s optical model can be validated against real LIGO data, specifically from stretches of data where the interferometer remains locked and the mirrors are allowed to reach a thermal steady-state. Data satisfying these criteria are not trivial to obtain, as whenever significant transient noise is introduced into the system, the interferometer loses lock. Ultimately, these optical simulations will improve our understanding of thermal effects in LIGO and may enable the development of more robust active thermal compensation techniques, mitigating the impact of thermal noise and allowing for even higher strain sensitivity needed for future gravitational wave detectors.

## 2 Background

### 2.1 LIGO Optical Setup

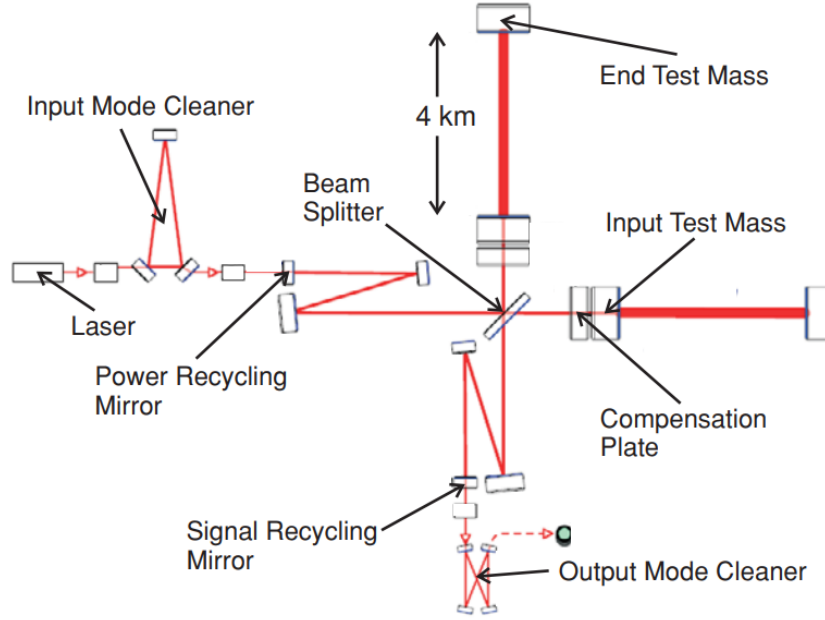


Figure 1: Diagram of LIGO’s dual recycled Michelson interferometer optical setup [6].

The LIGO detector is a Dual Recycled Michelson Interferometer (DRMI). Following the input laser path, the light travels through several key components [6]:

1. Input Mode Cleaner: Filters out higher-order laser spatial modes (see Sec 2.3).
2. Power Recycling Mirrors/Cavity: Forms a resonant cavity in the space before the beam splitter to amplify the injected laser power.
3. Beam Splitter: Divides and redirects the injected beam into the two Fabry-Perot arm cavities.

4. Output: After exiting the cavities, the beams recombine and are sent through a signal recycling mirror and an output mode cleaner before measurement by a photodetector.

When the interferometer arms are at the same length, it is tuned near a dark fringe, where the beams recombine destructively at the anti-symmetric port, minimizing the signal received by the photodetector. When a gravitational wave passes by the interferometer, the two Fabry-Perot cavities may be length contracted by different amounts causing a perturbed phase difference when the lasers recombine leading to an increase in output laser power.

## 2.2 Laser Modes

The spatial distribution of a laser beam can be decomposed into an orthonormal basis, most commonly either the Hermite-Gauss (HG) modes or Laguerre-Gauss (LG) modes. The derivation for these modes follows from solving the paraxial wave equation for an electromagnetic wave traveling in free space (Bond et al. [7]):

$$(\partial_x^2 + \partial_y^2)u - 2ik\partial_z u = 0 \quad (1)$$

One solution set to Equation (1) is the Hermite-Gauss modes, which are rectilinearly symmetric. In the equations below  $w(z)$  denotes the beam radius,  $q(z)$  denotes the complex beam parameter, and  $w_0$  is the minimum beam waist [7].

$$\text{HG}_{lm}(x, y, z) = u_l(x, z)u_m(y, z) \quad (2)$$

$$u_J(x, z) = \left( \frac{\sqrt{2/\pi}}{2^J J! w_0} \right)^{1/2} \left( \frac{q_0}{q(z)} \right)^{1/2} \left( \frac{-q^*(z)}{q(z)} \right)^{J/2} H_J \left( \frac{\sqrt{2}x}{w(z)} \right) e^{\frac{-ikx^2}{2q(z)}} \quad (3)$$

Another solution set to Equation 1 is the Laguerre-Gauss modes which are cylindrically symmetric.

$$u_{p,l}(r, \phi, z) = \left( \frac{2p!}{\pi(p+|l|)!w_0^2} \right)^{1/2} \left( \frac{q_0}{q(z)} \right) \left( \frac{-q^*(z)}{q(z)} \right)^{\frac{2p+|l|}{2}} \left( \frac{\sqrt{2}r}{w(z)} \right)^{|l|} L_p^{|l|} \left( \frac{2r^2}{w^2(z)} \right) e^{\frac{-ikr^2}{2q(z)}} e^{-il\phi} \quad (4)$$

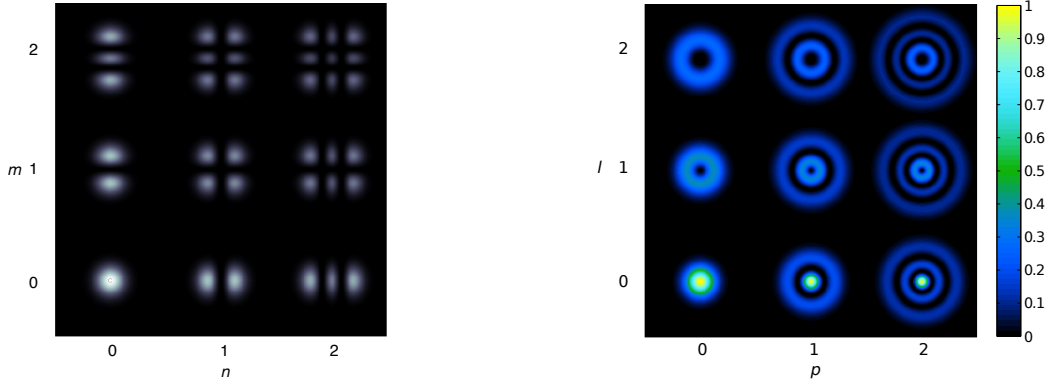


Figure 2: Normalized intensity distributions of Hermite-Gauss modes  $u_{nm}$  (left) and Laguerre-Gauss modes  $u_{pl}$  (right) [7].

The fundamental mode in both bases is a Gaussian intensity distribution which we can denote as the Transverse electromagnetic fundamental mode:  $\text{TEM}_{00}$ . This mode is relevant for the interferometer as we want the laser to remain in this fundamental mode throughout its optical path.

### 2.3 Optical Cavities

Optical cavities are combinations of reflecting optical components that allow for light to form standing waves. When the resonant wavelength of light is reflecting in an optical cavity, the amplitude/intensity will be increased, greatly increasing the laser power. These cavities are utilized throughout LIGO since the sensitivity of an interferometer-based GW detector is proportional to the laser power utilized [7].

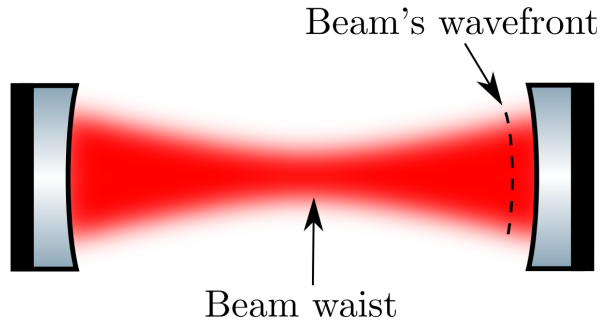


Figure 3: Diagram of optical cavity and corresponding spatial eigenmode [7].

Optical cavities also require the spatial wavefront of the light to match the geometry of the reflecting components to minimize loss in the gain. For example, in Figure 3 we can see that the beam wavefront must match the curvature of the mirror. Every optical cavity has an optical eigenmode, a combination of the basis modes described in Sec 2.2 that resonates in the cavity without loss.

One source of signal loss/noise in LIGO is mode-mismatch in an optical cavity. This is when the light in the cavity differs from the optical eigenmode. For example, thermal distortions in a cavity mirror’s coating can scatter a resonating eigenmode into higher-order, non-resonating modes.

## 2.4 Optical Thermal Modeling

The dominant thermal effects arise from two coupled physical mechanisms: thermo-elastic deformation and the thermo-optic effect. In thermo-elastic deformation, laser absorption in the mirror coating causes local heating which expands the substrate, physically deforming the mirror surface. In the thermo-optic effect, the same heating alters the local refractive index of the coating material, changing the effective optical path length for the reflected beam. Both effects produce a spatially varying phase shift across the beam profile, distorting the reflected wavefront away from the ideal spherical geometry required for cavity resonance [8].

