Machine Learning Meets Quantum State Preparation. The Phase Diagram of Quantum Control

Marin Bukov,[1,*] Alexandre G.R. Day,[1,†] Dries Sels,[1,2] Phillip Weinberg,[1] Anatoli Polkovnikov,[1] and Pankaj Mehta[1]

[1] *Department of Physics, Boston University, 590 Commonwealth Ave., Boston, MA 02215, USA*
[2] *Theory of quantum and complex systems, Universiteit Antwerpen, B-2610 Antwerpen, Belgium*

(Dated: May 2, 2017)

The ability to prepare a physical system in a desired quantum state is central to many areas of physics such as nuclear magnetic resonance, cold atoms, and quantum computing. However, preparing a quantum state quickly and with high fidelity remains a formidable challenge. Here we tackle this problem by applying cutting edge Machine Learning (ML) techniques, including Reinforcement Learning, to find short, high-fidelity driving protocols from an initial to a target state in complex many-body quantum systems of interacting qubits. We show that the optimization problem undergoes a spin-glass like phase transition in the space of protocols as a function of the protocol duration, indicating that the optimal solution may be exponentially difficult to find. However, ML allows us to identify a simple, robust variational protocol, which yields nearly optimal fidelity even in the glassy phase. Our study highlights how ML offers new tools for understanding nonequilibrium physics.

Reliably targeting specific quantum states of complex physical systems is a key issue in quantum mechanics. Moreover, quantum state manipulation is vital for nuclear magnetic resonance experiments [1], cold atomic systems [2, 3], trapped ions [4–6], quantum optics [7], superconducting qubits [8], nitrogen vacancy centers [9], and quantum computing [10]. However, our theoretical understanding of quantum state preparation remains limited, since the latter is intrinsically a nonequilibrium dynamical process. This severely restricts the applicability of analytical techniques and simulating the dynamics of large quantum many-body systems remains a formidable challenge. In this paper, we adopt a radically different approach to this problem based on machine learning (ML) [11–14]. ML has recently been applied successfully to several problems in equilibrium condensed matter physics [15, 16] and turbulent dynamics [17], and here we show that it provides deep insights into nonequilibrium quantum dynamics.

Many studies on quantum state preparation rely on the adiabatic theorem. It states that, by dynamically changing the external parameters of a Hamiltonian sufficiently slowly, it is possible to smoothly deform an eigenstate of the initial Hamiltonian to an eigenstate of the final Hamiltonian. The amount of time required for this process is set by the minimum energy gap and can become extremely long for many-body systems with small energy gaps in the spectrum, which poses fundamental limitations on adiabaticity-based approaches and, in some cases, the adiabatic limit may not even exist [18, 19]. This has motivated numerous alternative approaches to quantum state manipulation [20–34].

Despite these advances, surprisingly little is known about how to successfully load an interacting quantum system into a desired target state, especially at short times, or even when this is feasible in the first place. It was shown that, for single particles, there exists a "quantum speed limit" for how fast one can prepare a target state [2, 35]. However, if and how these results generalize to generic many-body quantum systems remains an open problem. Here we apply Reinforcement Learning [11, 13, 14] to find short, high-fidelity, driving protocols from an initial to a target state in multiple quantum systems – including many-body systems of strongly interacting qubits. A clear advantage of ML over more traditional approaches is that it is largely free from the usual biases developed by our intuition. In the case of quantum state preparation, ML allowed us to identify both complexity and simplicity in our optimization problem. Complexity manifests in the presence of a glassy landscape of high-fidelity protocols. Simplicity shows up in the existence of very simple, robust, ML-motivated variational protocols which give nearly optimal fidelity. Thus, ML techniques can guide us towards new, previously unexplored directions by adjusting our physical intuition.

Constrained Qubit Manipulation.—To illustrate the problem, consider a two-level system described by the Hamiltonian

$$H(t) = -S^z - h_x(t)S^x, \tag{1}$$

where S^α, are the spin-1/2 operators. This Hamiltonian comprises integrable many-body and free translational invariant systems, such as the transverse-field Ising model, graphene and topological insulators.

Reinforcement Learning (RL) is a branch of ML, where a computer agent learns how to perform and master a specific task, e.g. to prepare a specified target state, through feedback based on interactions with its environment. More specifically, we use modified versions of the Watkins Q-learning algorithm [11] to teach a computer to find the best driving protocol to prepare a quantum system in a target state $|\psi_*\rangle$ starting from an initial state $|\psi_i\rangle$ by controlling a time-dependent field $h_x(t)$. The agent can construct piecewise constant protocols by

