

ASTROPHYSICS

Improving cosmological reach of a gravitational wave observatory using Deep Loop Shaping

Jonas Buchli^{1*†}, Brendan Tracey^{1†}, Tomislav Andric^{2,3†}, Christopher Wipf^{4†}, Yu Him Justin Chiu^{1†}, Matthias Lochbrunner^{1†}, Craig Donner^{1†}, Rana X. Adhikari^{4*}, Jan Harms^{2,3*}, Iain Barr¹, Roland Hafner¹, Andrea Huber¹, Abbas Abdolmaleki¹, Charlie Beattie¹, Joseph Betzwieser⁴, Serkan Cabi¹, Jonas Degraeve¹, Yuzhu Dong¹, Leslie Fritz¹, Anchal Gupta⁴, Oliver Groth¹, Sandy Huang¹, Tamara Norman¹, Hannah Openshaw¹, Jameson Rollins⁴, Greg Thornton¹, George van den Driessche¹, Markus Wulfmeier¹, Pushmeet Kohli^{1*}, Martin Riedmiller¹, The LIGO Instrument Team[‡]

Improved low-frequency sensitivity of gravitational wave observatories would unlock study of intermediate-mass black hole mergers and binary black hole eccentricity and provide early warnings for multimessenger observations of binary neutron star mergers. Today's mirror stabilization control injects harmful noise, constituting a major obstacle to sensitivity improvements. We eliminated this noise through Deep Loop Shaping, a reinforcement learning method using frequency domain rewards. We proved our methodology on the LIGO Livingston Observatory (LLO). Our controller reduced control noise in the 10- to 30-hertz band by over 30x and up to 100x in subbands, surpassing the design goal motivated by the quantum limit. These results highlight the potential of Deep Loop Shaping to improve current and future gravitational wave observatories and, more broadly, instrumentation and control systems.

The gravitational wave (GW) detectors LIGO and Virgo have revolutionized astrophysics by detecting mergers of exotic objects, such as black holes (BHs) and neutron stars (NSs) (1–3). Currently, most of the detectable signal lies in the 30- to 2000-Hz band, leaving the low-frequency band (10 to 30 Hz) largely unexplored. Enhancing sensitivity in this band could lead to a substantial increase in cosmological reach and thus in the scientific capabilities of LIGO (Fig. 1A). The 10- to 30-Hz band is also important for the early (premerger) detection of binary neutron stars (BNSs), potentially doubling the warning time, which would enable real-time observation of neutron star collisions, the subsequent creation of heavy elements, and the birth of black holes (4–6). However, such sensitivity improvements are currently partially limited by injected control noise on the interferometer mirrors. Furthermore, as the control noise is a bottleneck to overall sensitivity improvements, there is currently little to be gained from improvements to other noise sources. We addressed this challenge with a new tailored reinforcement learning (RL) method and improved the alignment control of the LIGO mirrors. We lowered the injected control noise on the most demanding feedback control loop, the common-hard-pitch (θ_{CHP}) loop of the Livingston Observatory,

below the quantum back-action limit. By eliminating the harmful noise from this critical representative controller, we paved the path to improve LIGO's sensitivity.

The space-time strain associated with even the loudest GW signals produces a signal equivalent to only $\approx 10^{-19}$ meters of mirror motion. As a comparison, the environmental disturbance, due to Earth tides and seismic vibration, is roughly 13 orders of magnitude larger. To measure the weak GW signals, laser-interferometric GW detectors have hundreds of optomechanical degrees of freedom that require stabilization. Active control is used to achieve precise stabilization in the face of complex mirror dynamics and inherently unstable degrees of freedom. More specifically, the optomechanical response