The primary thermal distortion in the test mass can be approximated as a quadratic surface deformation, equivalent to a change in the mirror radius of curvature (RoC). This shifts the cavity eigenmode waist size and position, reducing the overlap integral between the circulating field and the ideal cavity mode. Beyond the quadratic approximation, higher-order surface deformations, arising from the non-uniform intensity profile of the laser beam and the finite aperture of the mirror substrate, produce more complex mode-coupling and are more difficult to model analytically. These higher-order effects are where the mismatch between current simulations and experimental data is most significant. It is difficult to quantify how these thermal effects manifest in experimental data because the initial thermal state is not well known. For example, whenever significant transient noise is introduced into the interferometer, the system may fall out of lock further complicating our understanding of the initial thermal state. Improved knowledge of the thermal initial state would allow simulations to make better predictions about the dynamic thermal state of the interferometer during its operation.

## 2.5 Thermal Compensation System (TCS)

To mitigate the effects of mirror thermal fluctuations, the Thermal Compensation System (TCS) was introduced. This instrument attaches to the mirrors and provides compensatory heating to the mirror surface to counteract thermal distortions caused by the laser. As the test masses absorb a fraction of the circulating laser power, they develop radial temperature gradients that cause thermo-elastic surface deformations in the Fabry-Perot arms and thermo-refractive substrate lenses in the recycling cavities [9]. To maintain optimal interferometer sensitivity, the TCS must constrain residual wavefront distortion to below 5.4 nm RMS, ensuring that round-trip scattering of the fundamental  $\text{TEM}_{00}$  mode remains strictly under 0.1% [9]. The system continuously monitors these thermal lenses using independent Hartmann Wavefront Sensors (HWS) [9]. To counteract the distortions, the TCS utilizes a combination of ring heater (RH) actuators to adjust the overall surface RoC, alongside  $\text{CO}_2$  laser projectors that apply spatially tunable heat profiles to the compensation plates to clean up higher-spatial-frequency residuals [9]. Incorporating the dynamics of these specific TCS sensors and actuators into our optical simulations will be helpful for accurately reproducing

the nonstationary thermal state of the interferometer.

## 2.6 Interferometer Lock

The interferometer only works at a single operating point, at which every cavity is resonant for the carrier light and the Michelson is held close to a dark fringe at the antisymmetric port. However, transient changes in the interferometer’s state, such as test mass pendulum motion, can move the interferometer from this point. The operating point must therefore be held by active feedback, which senses the deviation of each length from its set point and actuates on the optics to cancel it. An interferometer held in this state is said to be locked. Bringing it there from uncontrolled optics is lock acquisition, and dropping out of that state is lock loss.

Five degrees of freedom describe the length configuration [10, 11]: the differential and common combinations of the two arm lengths (DARM and CARM), the differential length of the short Michelson between the beamsplitter and the input test masses (MICH), and the lengths of the power- and signal-recycling cavities (PRCL and SRCL) [12].

Error signals for these degrees of freedom come from the Pound–Drever–Hall (PDH) technique and its extensions to coupled cavities [13, 14, 15, 7]. Locking is difficult because PDH signals are linear only very close to the operating point.

Models inherit the same constraint. In a steady-state frequency-domain simulation such as FINESSE [16, 17], the model adjusts the microscopic tuning of each degree of freedom until its modeled error signal is zero, reproducing the operating point that the real control system maintains. Any change to the optical parameters of the model moves that operating point, and the model has to be locked again before its outputs describe the detector. As in the real interferometer, the new operating point can be found only if the model starts close enough to it for the error signals to point the right way. This requirement shapes how the sampler evaluates the likelihood (Sec. 3.4).

## 2.7 Markov chain Monte Carlo (MCMC)

MCMC is a particularly useful technique for systems where the observable data exhibits a highly non-linear dependence on the underlying physical parameters. In the context of complex physical models, simulation-based Bayesian optimal experimental design relies heavily on Markov chain simulations to characterize these non-linear system responses [18, 19].

Within our FINESSE model, the relationship between the thermal state of the test masses and the resulting optical signal is inherently non-linear. As discussed in Section 2.3, thermal distortions in the mirror coatings alter the cavity geometry, leading to non-trivial mode-mismatch and scattering of the resonating optical eigenmode. The heating process itself is non-linear and knowledge of the thermal state is obscured by incomplete knowledge of the initial thermal state. By leveraging MCMC algorithms, we aim to capture this non-linear temperature dependence, isolating the specific thermal and design parameters that have the greatest impact on the signal variation observed in LIGO operational data.

Following conventions of Bayesian inference methods, we can introduce three spaces [18]:

- **The Design Space ( $\mathcal{D}$ ):** The set of all controllable experimental variables. A specific design, denoted as  $d \in \mathcal{D}$ , represents a unique configuration of the apparatus.
- **The Parameter Space ( $\Theta$ ):** The underlying physical states of nature or model parameters that the experiment seeks to infer. A specific parameter set, denoted as  $\theta \in \Theta$ , represents quantities the experimenter may not directly control but wishes to learn about.
- **Data Space ( $\mathcal{Y}$ ):**  $y \in \mathcal{Y}$  represents the observable measurements generated by the experiment.

The relationship between these spaces is governed by a forward model  $F(d, \theta)$ , which simulates the expected physical data  $y$ .

The goal of the experiment is to update our knowledge about the physical parameters  $\theta$  after observing data  $y$  collected under a specific design  $d$ . This update is governed by Bayes' theorem:

$$p(\theta|y, d) = \frac{p(y|\theta, d)p(\theta)}{p(y|d)} \quad (5)$$

Where  $p(\theta|y, d)$  is the posterior,  $p(y|\theta, d)$  is the likelihood, and  $p(\theta)$  is the prior.

## 3 Approach/Method

This work is hosted in the following GitLab repository:

<https://git.ligo.org/IF0sim/ligo-commissioning-modeling>

### 3.1 Finesse Model

The main optical model is built upon a base **finesse-ligo** LLO model with design parameters adjusted to match values from recent measurements. Additional components, such as coated apertures and baffles, were added to better match the physical interferometer. The main change to the model implemented during this project was the use of an analytical heating function to calculate mode scattering caused by a cylindrically symmetrical heating profile on the test masses. Prior to this change, the heating deformation was calculated with a slower, and more computationally expensive, Finite Element Analysis (FEA) method. We show in Figure 4 that this heating method, based on the paper by Hello and Vinet, is comparable to FEA methods at the laser power and ring heater power used in the thermal steady-state data [8]. This allowed for a significant reduction in the computational time for calculating heating deformations. Since the MCMC sampling continually varies ring heater power and laser powers over thousands of iterations, this speedup allows us to estimate parameters related to thermal effects in reasonable time.

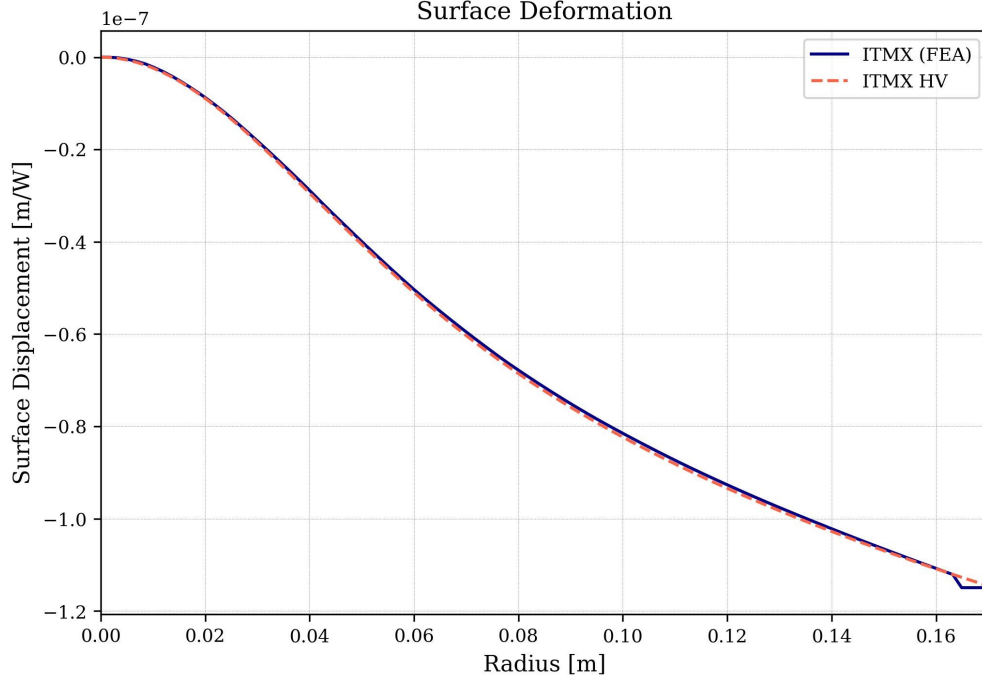


Figure 4: Comparison of Hello-Vinet heating method versus FEA heating method for ITMX surface deformations. The plot has units of  $m/W$  because this unit surface profile can be linearly rescaled based on the absorbed laser power. The clipping at the edges is due to finite coating diameter (or "flats") on test masses, which does not affect mode scattering.