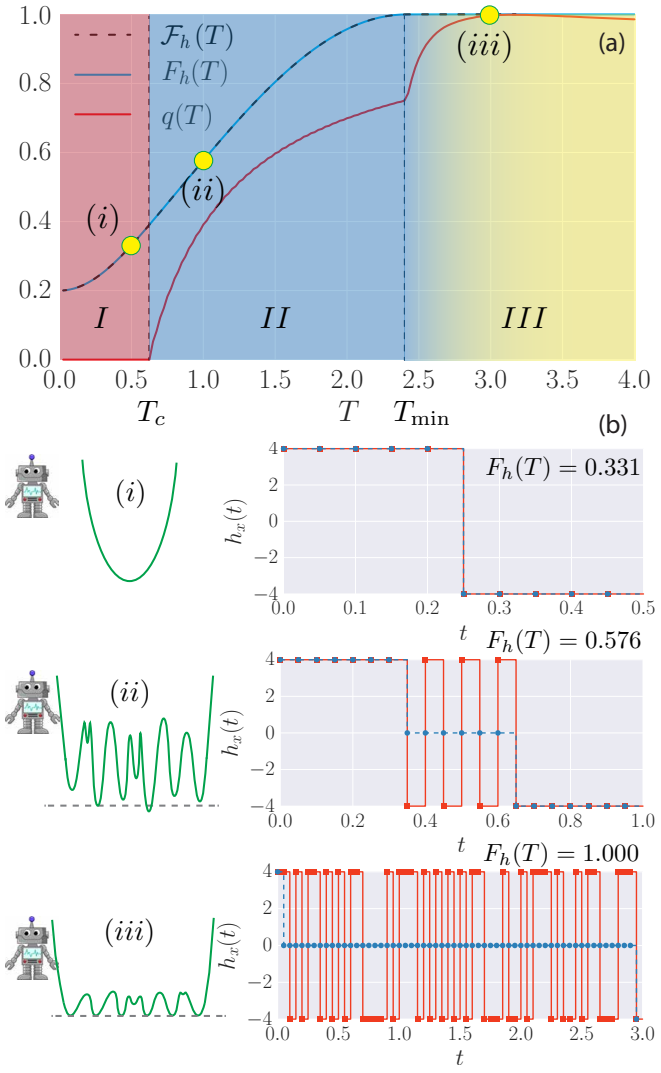


FIG. 1: (a) Phase diagram of the quantum state preparation problem for the qubit in Eq. (1) vs. protocol duration T, as determined by the order parameter $q(T)$ (red) and the maximum possible achievable fidelity $F_h(T)$ (blue), compared to the variational fidelity $\mathcal{F}_h(T)$ (black, dashed). Increasing the total protocol time T, we go from an overconstrained phase I, through a glassy phase II, to an underconstrained phase III. (b) Left: the (negative) fidelity landscape is shown schematically (green). Right: the optimal bang-bang protocol found by the RL agent at the points (i)–(iii) (red) and the variational protocol [36] (blue, dashed).

choosing the field step size δh_x at each time step δt from a set of allowed values. In order to make the agent learn, it is offered a reward for every protocol it constructs – the fidelity $F_h(T) = |\langle\psi_*|\psi(T)\rangle|^2$ for being in the target state after a fixed time T following the protocol $h_x(t)$. The goal of the agent is to maximize the total reward in a series of attempts. Starting with no prior knowledge about the physical system whatsoever, the agent keeps information about already tried protocols, based on which it constructs new, improved protocols through a sophisticated biased sampling algorithm (see Movie). We benchmark the results of RL using Stochastic Descent (SD) [36].

The initial $|\psi_i\rangle$ and target $|\psi_*\rangle$ states are chosen as the ground states of (1) at $h_x = -2$ and $h_x = 2$, respectively. Generically, one does not have access to infinite control fields; here we restrict to field $h_x(t) \in [-4, 4]$, see Fig. 1b. Pontryagin's maximum principle allows us to focus on bang-bang protocols (red), where $h_x(t) \in \{\pm 4\}$, although we verified that the method works also for quasi-continuous protocols with many steps δh_x [36]. The space of available protocols grows exponentially with the inverse step size δt^{-1}, resulting in the generic complexity of the optimization problem.

The control problem for the constrained qubit exhibits three distinct phases as a function of the ramp time T. If T is greater than a minimum time $T_{\min} \approx 2.4$ [set by the energy scales of the problem] known as the quantum speed limit [35], one can construct infinitely many protocols which prepare the target state with unit fidelity. This is a distinct feature of the two-level system where a single control field is sufficient to effectively cover the full Bloch sphere beyond the minimum time. The red line in Fig. 1b (iii) shows an optimal protocol of unit fidelity found by the agent for a total ramp time of the order of the minimum gap, whose Bloch sphere representation can be found in Fig. 2a (magenta line), c.f. Movie-3.

We find that for $T < T_{\min}$, there exists no protocol to prepare the target state with unit fidelity, and the problem of determining the best achievable fidelity becomes a challenge. In this regime, to the best of our knowledge, there exists no efficient method to construct the optimal driving protocol, and thus we choose to rely on ML algorithms. Figure 1a shows the best bang-bang protocol found by the agent (see Movie-2 and [36] for best protocol with quasi-continuous actions). This protocol has a remarkable feature: without any prior knowledge about the intermediate quantum state nor its Bloch sphere representation, the RL agent discovers that it is advantageous to first bring the state to the equator – which is a geodesic – and then effectively turns off the field $h_x(t)$, to enable the fastest possible precession about the $\hat{z}$-axis. After staying at the equator for as long as optimal, the agent rotates as fast as it can to bring the state as close as possible to the target, thus optimizing the final fidelity for the available protocol duration.

