

Quantum feedback control with a transformer neural network architecture

Pranav Vaidhyathan ¹, Florian Marquardt ^{2,3}, Mark T. Mitchison ^{4,5,*} and Natalia Ares ^{1,†}

¹Department of Engineering Science, University of Oxford, Oxford OX1 3PJ, United Kingdom

²Max Planck Institute for the Science of Light, Staudtstraße 2, 91058 Erlangen, Germany

³Department of Physics, Friedrich-Alexander-Universität Erlangen-Nürnberg, 91058 Erlangen, Germany

⁴School of Physics, Trinity College Dublin, College Green, Dublin 2 D02 K8N4, Ireland

⁵Department of Physics, King's College London, Strand, London WC2R 2LS, United Kingdom



(Received 21 December 2024; accepted 1 November 2025; published 24 February 2026)

Attention-based neural networks such as transformers have revolutionized various fields such as natural language processing, genomics, and vision. Here, we demonstrate the use of transformers for quantum feedback control through both a supervised and a reinforcement learning approach. In particular, due to the transformer's ability to capture long-range temporal correlations and training efficiency, we show that it can surpass some of the limitations of previous control approaches, e.g., those based on recurrent neural networks trained using a similar approach or policy-based reinforcement learning. We numerically show, for the example of state stabilization of a two-level system, that our bespoke transformer architecture can achieve near-unit fidelity to a target state in a short time even in the presence of inefficient measurement and Hamiltonian perturbations that were not included in the training set as well as the control of non-Markovian systems. We also demonstrate that our transformer can perform energy minimization of nonintegrable many-body quantum systems when trained for reinforcement learning tasks. Our approach can be used for quantum error correction, fast control of quantum states in the presence of colored noise, as well as real-time tuning, and characterization of quantum devices.

DOI: [10.1103/m429-jy1j](https://doi.org/10.1103/m429-jy1j)

Introduction. Quantum technologies depend crucially on our ability to precisely control quantum systems. Measurement-based feedback is an especially powerful approach to quantum control, which lies at the heart of quantum error correction and has myriad applications in the preparation and stabilization of quantum states in the presence of noise [1–3]. However, unlike noisy feedback control in the classical regime, quantum feedback faces an additional obstacle: Only partial information on the quantum state is available, even in principle, due to the inherently disturbing nature of quantum measurements. In general, therefore, optimizing control fields requires—in addition to the measurement record—an estimate of the quantum state to be explicitly computed from some model of the dynamics [4], adding overhead (e.g., extra memory or time costs) to the feedback loop.

Machine learning algorithms offer a promising route to solve this problem. Appropriately trained neural networks can provide a compact representation of the most important correlations between data, allowing for significantly more efficient feedback protocols, in principle. Recent work in this direction has demonstrated the power of both model-free [5–8] and

model-based [9,10] reinforcement learning for quantum feedback control. Although the former approach is a “black box” that is flexible enough to be applied to a range of scenarios, the latter can exploit the physics of the system to improve efficiency.

However, approaches using recurrent neural networks (RNN) do not scale well with long-range dependencies, such as an extensive measurement record. They also suffer from the problem of vanishing gradients [11]. This is due to the assumption that the hidden state within each recurrent unit encodes dependencies from the previous state. This adds an inherent Markovian inductive bias that is unsuitable for processes with memory [12]. Clearly, this poses a challenge for feedback control of non-Markovian open system dynamics [13], which arises naturally in many platforms. Yet, even for Markovian open quantum systems (i.e., those described by a Lindblad equation), the measurement record is a non-Markovian stochastic process [14] because measurement backaction causes the future evolution of the state to depend unavoidably on past measurement outcomes.

In recent years, transformers and attention-based models that were originally used to model natural language [15] have emerged as extremely versatile tools for various fields, ranging from genomics to robotics [16]. Due to the attention mechanism that encodes correlation between all aspects of the given input sequence, they have far outclassed recurrent neural networks (RNNs) and long short-term memory type RNNs (LSTMs) in various tasks. Recently, transformer-based approaches have also demonstrated their ability not only to adapt to model-based and model-free reinforcement learning

*Contact author: mark.mitchison@kcl.ac.uk

†Contact author: natalia.ares@eng.ox.ac.uk

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

tasks [17,18] but also to perform on par with state-of-the-art approaches for these tasks. As highlighted by Chen, *et al.*, these causally masked transformers simply output optimal actions and eliminate the requirement to fit value functions or calculate policy gradients [18].