### 3.2 Detector Data

We collected datasets that satisfy the thermal steady-state criteria with the LIGO Network Data Server protocol using the relevant methods in the GWpy library [20]. These datasets were then used for the MCMC model validation. However, for the purposes of validation, we can only make comparisons to columns which correspond to calibrated detectors. These calibration factors allow us to convert from ADC counts or channel voltage to physical values within the interferometer and generally have units of  $W/ct$ . Additionally, the uncertainty of these calibration factors plays a role in the sensitivity of our sampling. Only nine detectors satisfied our criteria for comparison purposes. This included realistic orders of magnitude, correct unit conversions, and correct observing period (O4). Collecting and choosing these calibrated factors took a great amount of work as there is not a centralized source for the most updated calibration factor per channel. The values were found across various aLOGs and are cited within the `L1_map.yaml` file in the repository. Future work on updating these values for as many detectors as possible would greatly aid detector characterization efforts.

| Segment (GPS Time)        | Interferometer Channel Name |                              |                              |     |
|---------------------------|-----------------------------|------------------------------|------------------------------|-----|
|                           | LSC-POP_A_<br>LF_OUTPUT     | IMC-IM4_TRANS_<br>SUM_OUT_DQ | LSC-POP_A_RF9_<br>Q_NORM_MON |     |
| [1422129869 - 1422217122) | 2509.52 ct                  | 56.66 ct                     | -41.17 ct                    | ... |
| [1422475262 - 1422635418) | 2511.95 ct                  | 56.69 ct                     | -41.88 ct                    | ... |
| [1422906429 - 1423042671) | 2509.22 ct                  | 56.62 ct                     | -39.29 ct                    | ... |
| [1423116358 - 1423230279) | 2505.82 ct                  | 56.62 ct                     | -42.36 ct                    | ... |
| [1424894576 - 1425055011) | 2474.96 ct                  | 56.45 ct                     | -38.55 ct                    | ... |
| [1425352967 - 1425459306) | 2478.77 ct                  | 56.44 ct                     | -44.21 ct                    | ... |
| [1425804561 - 1425894132) | 2488.43 ct                  | 56.38 ct                     | -36.32 ct                    | ... |
| [1434602740 - 1434726018) | 2211.49 ct                  | 56.52 ct                     | -10.40 ct                    | ... |
| [1435625437 - 1435775225) | 2207.02 ct                  | 56.46 ct                     | -4.64 ct                     | ... |
| [1436216024 - 1436374703) | 2205.43 ct                  | 56.47 ct                     | -4.64 ct                     | ... |
| [1438638690 - 1438728907) | 2209.58 ct                  | 56.43 ct                     | -6.01 ct                     | ... |
| [1439676408 - 1439782598) | 2208.66 ct                  | 56.52 ct                     | -6.95 ct                     | ... |
| [1441680250 - 1441767009) | 2210.70 ct                  | 56.46 ct                     | -7.18 ct                     | ... |
| [1441893597 - 1441985474) | 2212.81 ct                  | 56.53 ct                     | -6.93 ct                     | ... |
| [1444918046 - 1445009483) | 2207.75 ct                  | 56.60 ct                     | -11.30 ct                    | ... |

Table 1: Mean counts of select LLO channels during periods where interferometer was locked for more than 24 hours.

With a calibrated dataset in hand, we can begin to make comparisons between our thermal model and a given thermal steady-state period. Table 3 shows one preliminary comparison.

### 3.3 MCMC Sampling

To alleviate discrepancies between our FINESSE model and experimental data, we want to proceed by utilizing MCMC sampling to extract parameter sensitivity and identify discrepant parameters believed to be contributing to this error. Before utilizing the experimental detector data, we will study the accuracy/bias of these MCMC methods with a series of injection-recovery tests (Section 3.5). Nonetheless, we will begin by identifying how our FINESSE simulations fit into the Bayesian Inference framework explained in Section 2.7.

The data space,  $\mathcal{Y}$ , can be defined in one of two ways. For model validation,  $y \in \mathcal{Y}$  represents a set of power or gain measurements from calibrated LLO channels in a thermal steady-state period (i.e. a row of Table 1). In this case  $y$  is an explicit experimental measurement. For the injection recovery tests,  $y \in \mathcal{Y}$  is a set of simulated power detector measurements from a FINESSE model. In this case,  $y$  is a generated value. To have these cases be comparable, we must consider the fact that the experimental measurements are affected by a combination of sensing noise and calibration uncertainty, highlighted later in this section.

An element of the design space  $d \in \mathcal{D}$  represents controllable parameters in the interferometer and the entries are classical inputs into our FINESSE model. For example, the input laser power or ring heater power. Once we choose an appropriate experimental segment or synthetic injection, these variables are set once prior to sampling.

An element of the parameter space  $\theta \in \Theta$  represents uncontrollable parameters in the interferometer which can still be modified within our FINESSE model, and are continuously varied during the sampling stage. For example, the ITMX heating profile or PR2 mirror placement.

The forward model,  $F(d, \theta)$  is our FINESSE model which takes in an interferometer design vector and parameter vector to generate an output  $\bar{y}$  of simulated power measurements and gains. We then construct the log-likelihood from our forward model by assuming the experimental measurement noise on our power detectors follows a Gaussian distribution by the Central Limit Theorem and use a  $\chi^2$  statistic:

$$\ln p(y|d, \theta) = \ln \mathcal{L} = -\frac{1}{2} \sum_{i=1}^n \left( \frac{y_i - F(d, \theta)_i}{\sigma_i} \right)^2 = -\frac{1}{2} \sum_{i=1}^n \left( \frac{y_i - \bar{y}_i}{\sigma_i} \right)^2 \quad (6)$$

where  $\bar{y}_i$  is the  $i$ -th simulated detector output evaluated by the FINESSE model at proposed coordinate  $\theta$ , and  $\sigma_i$  is the standard deviation for that specific channel over the steady-state period added in quadrature with the calibration uncertainty. For injection-recovery, we choose  $\sigma_i$  to be on the same order as a specific steady-state period which we plan to compare to.

The prior,  $p(\theta)$ , can be constructed from our measurement uncertainty of a specific parameter. For example, if we consider PR2 mirror displacements along the optical path and this is known to 0.1 millimeters, we choose a truncated Gaussian prior with mean equal to the nominal placement and sigma equal to 0.1 mm. The probability of parameters being five sigma away or greater is set to 0. The choice of prior is critical as being able to confine our parameter space in this way greatly reduces the search space and allows for convergence in a tractable time period since likelihood evaluations can get computationally expensive.

The computational implementation of the sampling was carried out with the emcee python package [21, 22]. The package hosts a suite of affine-invariant sampling methods which require only the log-likelihood and log-prior. We chose to use an 80/20 split of the DE move and snooker move, since these shorten convergence times for highly correlated posteriors [24].

The convergence target for the sampling was per-parameter autocorrelation time. We used the  $\tau < N/50$  metric as recommended by the authors of the package [21].

### 3.4 Locking Sampler

The outputs of the model are meaningful only at a locked operating point, and the operating point itself depends on the parameters being sampled. Every likelihood evaluation must therefore re-lock the model at the proposed parameters  $\theta$  before any observable can be computed, and if  $\theta$  is too far from the state the lock starts from, no lock is found and the model makes no prediction at all. Some parameters alter the spatial mode structure, such as the radii of curvature of PR2 and PR3 or the ring-heater powers [7]. They therefore affect not only where the operating point lies but how easily the locks can reach it.

To address this, the FINESSE base model is built once, at the nominal parameter values, which are also the prior means: the thermal maps are applied, it is locked, and its arm loss

and DARM offset are fitted to the reference segment. Each likelihood evaluation starts from an independent copy of this locked model. It applies  $\theta$ , recomputes the thermal maps, and re-locks every enabled length loop from the nominal operating point. Once the locks have converged, the DARM offset is re-optimized so that the modeled carrier power at the output photodiodes matches its reference value. Finally, the observables are computed.

Starting every evaluation from the same reference state and maintaining the same computational locking method makes the likelihood a deterministic function of  $\theta$  alone. We could speed-up the sampler by starting each lock from the operating point of the emcee walker’s current position, which would shorten the lock for small steps, but it would make the outcome of the lock depend on chain history, so the sampler may no longer produce a well-defined posterior.

If the locking method does not converge within the max number of locking steps, we set  $\ln \mathcal{L} = -\infty$ ; the same is done when FINESSE reports that a cavity in the model has become geometrically unstable. For a successful lock, we assign  $\ln \mathcal{L}$  according to Equation 6. A proposal with zero posterior density is never accepted by the ensemble moves [21, 23, 24], so the walker keeps its current position.

We therefore choose hard prior bounds that lie inside a reachable locking space. The truncation of each prior is a compromise between two requirements: the support must be wide enough to include every physically plausible value and narrow enough that the model locks everywhere inside it. Because every evaluation re-locks from the nominal point, the bounds also cap the largest re-lock the sampler will ever request. If there are  $n$  parameters, this creates a bounding box with  $2n$  faces, that is a subset of  $\Theta$ . This bounding box is where the sampling occurs.