Decreasing the total ramp time T further, we find a second critical time $T_c \approx 0.6$, below which there exists a unique optimal protocol, even though the achievable fidelity can be quite limited (Movie-1), see Fig. 1b (i) and Fig. 2a.

Glassy Landscape.—The state preparation problem has a natural interpretation as an optimization problem. Fixing the total ramp time T, let us assign to each driving

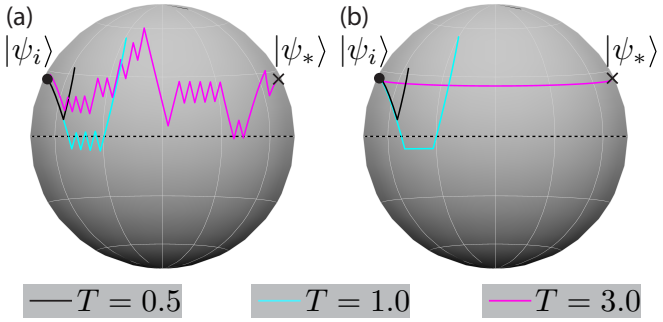


FIG. 2: Bloch sphere representation of the bang-bang protocols found by the RL agent (a) and variational protocols (b) from Fig. 1b. See also Movies 1–6.

protocol $h_x(t)$ a single number: the (negative) fidelity, $-F_h(T)$, obtained following a unitary evolution under the Hamiltonian (1). One can then consider $-F_h(T)$ as a potential landscape, the global minimum of which corresponds to the best driving protocol. The exploration problem of finding the optimal protocol then becomes reminiscent of finding the ground state configuration of classical spin model. To map out the landscape of local fidelity minima, $\{h_x^\alpha(t)\}$, we use SD, starting from a random bang-bang protocol configuration [36]. To study the correlations between the fidelity minima as a function of the total ramp time T, we define the order parameter $q(T)$, closely related to the Edwards-Anderson order parameter [37, 38], as

$$q(T) = \frac{1}{16 N_T} \overline{\sum_{t=1}^{N_T} \{h_x(t) - \overline{h_x}(t)\}^2}, \tag{2}$$

where $\overline{h_x}(t) = N_{\rm real}^{-1} \sum_{\alpha=1}^{N_{\rm real}} h_x^\alpha(t)$ is the sample-averaged protocol. If the minima $\{h_x^\alpha(t)\}_{\alpha=1}^{N_{\rm real}}$ are all uncorrelated, then $\overline{h_x}(t) \equiv 0$, and thus $q(T) = 1$. On the other hand, if the fidelity landscape contains only one minimum, then $\overline{h_x}(t) \equiv h_x(t)$ and $q(T) = 0$. The behaviour of $q(T)$, the maximum fidelity $F_h(T)$ together with a qualitative description of the corresponding fidelity landscapes are shown in Fig. 1.

For $T < T_c$, $q(T) = 0$ and the problem has a unique solution, suggesting that the fidelity landscape is convex. This overconstrained phase is labelled I in the phase diagram (Fig. 1a). At the critical time T_c the spin has enough time to reach the equator for the first time, and the system enters the glassy regime II. Since the state precession speed towards the equator depends on the maximum possible allowed field strength in the $\hat{x}$-direction, it follows that $T_c \to 0$ as $h_x \to \infty$. For $T > T_c$, the minima of the fidelity landscape proliferate, and the control problem undergoes a second-order transition to a glassy phase with many almost degenerate local minima, reflected by the order parameter $0 < q(T) < 1$. The minimization problem becomes non-convex, and a sophisticated algorithm is required to guarantee finding the global minimum. At $T = T_{\rm min}$, $q(T)$ exhibits a non-analyticity and the system enters the under-constrained phase III. In this phase, there is a proliferation of exactly degenerate, uncorrelated global minima, corresponding to protocols of unit fidelity, and the optimization task becomes easy again.

Many Coupled Qubits.—Our results raise the natural question of how difficult state preparation is in more complex quantum models, since there exists no theory for constructing optimal ramp protocols $h_x(t)$ in non-integrable many-body systems. To the end, consider a chain of L coupled qubits, which can be experimentally realized with superconducting qubits [8], cold atoms [39] and trapped ions [6]:

$$H(t) = -\sum_{j=1}^{L} \left(S_{j+1}^z S_j^z + h_z S_j^z + h_x(t) S_j^x \right). \tag{3}$$

We use periodic boundary conditions and restrict ourselves to the zero-momentum sector of positive parity, which contains the ground state. We set $h_z = 1$ to avoid the anti-ferromagnet to paramagnet quantum phase transition at $h_z = 0$, and choose the paramagnetic ground states of Eq. (3) at fields $h_x = -2$ and $h_x = 2$ for the initial and target state, respectively. The details of the control field $h_x(t)$ are the same as in the single qubit case, and we use the *many-body* fidelity as a measure of performance.