In this work, we aim to utilize the “unreasonable” effectiveness of the attention mechanism in order to perform closed-loop feedback control for continuously measured open quantum systems that also undergo evolution due to measurement backaction [4,19]. We demonstrate that an attention-based approach to quantum feedback control, with neural networks trained using both supervised and reinforcement learning, offers a robust and scalable solution that outperforms traditional methods used for quantum feedback control. Using transfer learning, we can demonstrate that the transformer also generalizes well to non-Markovian open quantum systems, which has yet to be demonstrated by existing methods.

Setup. We consider a quantum system undergoing a continuous weak measurement of the diffusive kind, e.g., homodyne readout in quantum optics [20,21] or electronic charge detection by a quantum point contact [22,23]. Let $\hat{\rho}_t$ denote the state of the system at time t , conditioned on the measurement record $\mathbf{r}_t$. An experimenter uses their knowledge of the measurement record to control the system by manipulating some control parameter λ_t entering its Hamiltonian $\hat{H}(\lambda_t)$. The conditional dynamics is then described by a stochastic master equation of the form [4]

$$d\hat{\rho}_t = \frac{1}{i\hbar}[\hat{H}(\lambda_t), \hat{\rho}_t]dt + \mathcal{D}[\hat{c}]\hat{\rho}_t dt + \sqrt{\eta}\mathcal{H}[\hat{c}]\hat{\rho}_t dW_t, \quad (1)$$

where the jump operator $\hat{c}$ describes the effect of coupling to the measuring device, the dissipation superoperator is $\mathcal{D}[\hat{c}]\hat{\rho} = \hat{c}\hat{\rho}\hat{c}^\dagger - \frac{1}{2}(\hat{c}^\dagger\hat{c}\hat{\rho} + \hat{\rho}\hat{c}^\dagger\hat{c})$, the innovation superoperator is $\mathcal{H}[\hat{c}]\hat{\rho} = \hat{c}\hat{\rho} + \hat{\rho}\hat{c}^\dagger - \text{Tr}[(\hat{c} + \hat{c}^\dagger)\hat{\rho}]\hat{\rho}$, the measurement efficiency is η , and the measurement noise in each small time step dt is described by independent Wiener increments dW_t with zero mean and variance $dW_t^2 = dt$ [24]. The measurement record r_t increments according to

$$dr_t = \text{Tr}[(\hat{c} + \hat{c}^\dagger)\hat{\rho}_t]dt + \frac{dW_t}{\sqrt{\eta}}. \quad (2)$$

Note that we use boldface notation to distinguish the history of the measurement record up to time t , $\mathbf{r}_t = (\dots, r_{t-2dt}, r_{t-dt}, r_t)$, from its instantaneous value, r_t . For simplicity, we consider a single jump operator c and control parameter λ_t , but our method can be generalized straightforwardly to the case of multiple jump operators and control parameters.

In a general feedback protocol, the control parameter for the next time step is determined by the entire past history of the measurement record; i.e., it is a functional $\lambda_{t+dt}[\mathbf{r}_t]$. In the simplest case of linear feedback [20], the control parameter is proportional to the measurement result, $\lambda_{t+dt} \propto dr_t$, but this approach permits a very limited class of protocols and also suffers badly from measurement inefficiencies [25,26]. More general state-based methods [27] decide the optimal feedback using an (implicit or explicit) model of the conditional quantum state, e.g., by solving Eq. (1) using experimentally obtained values for the measurement noise dW_t .

Alternatively, reinforcement learning creates an implicit model of the dynamics in terms of a probability distribution (policy function) $\pi_\theta(\lambda_{t+dt}|\mathbf{r}_t)$, which is represented as a neural network parametrized by some weights and biases θ . While recent successes of this approach have been demonstrated using the RNN architecture [5,10], here we take a different approach based on the transformer architecture [15].

Transformer model. Our model consists of a custom transformer encoder-decoder architecture (see Fig. 1), which we name QuantumEncoder and QuantumDecoder, to determine λ_t at each time step. At its core, a transformer is designed to process sequential data by capturing long-range dependencies and contextual information. Unlike traditional recurrent neural networks that process sequences step by step, transformers employ a mechanism called self-attention to attend to different parts of the input simultaneously.