We test both requirements before each sampler runs by evaluating the likelihood with each parameter moved in turn to a point 5% of the prior width inside either end of its support, the others held at their reference values. Then we check if these points lock. Finally we decrease the max number of locking steps until we find a minimum where locking can be achieved at every edge of the prior box.

### 3.5 Sampler Analysis

Before applying the sampler to experimental data, we must establish that it recovers parameters correctly on data whose answer we already know. This procedure is an injection-recovery test. We draw a set of true parameters  $\theta^* \sim p(\theta)$  from the prior, evaluate the forward model at that point to obtain  $\bar{y} = F(d, \theta^*)$ , perturb those values with noise drawn at the scale  $\sigma_i$  described above to produce a synthetic data vector  $y$ , and then run the sampler on  $y$  while withholding  $\theta^*$ . Since the true value is known by construction, the posterior can be checked against it directly.

Two distinct properties are of interest. The first is constraining power: whether the detector set carries information about a given parameter. This is calculated as the ratio of the prior width to the posterior width. If the two are equal, the likelihood is flat in that direction and the sampler has returned the prior unchanged; the parameter is not identifiable from the available channels. If the posterior is substantially narrower, the data are informative. This

distinction matters in our case because only a subset of the FINESSE detectors correspond to LLO channels with asserted calibrations, which restricts  $\mathcal{Y}$  considerably relative to what the model can in principle output.

The second property is the pull:

$$z_j = \frac{\theta_j^* - \hat{\theta}_j}{\sigma_{\theta_j}} \quad (7)$$

where  $\hat{\theta}_j$  is the posterior median of the  $j$ -th parameter and  $\sigma_{\theta_j}$  is half the 16–84 percentile range. A single pull is not informative on its own, since a value of order unity is the expected outcome for a correctly calibrated posterior. Across an ensemble of injections, however, the pulls should be distributed as a standard normal: a nonzero mean indicates bias, and a standard deviation away from one indicates that the credible intervals are systematically too narrow or too wide.

The stronger statement is obtained from the quantile of the truth within its own posterior. For the  $j$ -th parameter of a given injection,

$$q_j = \frac{1}{N} \sum_{k=1}^N \mathbb{1} \left[ \theta_j^{(k)} < \theta_j^* \right], \quad (8)$$

where  $\theta_j^{(k)}$  is the  $k$ -th of the  $N$  posterior samples and  $\mathbb{1}[\cdot]$  equals one when its argument is true and zero otherwise.  $q_j$  is therefore the fraction of posterior samples lying below the injected value. If the truths are drawn from the prior, the data are generated by the same forward model used in the likelihood, and the sampler returns the correct posterior, then  $q_j$  is uniformly distributed on  $[0, 1]$  across injections [25, 26].

Across an ensemble of  $M$  injections, the P–P plot shows, for each credible level  $p \in [0, 1]$ , the fraction of injections whose quantile does not exceed it,

$$F_j(p) = \frac{1}{M} \sum_{m=1}^M \mathbb{1} \left[ q_j^{(m)} \leq p \right], \quad (9)$$

which is the empirical cumulative distribution of the quantiles. A calibrated sampler gives  $F_j(p) = p$ , tracing the diagonal to within the sampling uncertainty of the ensemble size. P–P plots of this kind are a standard validation method for gravitational-wave parameter estimation [27].

## 4 Results

### 4.1 Injection-Recovery

#### 4.1.1 Single Injection

We can start by studying the results of a single injection-recovery test with max TEM<sub>ij</sub> capped to order 8. The inputs to the model were matched with the inputs/design parameters within Row 1 of Table 1. For example, setting the initial laser power based on

IMC-IM4\_TRANS\_SUM\_OUT\_DQ. Then, after manually perturbing  $\theta$ , we obtain the posterior plot in Figure 5, the associated parameter values in Table 2, and the convergence measurement in Figure 6.

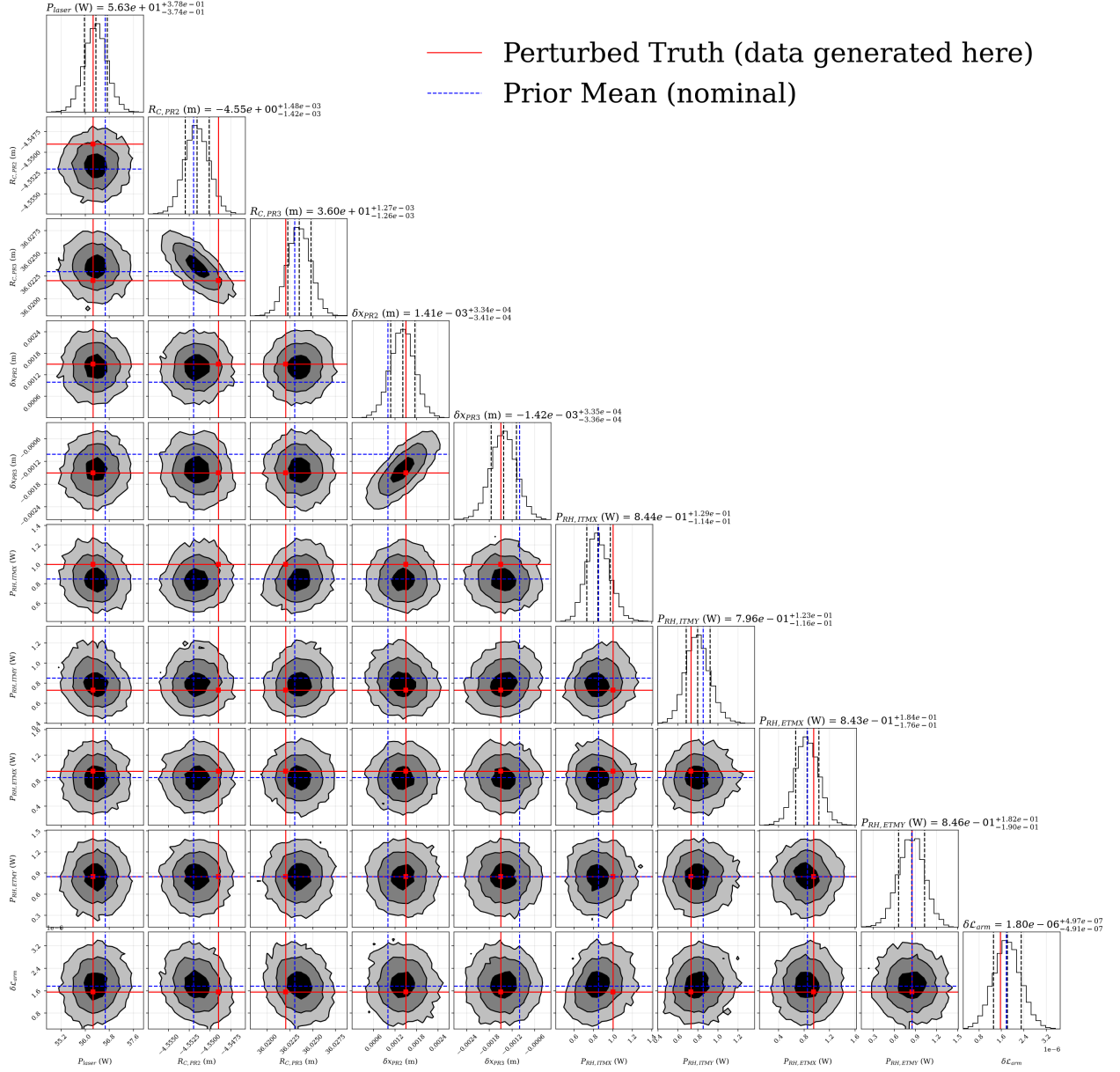


Figure 5: Posterior corner plot for 10 interferometer parameters generated from single injection-recovery test.