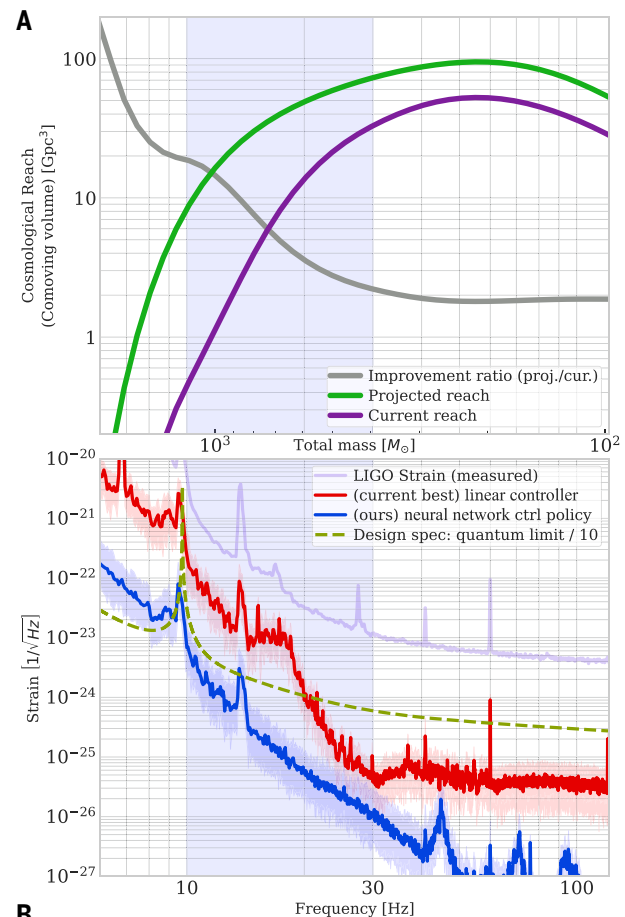


Fig. 1. Cosmological reach and strain noise from control. (A) The plot shows the volume in space explored with binary BH merger waveforms (21) for different cases of technical noise. The x axis in (A) is the total mass of the equal-mass binary pair. This corresponds to the x axis in (B), the frequency of the first quasinormal mode of a Schwarzschild BH with such a mass, as measured in the source frame. The purple trace shows the reach of LIGO as of March 2024. The green trace shows the volumetric improvement in the case where the technical noise is removed entirely. Many of the known technical noise sources are linked to controls. (B) LIGO's noise budget and controller performance. Lavender represents overall measured strain noise, red represents strain noise contribution from the currently operational linear controller for θ_{CHP} , and blue represents strain noise contribution from neural network RL policy as run on the LIGO Livingston Observatory on 5 Dec 2024 (mean, 10th and 90th percentiles of amplitude spectral density (ASD) of control action of neural network control policy). The dashed green line indicates the control design goal derived from the quantum back-action limit by applying a design margin of 10x; the control noise should drop below this curve. A detailed accounting of technical noise sources is available in (22).

¹Google DeepMind, London, UK. ²Gran Sasso Science Institute (GSSI), L'Aquila, Italy.

³Laboratori Nazionali del Gran Sasso, Assegi (INFN), Italy. ⁴LIGO Laboratory, Division of Physics, Math, and Astronomy, California Institute of Technology, Pasadena, CA, USA.

*Corresponding author. Email: buchli@google.com (J.B.); rana@caltech.edu (R.X.A.); jan.harms@gssi.it (J.H.); pushmeet@google.com (P.K.) †These authors contributed equally to this work. ‡LIGO Instrument Team authors and affiliations are listed in the supplementary materials.

of the interferometer (i.e., the plant) is subject to dynamic variations: Even low absorption of the high-power laser beam (~ 300 to 500 kW) causes thermal distortions in the mirrors, leading to offsets in sensor signals and changes in optomechanical resonant frequencies. In addition, the high-power laser also creates substantial forces and torques on the suspended mirrors, leading to optomechanical instabilities of several mechanical eigenmodes (7–9). These resonances are stabilized using feedback control, but any noise injected by the feedback controllers into the GW readout harms the peak astrophysical sensitivity and drowns out the GW signals themselves.