Figure 3 shows the phase diagram of the coupled qubits model. First, notice that while the overconstrained-to-glass critical point T_c survives, the quantum speed limit critical point $T_{\rm min}$ is (if existent at all) outside the short ramp times range of interest. Thus, the glassy phase extends over to long and probably infinite ramp times. Therefore, preparation of many-body states with high-fidelity remains a hard problem even for long protocols. Second, observe that, even though unit fidelity is no longer achievable, there exist nearly optimal protocols with extremely high many-body fidelity at short ramp times. This fact is striking because the Hilbert space of our system grows *exponentially* with L and we are using only one control field to manipulate exponentially many degrees of freedom. Furthermore, generically, we expect the *many-body* overlap between two states to be exponentially small in the system size L. Another remarkable feature of the optimal solution is that for the system sizes $L \geq 6$ both $q(T)$ and $-1/L \log F_h(T)$ converge to their thermodynamic limit with no visible finite size corrections [36]. This is related to the Lieb-Robinson bound for information propagation which suggests that information should spread over approximately $JT = 4$ sites for the ramp times considered.

Variational Theory.—An additional feature of the optimal bang-bang solution found by the agent is that the entanglement entropy of the half system generated dur-

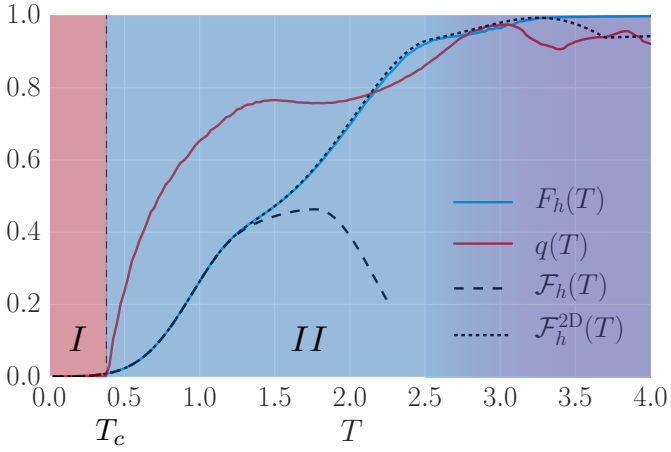


FIG. 3: Phase diagram of the many-body quantum state preparation problem. The order parameter (red) shows a kink at the critical time $T_c \approx 0.4$ when a phase transition occurs from an overconstrained phase (I) to a glassy phase (II). The best fidelity $F_h(T)$ (blue) is compared to the variational fidelity $\mathcal{F}_h(T)$ (dashed) and the 2D-variational fidelity $\mathcal{F}_h^{\rm 2D}(T)$ (dotted) [36].

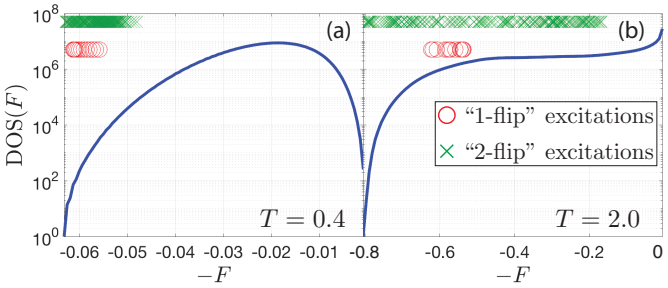


FIG. 4: Density of states (protocols) in the overconstrained phase at $T = 0.4$ (a) and the glassy phase at $T = 2.0$ (b) as a function of the fidelity F. The red circles and the green crosses show the fidelity of the "1-spin" flip and "2-spin" flip excitation protocols above the absolute ground state (i.e. the optimal protocol). The system size is $L = 6$ and each protocol has $N_T = 28$ bangs.

ing the evolution always remains small, satisfying an area law [36]. This implies that the system likely follows the ground state of some local, yet a-priori unknown, Hamiltonian [40]. This motivated us to use the best protocols found by ML to construct simple variational protocols consisting of a few bangs.