| Parameter                               | Prior                                                         | Posterior                     | Truth ( $\theta^*$ ) | Pull ( $z$ ) | $\frac{\sigma_{\text{prior}}}{\sigma_{\text{post}}}$ |
|-----------------------------------------|---------------------------------------------------------------|-------------------------------|----------------------|--------------|------------------------------------------------------|
| $P_{\text{laser}}$ (W)                  | $\mathcal{N}_{[\mu_0 \pm 5\sigma]}(\mu_0, 0.1)$               | $56.35^{+0.38}_{-0.37}$       | 56.26                | -0.23        | 0.27                                                 |
| $R_{C,PR2}$ (m)                         | $\mathcal{N}_{[\mu_0 \pm 5\sigma]}(\mu_0, 2 \cdot 10^{-3})$   | $-4.5516^{+0.0015}_{-0.0014}$ | -4.5490              | 1.79         | 1.38                                                 |
| $R_{C,PR3}$ (m)                         | $\mathcal{N}_{[\mu_0 \pm 5\sigma]}(\mu_0, 2 \cdot 10^{-3})$   | $36.0235^{+0.0013}_{-0.0013}$ | 36.0220              | -1.18        | 1.58                                                 |
| $\delta x_{PR2}$ (mm)                   | $\mathcal{N}_{[\mu_0 \pm 5\sigma]}(\mu_0, 0.5)$               | $1.41^{+0.33}_{-0.34}$        | 1.50                 | 0.26         | 1.48                                                 |
| $\delta x_{PR3}$ (mm)                   | $\mathcal{N}_{[\mu_0 \pm 5\sigma]}(\mu_0, 0.5)$               | $-1.42^{+0.34}_{-0.34}$       | -1.50                | -0.23        | 1.49                                                 |
| $P_{RH,ITMX}$ (W)                       | $\mathcal{N}_{[0,2]}(\mu_0, 0.05)$                            | $0.84^{+0.13}_{-0.11}$        | 1.00                 | 1.28         | 0.41                                                 |
| $P_{RH,ITMY}$ (W)                       | $\mathcal{N}_{[0,2]}(\mu_0, 0.05)$                            | $0.80^{+0.12}_{-0.12}$        | 0.73                 | -0.55        | 0.42                                                 |
| $P_{RH,ETMX}$ (W)                       | $\mathcal{N}_{[0,2]}(\mu_0, 0.05)$                            | $0.84^{+0.18}_{-0.18}$        | 0.95                 | 0.59         | 0.28                                                 |
| $P_{RH,ETMY}$ (W)                       | $\mathcal{N}_{[0,2]}(\mu_0, 0.05)$                            | $0.85^{+0.18}_{-0.19}$        | 0.85                 | 0.02         | 0.27                                                 |
| $\delta \mathcal{L}_{\text{arm}}$ (ppm) | $\mathcal{N}_{[\mu_0/7.5, 7.5\mu_0]}(\mu_0, \frac{\mu_0}{3})$ | $1.80^{+0.50}_{-0.49}$        | 1.56                 | -0.49        | 1.05                                                 |

Table 2: Injection-recovery results for a single sample. The pull is calculated with Eq. 7.  $\mathcal{N}_{[a,b]}$  is a Gaussian truncated to  $[a, b]$ .

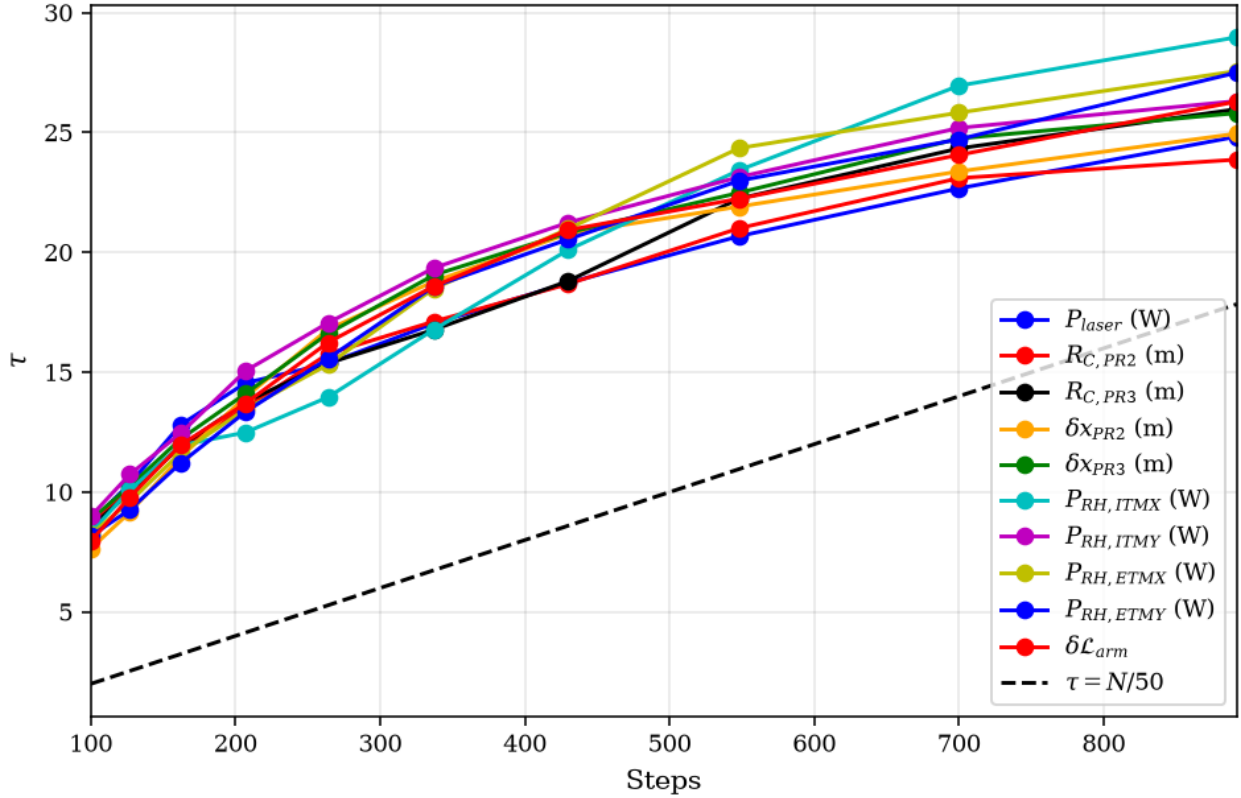


Figure 6: Plot of per parameter autocorrelation time  $\tau$ . To be confident that samples are independent from initial conditions, we ideally run the sampler until  $\tau < N_{\text{steps}}/50$  for every parameter. Due to computational constraints, we terminate sampler before this point for this test.

The first observation from this test is the sampler performance when recovering highly correlated parameters. For example consider the joint distribution between the RoC's of the PRC mirrors ( $R_{C,PR2}$ ,  $R_{C,PR3}$ ). We expect high correlation between adjacent mirrors, for example to maintain optimal mode matching in the cavity. The true mirror curvature was perturbed in a direction parallel to the parameter's correlation direction. This hinders the sampler's ability to recover the perturbation since the change in the  $\chi^2$  metric will be much weaker compared to if the perturbation was moved in an perpendicular direction. This perpendicular case can be seen in the joint distribution between the displacement parameters of the same mirrors ( $\delta x_{PR2}$ ,  $\delta x_{PR3}$ ). We can see that the posterior median settles on the perturbed truth for both mirror placements. This is quantitatively shown in the pull values of the 4 parameters in Table 2, where  $|z_{R_{C,PR2}}| > |z_{R_{C,PR3}}| > 1$  and  $|z_{\delta x_{PR3}}| < |z_{\delta x_{PR2}}| < 1$ .

The second observation is the high constraining power of the mirror parameters compared to ring heater powers. To address this for the experimental model validation, the max  $TEM_{ij}$  order was increased from 8 to 14 to potentially increase sensitivity to mode scattering/resonance.

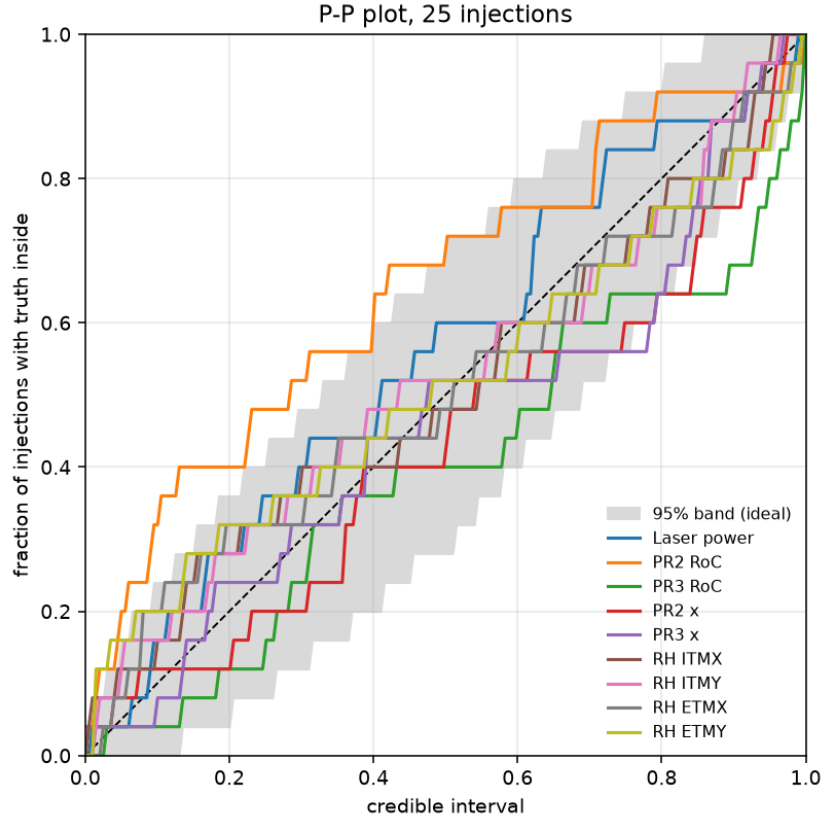


Figure 7: Probability–probability plot for 25 injection–recovery tests. Each curve is the empirical cumulative distribution of the quantile of the injected value within its own posterior; a calibrated sampler follows the diagonal. The grey band is the pointwise 95% region for an ideal sampler at this ensemble size. Most parameters lie within the band, while some sit above the diagonal at small credible intervals and below it at large ones, consistent with overly narrow posteriors. The PR2 and PR3 radii of curvature deviate in opposite directions, due to their degeneracy.