The QuantumEncoder processes the initial quantum state and the measurement record, embedding it into a higher-dimensional space and capturing dependencies through self-attention mechanisms. The QuantumDecoder takes the measurement record, with positional embeddings, as its input since this detailed sequential or spatial information is not provided by the encoder’s compressed representation of core features from its input (its latent space). By feeding the measurement record into the decoder, we allow the model to adaptively adjust the optimal control parameter at each time step based on the observed system dynamics. The decoder employs self-attention that is causally masked, preventing it from “seeing” future measurements, to ensure that predictions for the optimal control parameters at each time step are based only on current and past measurements. During training, the QuantumDecoder module takes the optimal parameter values and measurement record as input and learns to predict the next value λ_t in the sequence. We use the fidelity between the evolved state due to λ_t and target state as the loss function used to train our model. The output of the decoder’s last layer is passed through a linear transformation followed by a softmax function to obtain a probability distribution over the λ_t values. We also set up a sweep to optimize for optimal number of layers, the learning rate, optimizer, and the number of epochs (see the Supplemental Material [28]).

State stabilization in the two-level system. To demonstrate the effectiveness of the transformer-based approach, we showcase a numerical example of quantum state stabilization. The loss function in this case is the infidelity between the conditional state and some pure target state $|\psi_{\text{targ}}\rangle$,

$$L = 1 - \langle \psi_{\text{targ}} | \hat{\rho}_t | \psi_{\text{targ}} \rangle. \quad (3)$$

Our model system comprises a two-level system with Hamiltonian

$$\hat{H}(\lambda_t) = \frac{\hbar\varepsilon}{2}\hat{\sigma}_z + \frac{\hbar\lambda_t}{2}\hat{\sigma}_x, \quad (4)$$

undergoing a continuous measurement described by the jump operator $\hat{c} = \sqrt{\kappa}\hat{\sigma}_-$. Here, ε denotes a fixed energy bias and κ denotes the measurement rate.

To train the neural network, we prepare a dataset consisting of several initial states of the two-level system that are then evolved using the smesolve method from the QuTiP Python package used for simulating open quantum systems [29]. We

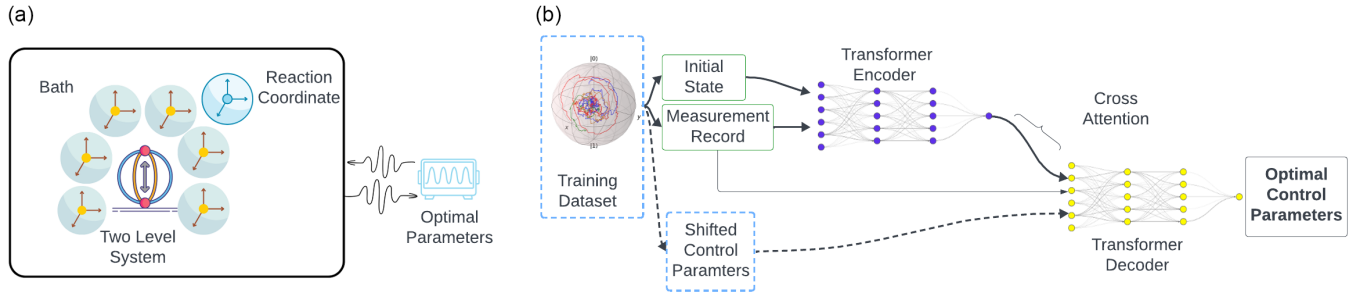


FIG. 1. Problem and architecture overview: (a) The two-level system (TLS) interacting with a bath in a non-Markovian manner that is embedded with a reaction coordinate (RC) that interacts in a Markovian manner. The measurement record obtained continuously is used to predict optimal values of the control parameters by the transformer. (b) The transformer's structure consists of an encoder and decoder architecture. During training, the encoder takes the initial state and the measurement record as input (green dotted boxes). The decoder takes the encoder output as part of the cross attention layer and the optimal parameters (blue dotted boxes), to autoregressively predict the optimal parameters for the next time steps. However, during inference, only the initial state and measurement record are given as input to the encoder (solid arrows). The decoder then predicts the optimal next values of the control parameters based on these data.

then train our transformer using the dataset consisting of a range of initial states, their associated measurement records, and a locally optimal control protocol, λ_t , that drives the system to the target state for each noise realization. This locally optimal control is found using the proportional and quantum state (PaQS) algorithm [27]. During the training phase, we always set $\varepsilon = 0$.

As seen in Fig. 2, the transformer generates feedback strategies that can stabilize the TLS in a coherent superposition state $|\psi_{\text{targ}}\rangle = (|0\rangle + i|1\rangle)/\sqrt{2}$ even with inefficient measurements (we take $\eta = 0.7$). The transformer approach is also robust against perturbations of the dynamics, as we show by introducing a significant bias $\varepsilon \neq 0$, which was absent during training. Further examples can be found in the Supplemental Material [28].