For the single qubit example, these variational protocols are shown in Fig. 1b (dashed blue lines) and represented on the Bloch sphere in Fig. 2b (see also Movies 4-6). The variational fidelity $\mathcal{F}_h(T)$ agrees nearly perfectly with the optimal fidelity $F_h(T)$ obtained using ML techniques, cf. Fig. 1a. We further demonstrate that our variational theory fully captures the physics of the two critical points T_c and $T_{\rm min}$ [36]. In the many-body case, the same one-parameter variational ansatz only describes the behaviour in the overconstrained phase, cf. Fig. 3 (dashed line), up to and including the critical point T_c, and fails for $T > T_c$. Nevertheless, a slightly modified, two-parameter variational ansatz, motivated again by the solutions found by the ML agent (see Movie), appears to be fully sufficient to capture the essential features of the optimal protocol much deeper into the glassy phase, as shown by the $\mathcal{F}_h^{\rm 2D}(T)$ curve in Fig. 3. This many-body variational theory features an additional pulse, reminiscent of spin-echo, which appears to control and suppress the generation of entanglement entropy during the drive [36]. Indeed, while the two-parameter ansatz is strictly better than the single-parameter protocol for all $T > T_c$, the difference between the two grows slowly as a function of time. It is only at a later time, $T \approx 1.3$, that the effect of the second pulse really kicks in, and we observe the largest entanglement in the system for the optimal protocol.

Glassy Behaviour.—It is shocking that the dynamics of a non-integrable many-body quantum system, associated with the optimal protocol, is so efficiently captured by such a simple, two-parameter variational protocol, even in the regimes where there is no obvious small parameter and where spin-spin interactions play a significant role. Upon closer comparison of the variational and the optimal fidelities, one can find regions in the glassy phase where the simple variational protocol outperforms the numerical 'best' fidelity, cf. Fig. 3.

To better understand this behavior, we choose a grid of $N_T = 28$ equally-spaced time steps, and compute all 2^{28} bang-bang protocols and their fidelities. The corresponding density of states (DOS) in fidelity space is shown in Fig. 4 for two choices of T in the overconstrained and glassy phase. This allows us to unambiguously determine the ground state of the fidelity landscape (i.e. the optimal protocol). Starting from this ground state, we then construct all excitations generated by local in time flips of the bangs of the optimal protocol. The fidelity of the "1-flip" excitations is shown using red circles in Fig. 4. Notice how, in the glassy phase, these 28 excitations have relatively low fidelities compared to the ground state and are surrounded by $\sim 10^6$ other states. This has profound consequences: as we are 'cooling' down in the glassy phase, searching for the optimal protocol and coming from a state high up in the fidelity landscape, if we miss one of the 28 elementary excitations, it becomes virtually impossible to reach the global ground state and the situation becomes much worse if we increase the system size. On the contrary, in the overconstrained phase, the smaller value of the DOS at the "1-flip" excitation ($\sim 10^2$) enables easily reaching the ground state.

The green crosses in Fig. 4 show the fidelity of the "2-flip" excitations. By the above argument, a "2-flip" algorithm would not see the phase as a glass for $T \lesssim$

2.5, yet it does so for $T \gtrsim 2.5$, marked by the different shading in Fig. 3. Correlated with this observation, we find a signature of a transition also in the improved two-parameter variational theory in the glassy phase [36]. In general, we expect the glassy phase to exhibit a series of phase transitions, reminiscent of the random k-SAT problems [41, 42].

Outlook.—The appearance of a glassy phase, which dominates the many-body physics, in the space of protocols of the quantum state preparation problem, has far-reaching consequences for condensed matter physics. Quantum computing relies heavily on our ability to prepare states with high fidelity, yet finding high efficiency state manipulation routines remains a formidable challenge. Highly controllable quantum emulators, such as ultracold atoms and ions, depend almost entirely on the feasibility to reach the correct target state, before it can be studied. We demonstrated how, by closely examining the protocols found by a computer agent, one can construct variational theories which capture almost all relevant features of the dynamics generated by the optimal protocol. Unlike the optimal bang-bang protocol, the simpler variational protocol is robust to small perturbations, while giving comparable fidelities. This implies the existence of nearly optimal protocols, which do not suffer from the exponential complexity of finding the global minimum of the entire optimisation landscape. We believe that, a common pattern between such effective theories can be revealed with time, which could help form the underlying principles for a theory of statistical physics away from equilibrium.

The existence of phase transitions in quantum control problems has profound consequences beyond physical systems. It is an open question whether this behavior is generic to other control problems. It is our hope that given the close connections between optimal control and RL, the physical interpretation of optimization problems in terms of a glassy phase will help advance our quest for better artificial intelligence.

Acknowledgements.—We thank J. Garrahan, M. Heyl, M. Schiró and D. Schuster for illuminating discussions. MB, PW and AP were supported by NSF DMR-1506340, ARO W911NF1410540 and AFOSR FA9550-16-1-0334. AD is supported by a NSERC PGS D. AD and PM acknowledge support from Simon's Foundation through the MMLS Fellow program. DS acknowledges support from the FWO as post-doctoral fellow of the Research Foundation – Flanders and CMTV. We used QuSpin for simulating the dynamics of the qubit systems [43]. The authors are pleased to acknowledge that the computational work reported on in this paper was performed on the Shared Computing Cluster which is administered by Boston University's Research Computing Services. The authors also acknowledge the Research Computing Services group for providing consulting support which has contributed to the results reported within this paper.