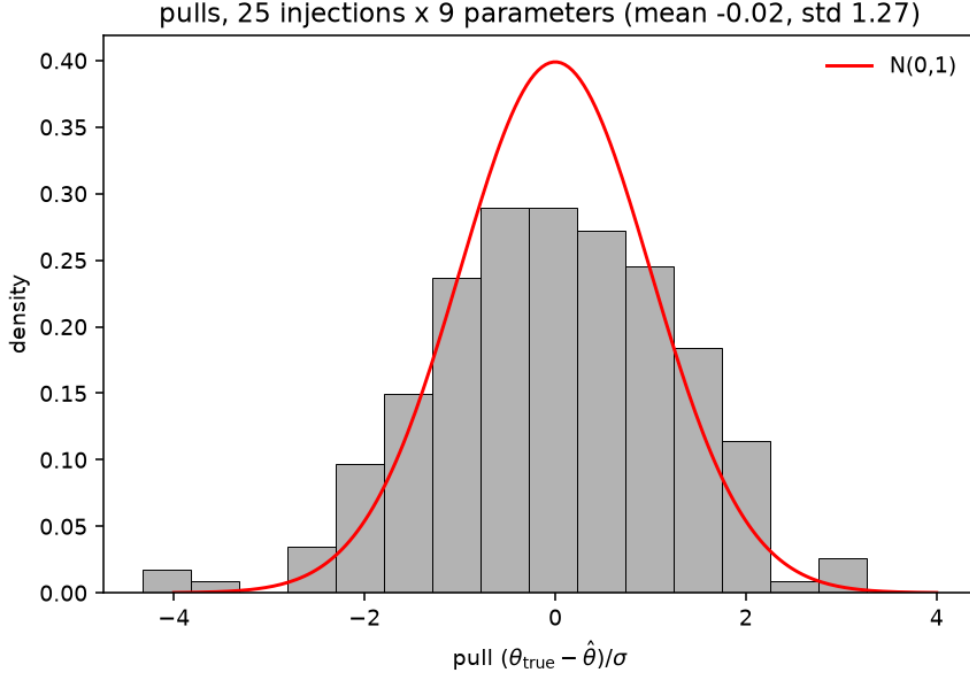


Figure 8: Distribution of pulls  $z = (\theta^* - \hat{\theta})/\sigma_\theta$  for 25 injection–recovery tests across nine sampled parameters (225 values), compared with the standard normal expected for a calibrated sampler. The distribution is centered at zero (mean  $-0.02$ ) but wider than  $\mathcal{N}(0, 1)$  (standard deviation 1.27), indicating that the posteriors are unbiased but their credible intervals are too narrow by approximately 25%. Pulls within a single injection are correlated through parameter degeneracies, so the values are not independent.

#### 4.1.2 Ensemble Injection

We used a FINESSE model with max  $\text{TEM}_{ij}$  capped to order 4 in order to save computation time. We also only ran each chain for a maximum of 400 steps. Design parameters for the model were matched using Row 1 of Table 1. We repeatedly draw samples from the multidimensional prior to perturb  $\theta$  and recover it to produce Figures 7 and 8.

The injection–recovery ensemble indicates that the sampler recovers the injected parameters without significant bias. The mean pull across all parameters is  $-0.02$ , and seven of the nine parameters have mean pulls within  $\pm 0.25$ , consistent with zero given the standard error of approximately 0.2 for this ensemble size. The exception is the pair of PR2 and PR3 radii of curvature, whose mean pulls of  $-0.55$  and  $+0.51$  are equal in magnitude and opposite in sign. This pattern is expected from their negative correlations as the detectors constrain a combination of the two curvatures more tightly than either individually, and the posterior distributes that combination between them differently in each injection.

The credible intervals, however, are systematically too narrow. The pull standard deviation exceeds 1 for every parameter, ranging from 1.06 to 1.41 with an average value of 1.27, and the fraction of pulls beyond  $2\sigma$  is between 8 and 20% against an expected 5%. The most likely cause is insufficient chain length. At an autocorrelation time of  $\tau \approx 28$  steps and 400 steps per chain, each injection yields roughly 550 effective samples, and the 16–84 percentile

interval estimated from so few samples does not resolve the tails of the posterior.

Approximately 60% of the attempted injections failed because the drawn parameters lay in regions where the interferometer model could not be locked. The ensemble is therefore drawn from the prior restricted to the lockable region, and the calibration statement applies to that region rather than to the full prior.

## 4.2 Model Validation

From Table 3, we can see that there are discrepancies in the model for various detectors. For example, the power recycling gain detectors for the 9 MHz sideband (PRG9) and 45 MHz sideband (PRG45) are significantly higher than the value from LLO, by factors of 2.8 and 1.6 respectively. The carrier power reflected from the PRM (a\_c0\_refl\_PRM) is low by a factor of 2.2, and the AS\_C power is 18% high. By contrast, Pin, PRG and a\_c0\_dcpd agree to within 1%. This agreement is not a prediction of the model since the arm loss is tuned to reproduce the measured PRG and the DARM offset is tuned to reproduce the measured DCPD power during the model build. Similarly, Pin enters as a direct input. The remaining channels are the ones that test the model.

Replacing the parameters with their posterior medians (Table 4, Table 5, Figure 9) reduces  $\chi^2$  from 2294 to 185. The largest improvements are in a\_c0\_refl\_PRM, which moves from a factor of 2.2 low to within 4%, and Pas\_c, which moves from 18% high to within 1%. PRG45 improves from 58% to 34% high. The improvement in the reflected and AS powers is obtained at the expense of the PRG, which moves from exact agreement to 10% high. This reflects a tension in the data. The observed arm power is not measured directly but derived as  $\frac{1}{2} P_{\text{in}} \text{PRG} \cdot G_{\text{arm}}$  with a fixed arm gain  $G_{\text{arm}} = 265$ , whereas the model's own arm gain,  $2P_{\text{arm}}/(P_{\text{in}} \text{PRG})$ , is approximately 254 at the nominal parameters and 246 at the posterior median. A model with this arm gain cannot match PRG and the derived arm power simultaneously, and the fit distributes the residual between them. PRG9 is essentially unchanged (112 to 109 against 39.7), indicating that none of the sampled parameters can reach it. The 9 MHz discrepancy might therefore originate from parameters outside of what was considered.

The remaining  $\chi^2 = 185$  is still large compared with the number of observables, so the model does not reproduce the data within the assumed uncertainties even at the posterior medians. The posterior widths should therefore be read as the statistical constraint conditional on the model being correct, not as the full uncertainty on the physical parameters. In addition, the injection-recovery tests (Section 3.5) indicate that the credible intervals at this chain length are approximately 25% too narrow.

### 4.2.1 Tension in the PRC Radii of Curvature

The largest parameter tensions from the nominal model occur in the radii of curvature of PR2 and PR3. Their posterior medians lie 7.5 and 9.8 posterior widths from the nominal values, compared with pulls of at most 2.6 for every other parameter. Measured using prior standard deviations instead, the shifts are 5.2 and 5.7, or  $-10.3$  mm and  $-11.3$  mm.

| Unit | Detector      | Model             | LLO               | LLO Channel                                  | Calibration ( $\frac{\text{Unit}}{\text{ct}}$ ) |
|------|---------------|-------------------|-------------------|----------------------------------------------|-------------------------------------------------|
| W    | Pin           | 53.3              | 53.4              | IMC-IM4_TRANS_SUM_OUTPUT                     | 1                                               |
|      | Parm          | $2.76 \cdot 10^5$ | $2.88 \cdot 10^5$ | $(1/2)*\text{Pin}*\text{PRG}*\text{ArmGain}$ | 0.92                                            |
|      | Pas_c         | 0.662             | 0.56              | ASC-AS_C_SUM_OUTPUT                          | 0.000125                                        |
|      | Pprc          | $2.19 \cdot 10^3$ | $2.31 \cdot 10^3$ | LSC-POP_A_LF_OUTPUT                          | 0.92                                            |
| mW   | a_c0_refl_PRM | 1.1               | 2.39              | LSC-REFL_A_LF_OUTPUT                         | 0.00025                                         |
|      | a_c0_dcpd     | 41                | 41                | OMC-DCPD_SUM_OUTPUT                          | 0.82                                            |
| a.u. | PRG           | 40.7              | 40.8              | LSC-PRC_GAIN_MON                             | 0.92                                            |
|      | PRG9          | 112               | 39.7              | LSC-POPAIR_B_RF18_I_MON                      | 0.019                                           |
|      | PRG45         | 23.1              | 14.6              | LSC-POPAIR_B_RF90_I_MON                      | 0.027                                           |

Table 3: Comparison of FINESSE model and LLO outputs prior to replacing parameter values with posterior medians. GPS time segment: [1422475262 ... 1422635418). Every FINESSE PD in the Detector column is matched with its corresponding physical detector. After multiplying the detector counts by the calibration factor in the Calibration column, we can compare Model outputs to LLO outputs in physical units.