Another beneficial feature of the transformer is the speed with which it outputs the optimal control parameter for the next time step. Table I compares the time taken to infer the entire trajectory by the transformer algorithm and the

modified PaQS algorithm, where the latter requires solving the stochastic master equation (1) at each time step. We observe a speedup of approximately 2 orders of magnitude in our numerical tests, which were performed using a standard laptop. This is because PaQS and other iterative numerical solvers require multiple evaluations of the right-hand side of Eq. (1) per gradient evaluation and per line search, whereas our deployed transformer avoids online integration of the stochastic master equation altogether. However, it should be noted that this speed advantage comes at the cost of a large memory required to store the neural network representation. This memory requirement is likely to prove the most significant bottleneck for integrating the transformer into optimized hardware such as graphics processing units (GPUs) or field-programmable gate arrays (FPGAs) [8].

In order to provide another example to demonstrate the flexibility and generalizability of our transformer-based approach, we apply it to the challenging problem of controlling non-Markovian quantum dynamics [30]. Specifically, we now consider our TLS to be coupled to a harmonic oscillator mode with angular frequency Ω and coupling strength g , leading to the Hamiltonian

$$\hat{H}(\lambda_t) = \frac{\hbar\varepsilon}{2}\hat{\sigma}_z + \frac{\hbar\lambda_t}{2}\hat{\sigma}_x + \hbar\Omega\hat{a}^\dagger\hat{a} + \hbar g\hat{\sigma}_z(\hat{a} + \hat{a}^\dagger). \quad (5)$$

We assume that the oscillator mode is coupled to a broadband environment that is continuously monitored via homodyne detection, leading to a stochastic master equation of the form

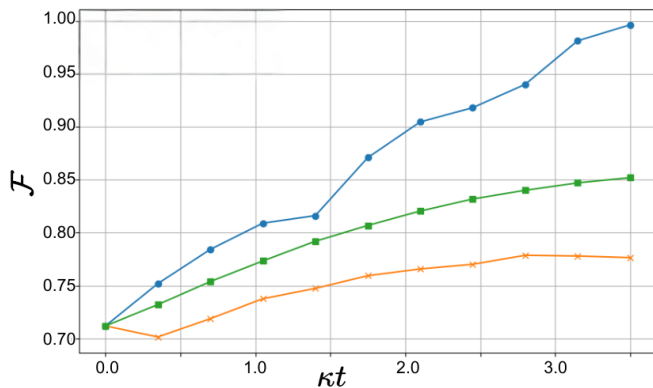


FIG. 2. Fidelity $\mathcal{F}$ with a target state as a function of time under feedback control. The initial state is $\hat{\rho}_0 = |\psi_0\rangle\langle\psi_0|$, where $|\psi_0\rangle = \alpha|0\rangle + \beta|1\rangle$ with $\alpha = \sqrt{\frac{7}{12}}$ and $\beta = \sqrt{\frac{5}{12}}$. The target state is $|\psi_{\text{targ}}\rangle = \frac{1}{\sqrt{2}}(|0\rangle + i|1\rangle)$. The performance of the transformer under imperfect measurement efficiency ($\eta = 0.7$) (blue, circles), and an increase in bias ($\varepsilon = 0.5$) (green, squares) is benchmarked against the fidelity improvement when randomly selecting λ_t values (orange, crosses).

TABLE I. The inference time in seconds of predicting optimal λ_t for a single trajectory with 100 discretized time steps during the evolution of the state governed by the stochastic master equation (1). We benchmark the inference time of the transformer against the time taken to calculate the optimal feedback operation using a gradient-based solver for the PaQS approach. The inference is run on a 2021 MacBook Pro with 16 GB of random access memory and an 8-core CPU.

Method	Inference time (s)
Hamiltonian modified PaQS	19.05
Quantum transformer inference	0.23