* Electronic address: mbukov@bu.edu
† Electronic address: agrday@bu.edu

[1] L. M. K. Vandersypen and I. L. Chuang, Rev. Mod. Phys. **76**, 1037 (2005).
[2] S. van Frank, M. Bonneau, J. Schmiedmayer, S. Hild, C. Gross, M. Cheneau, I. Bloch, T. Pichler, A. Negretti, T. Calarco, *et al.*, Scientific reports **6** (2016).
[3] P. B. Wigley, P. J. Everitt, A. van den Hengel, J. Bastian, M. A. Sooriyabandara, G. D. McDonald, K. S. Hardman, C. Quinlivan, P. Manju, C. C. Kuhn, *et al.*, Scientific reports **6** (2016).
[4] R. Islam, E. E. Edwards, K. Kim, S. Korenblit, C. Noh, H. Carmichael, G.-D. Lin, L.-M. Duan, C.-C. Joseph Wang, J. K. Freericks, and C. Monroe, Nature Communications **2**, 377 EP (2011), article.
[5] C. Senko, P. Richerme, J. Smith, A. Lee, I. Cohen, A. Retzker, and C. Monroe, Phys. Rev. X **5**, 021026 (2015).
[6] P. Jurcevic, B. P. Lanyon, P. Hauke, C. Hempel, P. Zoller, R. Blatt, and C. F. Roos, Nature **511**, 202 (2014), letter.
[7] C. Sayrin, I. Dotsenko, X. Zhou, B. Peaudecerf, T. Rybarczyk, S. Gleyzes, P. Rouchon, M. Mirrahimi, H. Amini, M. Brune, J.-M. Raimond, and S. Haroche, Nature **477**, 73 (2011).
[8] R. Barends, A. Shabani, L. Lamata, J. Kelly, A. Mezzacapo, U. Las Heras, R. Babbush, A. Fowler, B. Campbell, Y. Chen, *et al.*, Nature **534**, 222 (2016).
[9] B. B. Zhou, A. Baksic, H. Ribeiro, C. G. Yale, F. J. Heremans, P. C. Jerger, A. Auer, G. Burkard, A. A. Clerk, and D. D. Awschalom, Nat Phys **13**, 330 (2017), letter.
[10] M. A. Nielsen and I. Chuang, *Quantum computation and quantum information* (AAPT, 2002).
[11] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction* (MIT Press, Cambridge, MA, 2017).
[12] C. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics), 1st edn. 2006. corr. 2nd printing edn* (2007).
[13] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis, Nature **518**, 529 (2015), letter.
[14] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, and D. Hassabis, Nature **529**, 484 (2016), article.
[15] G. Carleo and M. Troyer, Science **355**, 602 (2017).
[16] J. Carrasquilla and R. G. Melko, Nat Phys **advance online publication** (2017), letter.
[17] G. Reddy, A. Celani, T. J. Sejnowski, and M. Vergassola, Proceedings of the National Academy of Sciences , 201606075 (2016).
[18] V. Khemani, R. Nandkishore, and S. L. Sondhi, Nature Physics **11**, 560 (2015).
[19] P. Weinberg, M. Bukov, L. D'Alessio, A. Polkovnikov, S. Vajna, and M. Kolodrubetz, arXiv , 1606.02229 (2016).
[20] A. Baksic, H. Ribeiro, and A. A. Clerk, Phys. Rev. Lett.

116, 230503 (2016).
[21] X. Wang, M. Allegra, K. Jacobs, S. Lloyd, C. Lupo, and M. Mohseni, Phys. Rev. Lett. **114**, 170501 (2015).
[22] R. R. Agundez, C. D. Hill, L. C. L. Hollenberg, S. Rogge, and M. Blaauboer, Phys. Rev. A **95**, 012317 (2017).
[23] S. Bao, S. Kleer, R. Wang, and A. Rahmani, arXiv , 1704.01423 (2017).
[24] G. M. Rotskoff, G. E. Crooks, and E. Vanden-Eijnden, Phys. Rev. E **95**, 012148 (2017).
[25] N. Leung, M. Abdelhafez, J. Koch, and D. I. Schuster, arXiv , arXiv:1612.04929v2 (2016).
[26] Z.-C. Yang, A. Rahmani, A. Shabani, H. Neven, and C. Chamon, arXiv , arXiv:1607.06473v2 (2016).
[27] C. Jarzynski, Phys. Rev. A **88**, 040101 (2013).
[28] M. Kolodrubetz, P. Mehta, and A. Polkovnikov, arXiv , 1602.01062v3 (2016).
[29] D. Sels and A. Polkovnikov, arXiv , 1607.05687 (2016).
[30] S. J. Glaser, T. Schulte-Herbrüggen, M. Sieveking, O. Schedletzky, N. C. Nielsen, O. W. Sørensen, and C. Griesinger, Science **280**, 421 (1998).
[31] H. Rabitz, R. de Vivie-Riedle, M. Motzkus, and K. Kompa, Science **288**, 824 (2000).
[32] N. Khaneja, R. Brockett, and S. J. Glaser, Phys. Rev. A **63**, 032308 (2001).
[33] S. E. Sklarz and D. J. Tannor, Phys. Rev. A **66**, 053619 (2002).
[34] N. Khaneja, T. Reiss, C. Kehlet, T. Schulte-Herbrggen, and S. J. Glaser, Journal of Magnetic Resonance **172**, 296 (2005).
[35] G. C. Hegerfeldt, Phys. Rev. Lett. **111**, 260501 (2013).
[36] See Supplemental Material.
[37] T. Castellani and A. Cavagna, Journal of Statistical Mechanics: Theory and Experiment **2005**, P05012 (2005).
[38] L. O. Hedges, R. L. Jack, J. P. Garrahan, and D. Chandler, Science **323**, 1309 (2009).
[39] J. Simon, W. S. Bakr, R. Ma, M. E. Tai, P. M. Preiss, and M. Greiner, Nature **472**, 307 (2011).
[40] L. Vidmar, D. Iyer, and M. Rigol, arXiv , arXiv:1512.05373v3 (2015).
[41] M. Mézard, G. Parisi, and R. Zecchina, Science **297**, 812 (2002).
[42] S. C. Morampudi, B. Hsu, S. L. Sondhi, R. Moessner, and C. R. Laumann, arXiv , arXiv:1704.00238 (2017).
[43] P. Weinberg and M. Bukov, SciPost Phys. **2**, 003 (2017).
[44] M. Tomka, T. Souza, S. Rosenberg, and A. Polkovnikov, arXiv , arXiv:1606.05890 (2017).
[45] C. Jarzynski, S. Deffner, A. Patra, and Y. Subasi, Phys. Rev. E **95**, 032122 (2017).