In simplified terms, the main control design challenge is that larger control action in lower frequencies provides better disturbance rejection but injects higher noise into the observation band. Conversely, lowering the control action reduces injected noise but results in insufficient disturbance rejection and possible loss of stability. Linear control systems theory shows fundamental limits to this trade-off (10, 11) under certain assumptions about the controller design. The ultimate aim of controller design is to shape the “closed-loop” behavior, i.e., the performance of the designed controller acting in a feedback loop with the plant.

There are many classical methods to achieve the desired closed-loop behavior. Early methods, i.e., the classic (open)-loop shaping methods, exploit the direct relationship between the open- and closed-loop transfer functions to design the controller. Since the 1980s, the focus of design has shifted from open-loop design to directly shaping closed-loop transfer functions, i.e., the sensitivity functions (12, 13), usually through optimization [e.g., convex optimization (14), H_∞ (15)]. These methods are more general and can take into

account a larger variety of design goals and constraints. Yet they still require strong assumptions, such as convexity and linearity. For GW detectors like LIGO, progress using traditional approaches has come to a plateau. Machine learning for interferometer control was presented in (16), with the primary aim to improve the contrast of the interferometer rather than to optimize closed-loop performance.

In this work, we present a new control design method, Deep Loop Shaping (DLS), to design controllers that satisfy specific demands on the system’s frequency domain behavior (Fig. 2). DLS has no constraints regarding the use of nonlinear models and control structures. It exploits the machinery of deep reinforcement learning to directly optimize frequency domain properties and shape the closed-loop behavior. We demonstrate DLS’s utility on the critical LIGO θ_{CHP} control loop, achieving state-of-the-art feedback control performance. The injected control noise was reduced by up to two orders of magnitude while maintaining mirror stability. Applying DLS more widely on LIGO can improve sensitivity. Furthermore, the method has wide applicability to control engineering; for example, highly unstable systems, vibration suppression, and noise cancellation all have strong frequency-dependent control demands.

The LIGO controls challenge

Angular sensing and control (ASC) is the challenge of maintaining the orientation of the interferometer mirrors. Stabilization is accomplished through a hybrid active-passive isolation system. Passive stabilization happens through a series of pendulums, from which the optics are suspended. These pendulums suppress seismic disturbances at frequencies above 10 Hz by several orders of magnitude. However,