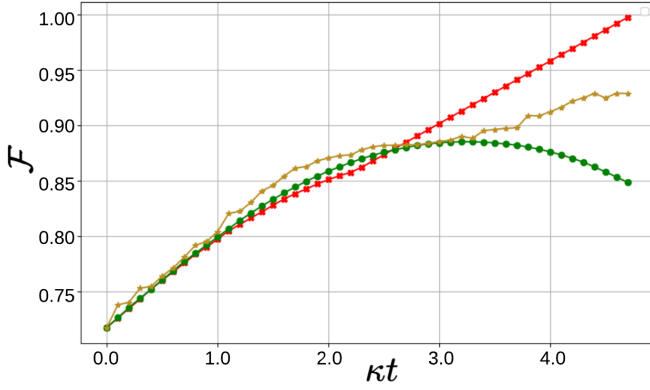


FIG. 3. The fidelity $\mathcal{F}$ as a function of time while benchmarking the performance of the transformer (red, crosses) as compared to a vanilla recurrent neural network (green, circles) and a gated recurrent unit recurrent neural network (yellow, stars). The context of 2000 measurement record samples is provided in the case of the non-Markovian setting accounted for by the reaction coordinate embedding. The coupling with the bath is provided by $g = 0.5$. The dimension of the reaction coordinate is truncated to $d = 6$.

(1) with $\hat{c} = \sqrt{\kappa}\hat{a}$. This situation can be realized, for example, by a superconducting qubit interfaced with a cavity resonator that is itself coupled to a waveguide [31]. This situation is well known to lead to non-Markovian dynamics for the qubit when the cavity linewidth κ is not too large, i.e., if $\kappa \lesssim g$ [32].

Alternatively, one can interpret the cavity mode as a “reaction coordinate,” which represents a collective mode of a structured reservoir whose spectral density is peaked at frequency Ω [33,34]. The extended open quantum system comprising the TLS and RC can thus be understood as a Markovian embedding of the original non-Markovian dynamics induced by the structured reservoir [35]. Meanwhile, the residual (broadband) environment represents far-field degrees of freedom that can be monitored without disrupting the non-Markovian character of the TLS evolution.

Since non-Markovian effects can affect the system dynamics in a much longer time horizon, the attention-based transformer model seems ideal to control non-Markovian systems due to the self-attention and multihead attention. We use transfer learning to fine-tune the transformer to predict optimal λ_t values for the reaction-coordinate setting with a smaller dataset. As seen in Fig. 3, the transformer learns to predict optimal λ_t even in this non-Markovian setting. To benchmark the transformer, we train a vanilla RNN and a gated recurrent unit recurrent neural network (GRU-RNN) with up to 60 time steps of the measurement record [36]. We train the RNNs on the given number of time steps to avoid the vanishing gradient problem as explored in the literature of deep learning [37]. We can observe that even though the vanilla RNN and GRU-RNN perform slightly better than the transformer for shorter time periods, where the measurement record provided is much smaller, the transformer outperforms the vanilla RNN and the GRU-RNN in the case of long context windows and non-Markovian closed-loop feedback control. This is likely due to the fact that the transformer can process and attend to arbitrary measurement record lengths and does

TABLE II. Comparison of ground-state and predicted energies for different numbers of spins. We set $g = 6.07$, $\lambda_0 = -0.097g$, $\lambda_T = 1.27g$, $h = 1.19g$, $\kappa = 0.148g$, and $gT = 5.56$.

Spins	Ground-state energy	Mean predicted energy
4	$-6.518g$	$-6.209g$
6	$-9.723g$	$-9.666g$
8	$-13.155g$	$-12.845g$

not suffer from the requirement of a sequential approach such as from RNNs.

Many-body state preparation. In the preceding examples, we demonstrated that our transformer can be trained via supervised learning, using known locally optimal protocols such as PaQS. We now turn to the example of approximate ground-state preparation of a many-body system via feedback control [38], where no such baseline is available and our transformer can no longer be trained in a supervised manner [39]. Consider an open N -qubit mixed-field Ising chain governed by the Hamiltonian

$$\hat{H}(\lambda_t) = \lambda_t \sum_{i=1}^N \hat{\sigma}_i^z + g \sum_{i=1}^{N-1} \hat{\sigma}_i^x \hat{\sigma}_{i+1}^x + h \sum_{i=1}^N \hat{\sigma}_i^x, \quad (6)$$

where g is the nearest-neighbor coupling strength, h is a static longitudinal field, and the transverse field λ_t is the control parameter. Our goal is to sweep this parameter from an initial value $\lambda_0 \approx 0$ —for which the tensor-product ground state is generally easy to prepare—to a final value $\lambda_T \gtrsim g$, within a fixed time T , while ideally leaving the system close to the ground state of the final Hamiltonian. This is generally challenging in a finite-time protocol due to the creation of nonadiabatic excitations [40,41]. Moreover, the Hamiltonian in Eq. (6) is nonintegrable [42,43] and thus finding locally optimal control solutions becomes unfeasible as N grows larger.