SUPPLEMENTAL MATERIAL

DETAILS OF THE NUMERICAL ALGORITHMS

Reinforcement Learning

Reinforcement Learning (RL) is a subfield of Machine Learning (ML) where a computer agent is taught to perform and master a specific task by performing a series of actions in order to maximize a reward function (in our case, the final fidelity). As already mentioned in the main text, we use a modified version of Watkins online Q-learning algorithm with linear function approximation and eligibility traces [11] to teach a RL agent to find protocols of optimal fidelity. Below, we briefly summarize the details of the procedure. For a detailed description of the algorithm, we refer the reader to Ref. [11]. A Python implementation of the algorithm is available on Github.

The fidelity optimization problem is defined as an episodic, undiscounted Reinforcement Learning task. Each episode takes a fixed number of steps $N_T = T/\delta t$, where T is the total ramp time, and δt is the physical time step (protocol time step). We define the state $\mathcal{S}$, action $\mathcal{A}$ and reward $\mathcal{R}$ spaces, respectively, as

$$\mathcal{S} = \{s = (t, h_x(t))\}, \qquad \mathcal{A} = \{a = \delta h_x\}, \qquad \mathcal{R} = \{r \in [0, 1]\}. \tag{4}$$

The state space consists of all tuples $(t, h_x(t))$ of time t with the corresponding magnetic field $h_x(t)$. Notice that no information about the quantum state whatsoever is encoded in the RL state. Thus, the RL agent is able to learn circumventing the difficulties associated with the notions of quantum physics. Including time t to the state is not a common procedure in Q-learning, but is required in order for the agent to be able to estimate how far away it is from the episode's end.

The action space $\mathcal{A}$ consists of all jumps in the protocol $h_x(t)$, denoted by δh_x. Thus, protocols are constructed as piecewise-constant functions. We restrict the available actions of the RL agent such that at all times the field $h_x(t)$ is in the interval $[-4, 4]$. The bang-bang and quasi-continuous protocols discussed in the next section are examples of the family of protocol functions we allow in the simulation. Pontryagin's maximum principle in Optimal Control applied to our problem postulates that, given any protocol of certain fidelity, there exists a bang-bang protocol of the same fidelity. Thus, allowing for bang-bang protocols, our study is entitled to capture the optimal solution as well. Generally, this follows from the Trotter-Suzuki decomposition, which implies that any time evolution, not only optimal, can be approximated by a series of bangs with an arbitrary accuracy.

Last but not least, the reward space $\mathcal{R}$ is the space of all real numbers in the interval $[0, 1]$. The rewards for the agent are given only at the end of each episode, according to:

$$r(t) = \begin{cases} 0, & \text{if } t < T \\ |\langle\psi_*|\psi(T)\rangle|^2, & \text{if } t = T \end{cases} \tag{5}$$

This reflects the fact that we are not interested in which quantum state the physical system is in during the ramp; all that matters is that we maximize the final fidelity.

An essential part of the RL problem is to define the environment, with which the agent interacts in order to learn. We choose this to consist of the Schrödinger initial value problem, together with the target state:

$$\text{Environment} = \{i\partial_t|\psi(t)\rangle = H(t)|\psi(t)\rangle, \quad |\psi(0)\rangle = |\psi_i\rangle, \quad |\psi_*\rangle\}, \tag{6}$$

where $H(t)$ is the Hamiltonian whose time dependence is defined through the magnetic filed $h_x(t)$ which the agent is constructing during the episode via online Q-learning updates.