Additionally, the sampler accepts these shifts despite their computed cost. The prior penalty is  $\frac{1}{2}(5.2^2 + 5.7^2) \approx 29$  in the log-posterior, equivalent to  $\Delta\chi^2 \approx 58$ . This is only favorable if the likelihood improves by more than that amount, and the total  $\chi^2$  was computed to change from 2294 at the nominal parameter set to 185 at the posterior median parameter set. The recycling-cavity curvatures are therefore the parameters the data move most strongly, and they are the most plausible source of the improvement in Table 4.

A physical interpretation for this is that PR2 and PR3 form the mode structure of the power-recycling cavity. Their curvatures determine how well the input beam couples into the fundamental mode that resonates in the arms. The detectors that improve most after the fit are the ones that sample this cavity directly. The reflected carrier power a\_c0\_refl\_PRM changes from a factor of 2.2 low to within 4%. PRG45 improves from 58% to 34% high, and the pick-off power Pprc moves from 5% low to 4% high.

The posterior also shows that the data do not constrain the two curvatures independently. PR2 and PR3 are strongly anti-correlated (Figure 9), so the measurements fix a combination of the two. Further analysis is needed to resolve this discrepancy regarding the PRC RoCs.

| Unit | Detector      | Model             | LLO               | LLO Channel              | Calibration ( $\frac{\text{Unit}}{\text{ct}}$ ) |
|------|---------------|-------------------|-------------------|--------------------------|-------------------------------------------------|
| W    | Pin           | 53.3              | 53.4              | IMC-IM4_TRANS_SUM_OUTPUT | 1                                               |
|      | Parm          | $2.94 \cdot 10^5$ | $2.88 \cdot 10^5$ | (1/2)*Pin*PRG*ArmGain    | 0.92                                            |
|      | Pas_c         | 0.563             | 0.56              | ASC-AS_C_SUM_OUTPUT      | 0.000125                                        |
|      | Pprc          | $2.4 \cdot 10^3$  | $2.31 \cdot 10^3$ | LSC-POP_A_LF_OUTPUT      | 0.92                                            |
| mW   | a_c0_refl_PRM | 2.29              | 2.39              | LSC-REFL_A_LF_OUTPUT     | 0.00025                                         |
|      | a_c0_dcpd     | 41                | 41                | OMC-DCPD_SUM_OUTPUT      | 0.82                                            |
| a.u. | PRG           | 44.9              | 40.8              | LSC-PRC_GAIN_MON         | 0.92                                            |
|      | PRG9          | 109               | 39.7              | LSC-POPAIR_B_RF18_I_MON  | 0.019                                           |
|      | PRG45         | 19.5              | 14.6              | LSC-POPAIR_B_RF90_I_MON  | 0.027                                           |

Table 4: Comparison of FINESSE model and LLO outputs after replacing parameter values with posterior medians.

| Parameter                               | Prior                                                         | Posterior                     | Nominal ( $\mu_0$ ) | Pull ( $z$ ) | $\frac{\sigma_{\text{prior}}}{\sigma_{\text{post}}}$ |
|-----------------------------------------|---------------------------------------------------------------|-------------------------------|---------------------|--------------|------------------------------------------------------|
| $P_{\text{laser}}$ (W)                  | $\mathcal{N}_{[\mu_0 \pm 5\sigma]}(\mu_0, 0.5)$               | $53.320^{+0.087}_{-0.086}$    | 53.359              | 0.45         | 5.77                                                 |
| $R_{C,\text{PR2}}$ (m)                  | $\mathcal{N}_{[\mu_0 \pm 20\sigma]}(\mu_0, 2 \cdot 10^{-3})$  | $-4.5623^{+0.0014}_{-0.0014}$ | -4.5520             | 7.49         | 1.45                                                 |
| $R_{C,\text{PR3}}$ (m)                  | $\mathcal{N}_{[\mu_0 \pm 20\sigma]}(\mu_0, 2 \cdot 10^{-3})$  | $36.0117^{+0.0012}_{-0.0011}$ | 36.0230             | 9.75         | 1.73                                                 |
| $\delta x_{\text{PR2}}$ (mm)            | $\mathcal{N}_{[\mu_0 \pm 5\sigma]}(\mu_0, 0.5)$               | $0.26^{+0.32}_{-0.32}$        | 1.00                | 2.33         | 1.57                                                 |
| $\delta x_{\text{PR3}}$ (mm)            | $\mathcal{N}_{[\mu_0 \pm 5\sigma]}(\mu_0, 0.5)$               | $-0.27^{+0.32}_{-0.32}$       | -1.00               | -2.30        | 1.57                                                 |
| $P_{\text{RH,ITMX}}$ (W)                | $\mathcal{N}_{[0, 2]}(\mu_0, 0.2)$                            | $0.687^{+0.042}_{-0.042}$     | 0.796               | 2.61         | 4.78                                                 |
| $P_{\text{RH,ITMY}}$ (W)                | $\mathcal{N}_{[0, 2]}(\mu_0, 0.2)$                            | $0.532^{+0.040}_{-0.039}$     | 0.609               | 1.92         | 5.03                                                 |
| $P_{\text{RH,ETMX}}$ (W)                | $\mathcal{N}_{[0, 2]}(\mu_0, 0.2)$                            | $0.595^{+0.045}_{-0.045}$     | 0.583               | -0.27        | 4.45                                                 |
| $P_{\text{RH,ETMY}}$ (W)                | $\mathcal{N}_{[0, 2]}(\mu_0, 0.2)$                            | $0.143^{+0.044}_{-0.044}$     | 0.127               | -0.36        | 4.55                                                 |
| $\delta \mathcal{L}_{\text{arm}}$ (ppm) | $\mathcal{N}_{[\mu_0/7.5, 7.5\mu_0]}(\mu_0, \frac{\mu_0}{3})$ | $1.97^{+0.49}_{-0.48}$        | 1.63                | -0.70        | 1.13                                                 |

Table 5: Posterior estimates from LLO thermal steady-state data, segment [1422475262 ... 1422635418). The injection-recovery tests (Section 3.5) indicate that the posterior widths at this chain length are approximately 25% too narrow, so the widths may have to be interpreted with a dilation by a factor of about 1.27. The pull is computed with Eq. 7 using the nominal value  $\mu_0$  in place of  $\theta^*$ . Note that the truncation of the  $R_{C,\text{PR2}}$  and  $R_{C,\text{PR3}}$  priors had to be extended from  $\pm 5\sigma$  to  $\pm 20\sigma$  in order to prevent the sampler from continuously colliding with the boundary of the prior.

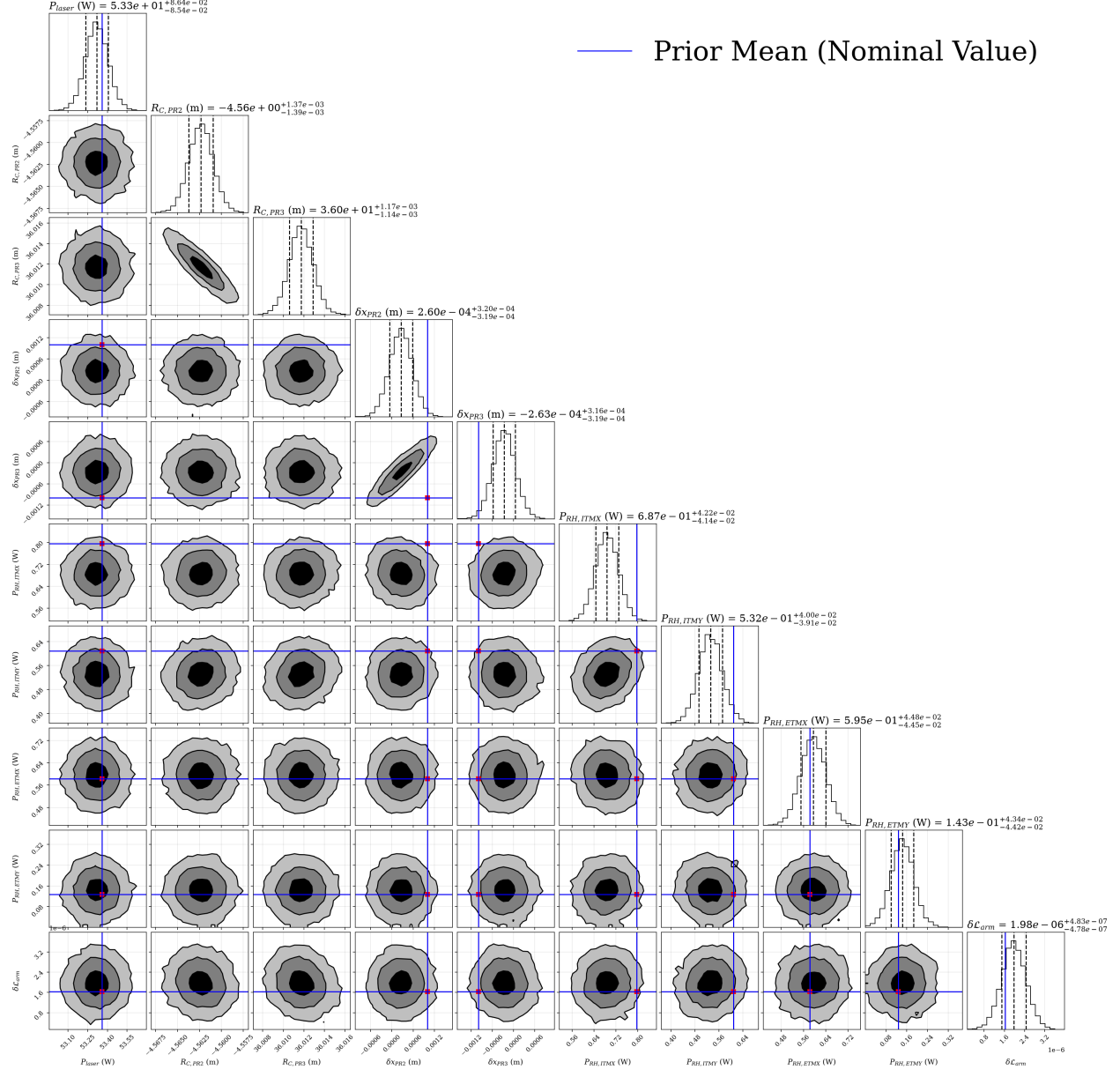


Figure 9: Posterior corner plot for 10 interferometer parameters matched to GPS time segment: [1422475262 ... 1422635418].