The feedback is now conditioned on a homodyne measurement of the collective spin operator $\hat{c} = \sqrt{\kappa} \sum_i \hat{\sigma}_i^z$, which produces diffusive dynamics as given by Eq. (1), and an additional measurement of the mean energy $\text{tr}[\hat{H}(\lambda_T)\hat{\rho}_T]$ at the final time. The latter provides the loss function to be minimized by the control protocol for a fixed time T and end points $\{\lambda_0, \lambda_T\}$. Crucially, since labeled optimal trajectories are now absent, we utilize a model-free reinforcement learning type approach [17] that allows our model to learn optimal control pulses through trial and error by direct interaction and observation of rewards. We use the strategy of an online iteratively refined decision transformer training mechanism [44]. Our inputs remain consistent with the previous examples; however, we also include the return-to-go award based on the final energy. Table II shows the mean final energies reached by our trained transformer, demonstrating good performance even as the system size grows; e.g., we observe a maximum discrepancy with the true ground-state energy of 2% for eight spins. These results demonstrate the flexibility of our transformer architecture, as it can be adapted to situations where model-based supervised learning is challenging or impossible.

Conclusion. In this work, we have presented an approach for closed-loop adaptive feedback control of open quantum systems using attention-based transformer neural networks.

We have demonstrated that our quantum transformer model can effectively learn to predict optimal control parameters based on the initial state and measurement record of a two-level quantum system. The transformer architecture, with its self-attention mechanism, enables capturing long-range dependencies in the measurement record, outperforming traditional methods like recurrent neural networks for feedback control in non-Markovian systems where different temporal aspects of the measurement record may affect the state evolution. Due to these reasons, transformer models are an ideal candidate for applications in quantum prediction [45] as well as control, which we have demonstrated in this work.

We have shown the robustness and scalability of our approach under various conditions, such as imperfect measurement efficiency and perturbations in the Hamiltonian. Furthermore, using transfer learning, we have successfully applied our transformer model to the challenging task of controlling non-Markovian open quantum systems, which is a significant advancement in the field. The attention-based approach to quantum feedback control offers several advantages, including faster inference times compared to state-of-the-art methods like the Hamiltonian modified PaQS. The transformer's ability to handle long context windows and its scalability make it a promising tool for controlling complex quantum systems. Furthermore, we demonstrated the transformer's ability to perform reinforcement learning when optimal parameters for supervised training are inaccessible, highlighting the promise of our approach for tackling challenging many-body control problems. In conclusion, our work demonstrates the effectiveness of attention-based transformer models for closed-loop feedback control of Markovian,

non-Markovian, and many-body open quantum systems. This approach opens up possibilities for the development of robust and efficient quantum control techniques, which are crucial for the advancement of quantum technologies. Future work could explore the application of this method to even more complex quantum systems and investigate its performance in experimental settings.

Acknowledgments. The authors acknowledge useful discussions with Professor Gerard Milburn. N.A. acknowledges support from the European Research Council (Grant Agreement No. 948932) and the Royal Society (URF-R1-191150). M.T.M. was supported by a Royal Society University Research Fellowship. The research of F.M. was partially supported by the Munich Quantum Valley, which is supported by the Bavarian state government with funds from the Hightech Agenda Bayern Plus. This project was cofunded by the European Union and UK Research & Innovation (Quantum Flagship project ASPECTS, Grant Agreement No. 101080167). P.V. was supported by the United States Army Research Office under Award No. W911NF-21-S-0009-2. The authors would like to acknowledge the use of the University of Oxford Advanced Research Computing (ARC) facility in carrying out this work [46].

Views and opinions expressed are, however, those of the authors only and do not necessarily reflect those of the European Union, Research Executive Agency or UK Research & Innovation. Neither the European Union nor UK Research & Innovation can be held responsible for them.

Data availability. The data that support the findings of this Letter are openly available [47].