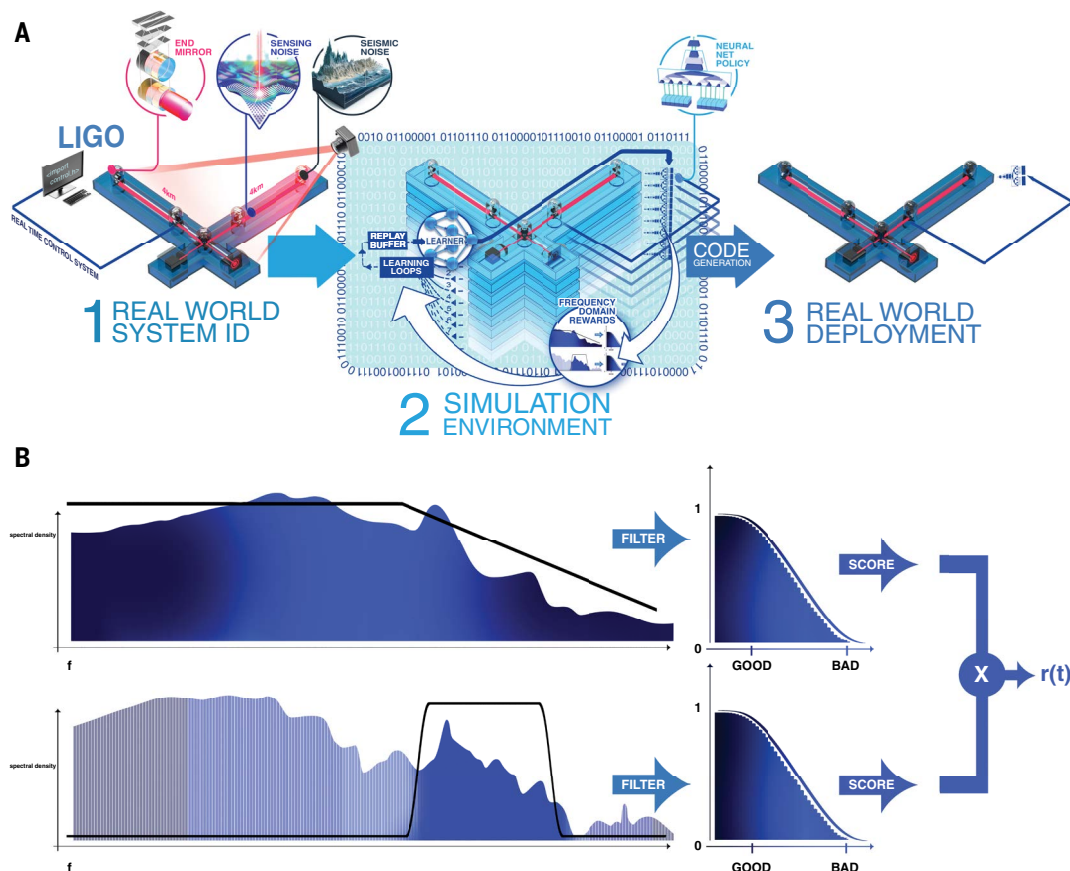


Fig. 2. DLS: Reinforcement learning with frequency domain rewards. (A) (1) A model is identified from plant measurements. (2) The identified model is used as a learning environment. Frequency domain rewards are used to compute rewards. (3) The optimized control policy is deployed on the plant. (B) Illustration of the frequency domain rewards and the multiplicative scoring.

active stabilization is required to reject seismic disturbances at frequencies below ~ 3 Hz. This stabilization is accomplished by a set of actuators that produce a torque on the suspended optics at the penultimate stage of the suspension system. Additionally, there are disturbances caused by the radiation-pressure forces of the high-power laser beam (17) that couple the angular motions of the cavity mirrors. The sensors used to measure the angular motion have a good signal-to-noise ratio at frequencies below a few hertz to enable active stabilization, but in the 10- to 30-Hz band, the sensor noise is orders of magnitude larger than the signal related to angular motion, and so motion in this band is not visible from the angular sensor and is only seen in the interferometer strain spectrum. The active controller injects sensing noise in this frequency band through the actuator to the test mass. This effect is the primary cause of test mass angular motion at frequencies above 10 Hz (8, 18). Avoiding the injection of noise as much as possible and at the same time guaranteeing rejection of seismic disturbances is the main design goal for ASC controllers. We address this problem, which comprises the most challenging control loops in a GW observatory, with DLS.

The θ_{CHP} loop

In this work, we primarily focus on the ASC control loop, “common hard pitch” (θ_{CHP}). “Hard pitch” refers to the stiffer of the two optomechanical pitch eigenmodes of the arm cavities. “Common” signifies a relation of modes between the cavities of the two arms (8). The θ_{CHP} degree of freedom is the most difficult of the entire ASC system to stabilize and optimize. Reducing the control noise in this loop well below the quantum limit would remove this source of noise as a blocking issue for improved astrophysical sensitivity.

Closed-loop shaping as a reinforcement learning problem

In this work, we formulated the θ_{CHP} closed-loop control as an optimal control problem and found approximate solutions through reinforcement learning (RL). ASC requirements are naturally expressed as functions of the system response in the frequency domain, i.e., as desired spectra of state-space signals. We introduced a new reward scheme based on frequency domain behavior to enforce the desired closed-loop shaping of the control policy in such a way that RL could discover an effective control policy. It is similar to traditionally used methods of shaping sensitivity functions, but RL removes restrictions on the reward definition and system dynamics. RL can also discover nonlinear policies represented with deep neural networks that can serve as drop-in replacements for the existing hand-crafted controllers and enables improved performance without compromising robustness.