## 5 Conclusions

This project built a thermal model of LLO in FINESSE and paired it with an MCMC sampler to diagnose where the model and the interferometer disagree. The model replaces finite-element heating calculations with the analytic Hello-Vinet solution for self-heating and ring-heater deformation, which agrees with the finite-element result at the powers of the thermal steady-state data. It is fast enough that the thermal state can be recomputed at every likelihood evaluation. The sampler re-locks an independent copy of a single reference model at every proposal, which keeps the likelihood a deterministic function of the parameters.

The prior bounds are chosen to lie inside the region where the model can be locked.

The injection–recovery tests show that the sampler recovers injected parameters without significant bias: the mean pull across 25 injections and nine parameters is  $-0.02$ . The credible intervals, however, are systematically too narrow, with a pull standard deviation of 1.27. We attribute this provisionally to the short chains used for the ensemble, which yield of order 550 effective samples per injection. The tests also show that the PR2 and PR3 radii of curvature are strongly degenerate: the detectors constrain a combination of the two much more tightly than either individually.

Applied to LLO data from a thermal steady-state segment, the sampler reduces  $\chi^2$  from 2294 at the nominal parameters to 185 at the posterior medians. The largest improvements are in the carrier power reflected from the PRM and the AS\_C power, which move from discrepancies of a factor of 2.2 and 18% to within 4% and 1% respectively. The change responsible is primarily a shift of approximately 10 mm in each of the PR2 and PR3 radii of curvature, about five prior standard deviations. This is consistent with the recycling-cavity geometry absorbing a mode-matching error, since the channels that improve are those that sample that cavity directly. The ITM ring heater powers are also constrained well by the data, to about one fifth of their prior width, and settle approximately 13% below their values read from the dataset.

The fit also shows where the current model is incomplete. The remaining  $\chi^2 = 185$  is large compared with the number of observables, so the model cannot reproduce the data within the assumed uncertainties. The 9 MHz recycling gain is unchanged by the fit (109 against a measured 39.7), so its discrepancy lies outside the sampled parameters. The PRG and the derived arm power cannot be matched simultaneously, because the model’s arm gain (246–254) differs from the fixed value of 265 used to derive the measured arm power.

These results suggest several next steps:

- Rerun the injection ensemble with longer chains, and at the mode order used for validation, to confirm whether the narrow credible intervals are due to short chain length.
- Sample the input beam waist and position alongside the PR2 and PR3 curvatures by using measurements of beam spot size on the test masses as observables. If the curvature shifts shrink when the input mode is free, the tension may originate in the assumed model input beam rather than the mirrors. The inferred shifts could also be compared with updated measurement uncertainty on the PR2 and PR3 radii.
- Add observables that respond more directly to the recycling-cavity mode and extend the set of calibrated channels. A centralized record of current calibration factors would make this considerably easier.
- Repeat the analysis across the other thermal steady-state segments. Parameters that are physical should be stable between segments, while those compensating for model deficiencies may not be. The drop in POP power between the early and late segments in Table 1 indicates a change in the interferometer state that such a comparison should capture.

## References

- [1] B. P. Abbott et al. (LIGO Scientific Collaboration and Virgo Collaboration), *Observation of gravitational waves from a binary black hole merger*. Phys. Rev. Lett. 116, 061102 (2016).
- [2] R. X. Adhikari, *Gravitational radiation detection with laser interferometry*. Rev. Mod. Phys. 86, 121 (2014).
- [3] V. B. Braginsky and S. P. Vyatchanin, *Thermodynamical fluctuations in optical mirror coatings*. Phys. Lett. A 312, 244–255 (2003).
- [4] G. M. Harry et al., *Thermal noise from optical coatings in gravitational wave detectors*. Appl. Opt. 45, 1569–1574 (2006).
- [5] L. Sun et al., *Characterization of systematic error in Advanced LIGO calibration*. Class. Quantum Grav. 37, 225008 (2020).
- [6] G. M. Harry (for the LIGO Scientific Collaboration), *Advanced LIGO: The next generation of gravitational wave detectors*. Class. Quantum Grav. 27, 084006 (2010).
- [7] C. Bond, D. Brown, A. Freise, and K. A. Strain, *Interferometer techniques for gravitational-wave detection*. Living Rev. Relativ. 19, 3 (2016).
- [8] P. Hello and J.-Y. Vinet, *Analytical models of thermal aberrations in massive mirrors heated by high power laser beams*. J. Phys. (France) 51, 1267–1282 (1990).
- [9] A. F. Brooks et al., *Overview of Advanced LIGO adaptive optics*. Appl. Opt. 55, 8256–8265 (2016).
- [10] P. Fritschel et al., *Readout and control of a power-recycled interferometric gravitational-wave antenna*. Appl. Opt. 40, 4988–4998 (2001).
- [11] K. Izumi and D. Sigg, *Advanced LIGO: Length sensing and control in a dual recycled interferometric gravitational wave antenna*. Class. Quantum Grav. 34, 015001 (2017).
- [12] J. Aasi et al. (LIGO Scientific Collaboration), *Advanced LIGO*. Class. Quantum Grav. 32, 074001 (2015).
- [13] R. V. Pound, *Electronic frequency stabilization of microwave oscillators*. Rev. Sci. Instrum. 17, 490–505 (1946).
- [14] R. W. P. Drever et al., *Laser phase and frequency stabilization using an optical resonator*. Appl. Phys. B 31, 97–105 (1983).
- [15] E. D. Black, *An introduction to Pound–Drever–Hall laser frequency stabilization*. Am. J. Phys. 69, 79–87 (2001).
- [16] A. Freise et al., *Frequency-domain interferometer simulation with higher-order spatial modes*. Class. Quantum Grav. 21, S1067–S1074 (2004).

- [17] D. Brown et al., *FINESSE 3: Frequency-domain interferometer simulation software* (2025). <https://doi.org/10.5281/zenodo.12662017>, <https://finesse.ifosim.org>.
- [18] X. Huan and Y. M. Marzouk, *Simulation-based optimal Bayesian experimental design for nonlinear systems*. J. Comput. Phys. 232, 288–317 (2013).
- [19] P. Müller, B. Sansó, and M. De Iorio, *Optimal Bayesian design by inhomogeneous Markov chain simulation*. J. Am. Stat. Assoc. 99, 788–798 (2004).
- [20] D. M. Macleod, J. S. Areeda, S. B. Coughlin, T. J. Massinger, and A. L. Urban, *GWpy: A Python package for gravitational-wave astrophysics*. SoftwareX 13, 100657 (2021).
- [21] D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman, *emcee: The MCMC hammer*. Publ. Astron. Soc. Pac. 125, 306–312 (2013).
- [22] J. Goodman and J. Weare, *Ensemble samplers with affine invariance*. Commun. Appl. Math. Comput. Sci. 5, 65–80 (2010).
- [23] C. J. F. ter Braak, *A Markov chain Monte Carlo version of the genetic algorithm Differential Evolution: Easy Bayesian computing for real parameter spaces*. Stat. Comput. 16, 239–249 (2006).
- [24] C. J. F. ter Braak and J. A. Vrugt, *Differential Evolution Markov chain with snooker updater and fewer chains*. Stat. Comput. 18, 435–446 (2008).
- [25] S. R. Cook, A. Gelman, and D. B. Rubin, *Validation of software for Bayesian models using posterior quantiles*. J. Comput. Graph. Stat. 15, 675–692 (2006).
- [26] S. Talts, M. Betancourt, D. Simpson, A. Vehtari, and A. Gelman, *Validating Bayesian inference algorithms with simulation-based calibration*. arXiv:1804.06788 (2018).
- [27] J. Veitch et al., *Parameter estimation for compact binaries with ground-based gravitational-wave observations using the LALInference software library*. Phys. Rev. D 91, 042003 (2015).