-
- [1] J. Zhang, Y.-X. Liu, R.-B. Wu, K. Jacobs, and F. Nori, Quantum feedback: Theory, experiments, and applications, *Phys. Rep.* **679**, 1 (2017).
 - [2] M. Puviani, S. Borah, R. Zen, J. Olle, and F. Marquardt, Non-Markovian feedback for optimized quantum error correction, *Phys. Rev. Lett.* **134**, 020601 (2025).
 - [3] H. M. Wiseman, S. Mancini, and J. Wang, Bayesian feedback versus Markovian feedback in a two-level atom, *Phys. Rev. A* **66**, 013807 (2002).
 - [4] H. M. Wiseman and G. J. Milburn, *Quantum Measurement and Control*, 1st ed. (Cambridge University Press, Cambridge, UK, 2009).
 - [5] T. Fösel, P. Tighineanu, T. Weiss, and F. Marquardt, Reinforcement learning with neural networks for quantum feedback, *Phys. Rev. X* **8**, 031084 (2018).
 - [6] Z. T. Wang, Y. Ashida, and M. Ueda, Deep reinforcement learning control of quantum cartpoles, *Phys. Rev. Lett.* **125**, 100401 (2020).
 - [7] V. V. Sivak, A. Eickbusch, H. Liu, B. Royer, I. Tsioutsios, and M. H. Devoret, Model-free quantum control with reinforcement learning, *Phys. Rev. X* **12**, 011059 (2022).
 - [8] K. Reuer, J. Landgraf, T. Fösel, J. O'Sullivan, L. Beltrán, A. Akin, G. J. Norris, A. Remm, M. Kerschbaum, J.-C. Besse, F. Marquardt, A. Wallraff, and C. Eichler, Realizing a deep reinforcement learning agent for real-time quantum feedback, *Nat. Commun.* **14**, 7138 (2023).
 - [9] S. Borah, B. Sarma, M. Kewming, G. J. Milburn, and J. Twamley, Measurement-based feedback quantum control with deep reinforcement learning for a double-well nonlinear potential, *Phys. Rev. Lett.* **127**, 190403 (2021).
 - [10] R. Porotti, V. Peano, and F. Marquardt, Gradient-ascent pulse engineering with feedback, *PRX Quantum* **4**, 030305 (2023).
 - [11] Y. Bengio, P. Simard, and P. Frasconi, Learning long-term dependencies with gradient descent is difficult, *IEEE Trans. Neural Netw.* **5**, 157 (1994).
 - [12] E. Genois, J. A. Gross, A. Di Paolo, N. J. Stevenson, G. Koolstra, A. Hashim, I. Siddiqi, and A. Blais, Quantum-tailored machine-learning characterization of a superconducting qubit, *PRX Quantum* **2**, 040355 (2021).
 - [13] H.-P. Breuer, E.-M. Laine, J. Piilo, and B. Vacchini, Colloquium: Non-Markovian dynamics in open quantum systems, *Rev. Mod. Phys.* **88**, 021002 (2016).
 - [14] G. T. Landi, M. J. Kewming, M. T. Mitchison, and P. P. Potts, Current fluctuations in open quantum systems: Bridging the gap between quantum continuous measurements and full counting statistics, *PRX Quantum* **5**, 020201 (2024).
 - [15] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, Attention is all you need,