RL designs controllers by adapting a parameterized state-action mapping. Our specific choice of learning algorithm is maximum a posteriori policy optimization (MPO) (19). We used a small multilayer perceptron (MLP) with a dilated convolution input layer for the policy network, which executes sufficiently fast for control. The critic network is a long-short-term-memory (LSTM) network with input and output MLPs, as the critic is not needed for deployment.

Frequency domain rewards

RL naturally lends itself to reward descriptions formulated in the time domain, e.g., scoring events that happen at certain times. Instead, we directly formulated the ASC requirements as rewards in the frequency domain (Fig. 2B). To do so, we designed linear filters for the θ_{CHP} response signal whose transfer functions each select a certain frequency band of the signal. We used a low-pass filter to reward pitch alignment, a band-pass filter to reduce control action in the 8- to 30-Hz band, and an additional band-pass filter for frequencies >40 Hz to avoid high-frequency artifacts. A high output from a filter at a given timestep corresponds to a large historic response in the measured frequency band. These per-timestep response

measures can then be used to construct a reward for RL. Specifically, we computed the RL reward by passing the filter outputs through a sigmoid function to compute a (per-filter) score in $[0,1]$, with a value of 1 when the specification was fulfilled and fading to 0 as the response worsened. These individual filter scores were multiplied to yield the per-timestep reward, then used by the RL method to choose a policy that minimizes the discounted sum of this reward over time. This formulation of multiplying rewards can loosely be understood as a soft logical-AND; i.e., we wanted all properties to be fulfilled for high reward.

Training and deployment

We trained nonlinear control policies with RL against a linear stochastic state-space simulation of the plant dynamics (i.e., optomechanical response of the interferometer) identified from measurement data of the plant. We used domain randomization to add robustness to the learned policies. Specifically, at the beginning of each episode, we randomized the angular instability pole frequency and sampled variations in the seismic noise, including the overall noise strength.

At the conclusion of training, we performed several steps to ready the policy for hardware testing. First, a deterministic policy was created by using only the mean of the policy Gaussian. Second, we validated this deterministic policy across a selected set of disturbances and nonnominal plant parameters. We examined the reward achieved as well as measured key performance criteria, such as root mean square of the control effort in the observation band (10 to 30 Hz), and visually inspected the error and control spectra. With performance confirmed, we “exported” the policy for the hard real-time control required for execution on LIGO without further training or adaptation on the plant.

We deployed the control policies directly in the existing control infrastructure of the interferometer (20). As such, the RL-trained policies were drop-in replacements of the existing single-input, single-output (SISO) controllers. In particular, the LIGO control system uses somewhat arbitrary “counts” as the units for ASC inputs and outputs, and we adopted these conventions for the controller and the controller-simulator interface. We have also reported our results in these units for technical reasons.

Deployment on gravitational wave observatory hardware

We ran the deployed policies on the LIGO Livingston Observatory (LLO). In the experiments, the control of θ_{CHP} was under the sole authority of a neural network-based control policy. We measured the ASC noise during policy execution as well as comparison spectra from the standard controller before and after the nonlinear policy. In Fig. 1B we compare the performance of the neural network policy against the standard controller for a >10 min stretch on 5 December 2024. The figure shows the projection of the measured angular control noise into the GW readout. Additional details of this experiment are shown in the supplementary materials.

We found excellent performance for the neural network policy. In the crucial 3- to 30-Hz band, we see a reduction of noise of up to two orders of magnitude. At the same time, the neural network policy shows similar control authority as the linear controller in the control band (<3 Hz). The control noise added by the neural network policy is well below the fundamental thermodynamic noise and quantum back-action noise in the whole band of interest. These results show that the neural network policy has effectively removed the issue of noise injected by active control as a limit to the astrophysical sensitivity.