- in *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems* (Curran Associates, Inc., 2017), Vol. 30, pp. 6000–6010.
- [16] U. Kamath, K. Graham, and W. Emara, *Transformers for Machine Learning: A Deep Dive* (Chapman & Hall/CRC, Boca Raton, Florida, USA, 2022).
- [17] M. Janner, Q. Li, and S. Levine, Offline reinforcement learning as one big sequence modeling problem, in *NIPS'21: Proceedings of the 35th International Conference on Neural Information Processing Systems* (Curran Associates, Inc., 2021), Vol. 34, pp. 1273–1286.
- [18] L. Chen, K. Lu, A. Rajeswaran, K. Lee, A. Grover, M. Laskin, P. Abbeel, A. Srinivas, and I. Mordatch, Decision transformer: Reinforcement learning via sequence modeling, in *Advances in Neural Information Processing Systems* (NeurIPS, 2021), Vol. 34, pp. 15084–15097.
- [19] K. Jacobs and D. A. Steck, A straightforward introduction to continuous quantum measurement, *Contemp. Phys.* **47**, 279 (2006).
- [20] H. M. Wiseman and G. J. Milburn, Quantum theory of optical feedback via homodyne detection, *Phys. Rev. Lett.* **70**, 548 (1993).
- [21] H. M. Wiseman, Quantum theory of continuous feedback, *Phys. Rev. A* **49**, 2133 (1994).
- [22] A. N. Korotkov, Continuous quantum measurement of a double dot, *Phys. Rev. B* **60**, 5737 (1999).
- [23] H.-S. Goan, G. J. Milburn, H. M. Wiseman, and H. Bi Sun, Continuous quantum measurement of two coupled quantum dots using a point contact: A quantum trajectory approach, *Phys. Rev. B* **63**, 125326 (2001).
- [24] Y. Jiang, X. Wang, L. Martin, and K. B. Whaley, Optimality of feedback control for qubit purification under inefficient measurement, *Phys. Rev. A* **102**, 022612 (2020).
- [25] J. Wang and H. M. Wiseman, Feedback-stabilization of an arbitrary pure state of a two-level atom, *Phys. Rev. A* **64**, 063810 (2001).
- [26] M. T. Mitchison, J. Goold, and J. Prior, Charging a quantum battery with linear feedback control, *Quantum* **5**, 500 (2021).
- [27] S. Zhang, L. S. Martin, and K. B. Whaley, Locally optimal measurement-based quantum feedback with application to multiqubit entanglement generation, *Phys. Rev. A* **102**, 062418 (2020).
- [28] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/m429-jy1j> for more examples and details on the model training, architecture, and performance.
- [29] J. Johansson, P. Nation, and F. Nori, QuTiP 2: A Python framework for the dynamics of open quantum systems, *Comput. Phys. Commun.* **184**, 1234 (2013).
- [30] C.-F. Li, G.-C. Guo, and J. Piilo, Non-Markovian quantum dynamics: What is it good for?, *Europhys. Lett.* **128**, 30001 (2020).
- [31] A. Blais, R.-S. Huang, A. Wallraff, S. M. Girvin, and R. J. Schoelkopf, Cavity quantum electrodynamics for superconducting electrical circuits: An architecture for quantum computation, *Phys. Rev. A* **69**, 062320 (2004).
- [32] H.-P. Breuer and F. Petruccione, *The Theory of Open Quantum Systems* (Oxford University Press, Oxford, 2002).
- [33] J. Iles-Smith, N. Lambert, and A. Nazir, Environmental dynamics, correlations, and the emergence of noncanonical equilibrium states in open quantum systems, *Phys. Rev. A* **90**, 032114 (2014).
- [34] D. Tamascelli, A. Smirne, S. F. Huelga, and M. B. Plenio, Nonperturbative treatment of non-Markovian dynamics of open quantum systems, *Phys. Rev. Lett.* **120**, 030402 (2018).
- [35] M. P. Woods, R. Groux, A. W. Chin, S. F. Huelga, and M. B. Plenio, Mappings of open quantum systems onto chain representations and Markovian embeddings, *J. Math. Phys.* **55**, 032101 (2014).
- [36] K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, Learning phrase representations using RNN encoder-decoder for statistical machine translation, in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (Association for Computational Linguistics, Doha, Qatar, 2014), pp. 1724–1734.
- [37] R. Pascanu, T. Mikolov, and Y. Bengio, On the difficulty of training recurrent neural networks, in *Proceedings of the 30th International Conference on Machine Learning*, edited by S. Dasgupta and D. McAllester (PMLR, Atlanta, GA, 2013), Vol. 28, pp. 1310–1318.
- [38] F. Metz and M. Bukov, Self-correcting quantum many-body control using reinforcement learning with tensor networks, *Nat. Mach. Intell.* **5**, 780 (2023).
- [39] Y.-H. Zhang and M. Di Ventura, Transformer quantum state: A multipurpose model for quantum many-body problems, *Phys. Rev. B* **107**, 075147 (2023).
- [40] M. Kolodrubetz, D. Sels, P. Mehta, and A. Polkovnikov, Geometry and non-adiabatic response in quantum and classical systems, *Phys. Rep.* **697**, 1 (2017).
- [41] D. Guéry-Odelin, A. Ruschhaupt, A. Kiely, E. Torrontegui, S. Martínez-Garaot, and J. G. Muga, Shortcuts to adiabaticity: Concepts, methods, and applications, *Rev. Mod. Phys.* **91**, 045001 (2019).
- [42] H. Kim and D. A. Huse, Ballistic spreading of entanglement in a diffusive nonintegrable system, *Phys. Rev. Lett.* **111**, 127205 (2013).
- [43] J. F. Rodríguez-Nieva, C. Jonay, and V. Khemani, Quantifying quantum chaos through microcanonical distributions of entanglement, *Phys. Rev. X* **14**, 031014 (2024).
- [44] Q. Zheng, A. Zhang, and A. Grover, Online decision transformer, in *Proceedings of Machine Learning Research* (Baltimore, Maryland, USA, 2022), pp. 27042–27059.
- [45] L. E. H. Rodríguez and A. A. Kananenka, A short trajectory is all you need: A transformer-based model for long-time dissipative quantum dynamics, *J. Chem. Phys.* **161**, 171101 (2024).
- [46] <http://dx.doi.org/10.5281/zenodo.22558>.
- [47] Code for the network architecture and data generation, GitHub (2025), <https://github.com/pranavjv/transformerfeedbackcontrol>.