In the supplementary materials, we present additional results from April and August 2024, with total time on the instrument of well over 1 hour. The sustained control of the unstable θ_{CHP} mode demonstrates robustness of the neural network policy to normal seismic activity. We

saw a good match between the training simulation and the real plant under the tested conditions for frequencies >0.1 Hz, which increases confidence in our results. We additionally compared the control policy against the incumbent linear controller in terms of statistical measures, such as non-Gaussianity and nonstationarity. We found that, although the policy does exhibit some nonstationarity, the overall reduction in noise still leads to a clear benefit for signal detection.

For comparison, we derived controllers with convex optimization and show a series of simulation-based results in the supplementary materials. Although these optimized linear controllers have similar predicted performance, they are not fit for deployment in the high-stakes environment of the real observatory. In particular, they are open-loop unstable, and their disturbance rejection behavior is highly aggressive, in contrast to the neural network policies. In addition to experiments on LLO, we used the same methodology on the mode-cleaner of the Caltech prototype and similarly found that DLS is capable of reducing noise in a band of interest while maintaining overall control.

REFERENCES AND NOTES

- LIGO Scientific Collaboration; Virgo Collaboration; *Phys. Rev. Lett.* **116**, 061102 (2016).
- LIGO Scientific Collaboration; Virgo Collaboration; *Phys. Rev. Lett.* **119**, 161101 (2017).
- LIGO Scientific Collaboration; Virgo Collaboration; KAGRA Collaboration; *Phys. Rev. X* **13**, 041039 (2023).
- H. Yu, R. X. Adhikari, R. Magee, S. Sachdev, Y. Chen, *Phys. Rev. D* **104**, 062004 (2021).
- B. Banerjee *et al.*, *Astron. Astrophys.* **678**, A126 (2023).
- A. Tohuvavohu *et al.*, *Astrophys. J. Lett.* **975**, L19 (2024).
- J. A. Sidles, D. Sigg, *Phys. Lett. A* **354**, 167–172 (2006).
- L. Barsotti, M. Evans, P. Fritschel, *Class. Quantum Gravity* **27**, 084026 (2010).
- V. Braginsky, S. Strigin, S. Vyatchanin, *Phys. Lett. A* **287**, 331–338 (2001).
- K. J. Aström, R. Murray, *Feedback Systems: An Introduction for Scientists and Engineers* (Princeton Univ. Press, ed. 2, 2021); <https://books.google.com/books?id=l50DEAAQBAJ>.
- G. Stein, *IEEE Control Syst.* **23**, 12–25 (2003).
- H. W. Bode, *Network Analysis and Feedback Amplifier Design* (D. Van Nostrand Company, 1945).
- G. Zames, *IEEE Trans. Automat. Contr.* **26**, 301–320 (1981).
- C. Barratt, S. Boyd, in *Control and Dynamical Systems: Digital and Numeric Techniques and Their Applications in Control Systems, Part I*, vol. 55, C. T. Leondes, Ed. (Academic Press, 1993), pp. 1–24.
- J. C. Doyle, K. Glover, P. P. Khargonekar, B. A. Francis, *IEEE Trans. Automat. Contr.* **34**, 831–847 (1989).
- N. Mukund *et al.*, *Phys. Rev. Appl.* **20**, 064041 (2023).
- K. L. Dooley *et al.*, *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* **30**, 2618–2626 (2013).
- H. Yu *et al.*, *Phys. Rev. Lett.* **120**, 141102 (2018).
- A. Abdolmaleki *et al.*, Maximum a Posteriori Policy Optimisation, *International Conference on Learning Representations*, Vancouver, CA, 30 April to 3 May 2018. (ICLR, 2018).
- R. Bork *et al.*, *SoftwareX* **13**, 100619 (2021).
- S. Khan *et al.*, *Phys. Rev. D* **93**, 044007 (2016).
- E. Capote *et al.*, Advanced LIGO Detector Performance in the Fourth Observing Run. [gr-qc] (2024).
- M. Hoffman *et al.*, Acme: A research framework for distributed reinforcement learning. arXiv:2006.00979 [cs.LG] (2020).
- F. Yang *et al.*, Launchpad: A Programming Model for Distributed Machine Learning Research (2021) [cs.DC], arXiv:2106.04516.
- google-deepmind, dm env: A python interface for reinforcement learning environments, GitHub (2019); https://github.com/deepmind/dm_env.
- T. Hennigan, T. Cai, T. Norman, L. Martens, I. Babuschkin, Haiku: Sonnet for JAX (2024); <https://github.com/deepmind/dm-haiku>.
- A. Cassirer *et al.*, Reverb: A Framework For Experience Replay. arXiv:2102.04736 [cs.LG] (2021).
- J. Harms, Lightsaber, <https://github.com/janosch314/Lightsaber> (2025).
- LIGO Scientific Collaboration, Identified Plant Model & Selected ASC Experimental Data of LIGO Livingston, Zenodo (2025); <https://doi.org/10.5281/zenodo.15793015>.

ACKNOWLEDGMENTS

We thank J. Dean for strategic help and inspiration at the start of the project. **Funding:** The authors gratefully acknowledge the support of the US National Science Foundation (NSF) for the construction and operation of the LIGO Laboratory and Advanced LIGO as well as the Science and Technology Facilities Council of the UK and the Max Planck Society for support of the construction of Advanced LIGO. Additional support for Advanced LIGO was provided by the Australian Research Council. LIGO was constructed by the California Institute of Technology and Massachusetts Institute of Technology with funding from the NSF and operates under cooperative agreement no. PHY-18671764464. Advanced LIGO was built under grant no. PHY-18680823459. **Author contributions:** R.X.A., J.Bu., S.C., J.H., M.R., and B.T. conceived the project. R.X.A., J.Bu., C.D., A.H., J.H., and B.T. led the project. T.A., R.X.A., I.B., J.Bu., J.Be., C.D., G.v.d.D., J.D., A.G., J.H., M.L., B.T., G.T., and C.W. developed the physics simulations. T.A., I.B., J.Bu., Y.H.J.C., C.D., J.D., M.L., B.T., and C.W. integrated the physics simulations with the learning framework. A.A., J.Bu., R.H., S.H., M.L., M.W., and B.T. developed the learning framework and performed learning experiments. C.D., T.N., J.R., and C.W. developed the real-time neural network interface. R.X.A., J.Be., J.Bu., C.D., A.G., B.T., and C.W. integrated the real-time neural network with the control system and ran experiments on LLO and the California Institute of Technology 40m prototype. C.B., J.Bu., Y.H.J.C., C.D., Y.D., O.G., M.L., and C.W. developed data curation tools. R.X.A., I.B., J.Bu., Y.H.J.C., B.T., and C.W. developed and ran the data analysis. L.F., P.K., H.O., and M.R. consulted for the project. T.A., R.X.A., J.Bu., J.H., B.T., and C.W. wrote the manuscript. The LIGO Instrument Team maintains and runs the LIGO Observatory. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The learning algorithm used in the actor-critic RL method is MPO (19), a reference implementation of which is available under an open-source license (23). Additionally, the software libraries launchpad (24), dm env (25), Jax/Haiku (26), and reverb (27) were used, which are also open source. Simulations were implemented in Lightsaber (28) and advLigoRTS (20). The identified LLO model and experimental data are available at (29). **License information:** Copyright © 2025 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. <https://www.science.org/about/science-licenses-journal-article-reuse>

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adw1291
Supplementary Text; Figs. S1 to S16; Tables S1 to S3; References (30–70)

Submitted 28 January 2025; accepted 7 July 2025

10.1126/science.adw1291

Correction (24 October 2025): Reference 16 was inadvertently removed during the review process; it was added back in, and all subsequent references were renumbered. Additionally, the date of experiment was listed incorrectly in the main text and supplementary materials, and one of the non-byline author affiliations in the supplementary materials was accidentally duplicated; both of these errors were corrected.