Let us now briefly illustrate the protocol construction algorithm: for instance, if we start in the initial RL state $s_0 = (t = 0, h = -4)$, and take the action $a = \delta h_x = 8$, we go to the next RL state $s_1 = (\delta t, +4)$. As a result of the interaction with the environment, the initial quantum state is evolved forward in time for one time step (from time $t_0 = 0$ to time $t_1 = \delta t$) with the constant Hamiltonian $H(h = 4)$: $|\psi(\delta t)\rangle = \mathrm{e}^{-iH(h=4)\delta t}|\psi_i\rangle$. After each step we compute the local reward according to Eq. (5), and update the Q-function, even though the instantaneous reward at that step might be zero. This procedure is repeated until the end of the episode is reached at $t = T$. In general, one can imagine this process as a state-action-reward chain

$$s_0 \to a_0 \to r_0 \to s_1 \to a_1 \to r_1 \to s_2 \to \cdots \to s_{N_T}.$$

The total return for a state-action reward chain is defined as the sum of all local rewards during the episode: $R = \sum_{i=0}^{N_T} r_i$. The central object in Watkins Q-Learning is the $Q(s, a)$ function which is given by the expected return at

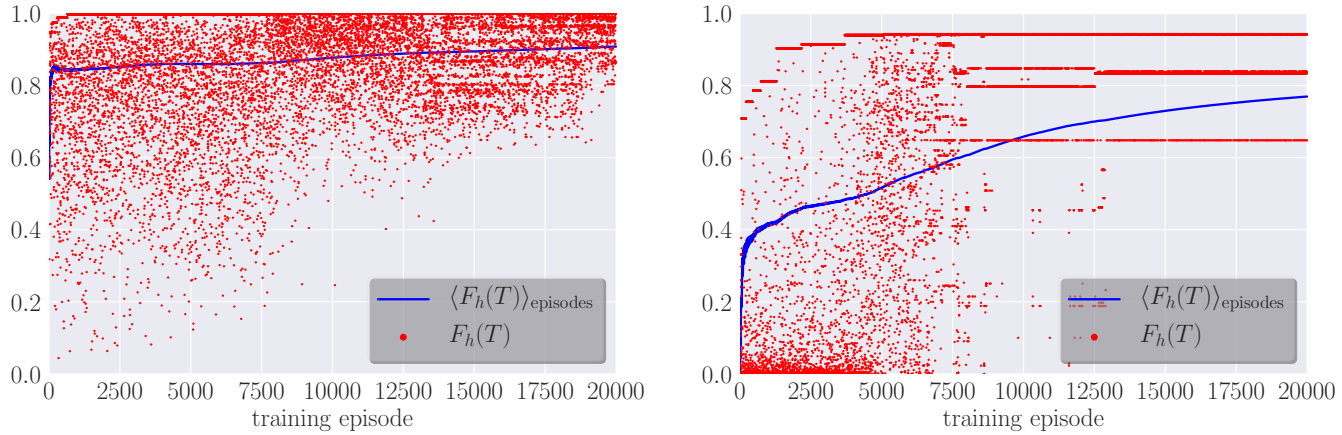


FIG. 5: Learning curves of the RL agent for $L = 1$ at $T = 2.4$ (left) [see Movie] and $L = 10$ at $T = 3.0$ (right) [see Movie]. The red dots show the instantaneous reward (i.e. fidelity) at every episode, while the blue line the cumulative episode-average. The ramp-up of the RL temperature β_{RL} suppresses exploration over time which leads to a gradually increasing average fidelity. The time step is $\delta t = 0.05$.

the end of each episode for starting in a state s and taking the action a. It satisfies the Bellman optimality equation, the solution of which cannot be obtained in a closed form. The usual way of solving the Bellman equation numerically is to make use of temporal difference learning, which results in the following Q-Learning update rule [11]

$$Q(s_i, a_i) \longleftarrow Q(s_i, a_i) + \alpha \left[r_i + \max_a Q(s_{i+1}, a) - Q(s_i, a_i) \right], \tag{7}$$

where the learning rate $\alpha \in (0, 1)$. Whenever $\alpha \approx 1$, the convergence of the update rule (7) can be slowed down or even precluded, in cases where the error $\delta_t = r_i + \max_a Q(s_{i+1}, a) - Q(s_i, a_i)$ becomes significant. On the contrary, $\alpha \approx 0$ corresponds to very slow learning. Thus, the optimal value for the learning rate lies in between, and is determined empirically for the problem under consideration.

To allow for the efficient implementation of nearly continuous drives (used to generate the quasi-continuous protocols), we employ a linear function approximation to the Q-function, using equally-spaced tilings along the entire range of $h_x(t) \in [-4, 4]$ [11]. This allows the RL agent to generalize, i.e. gain information about the fidelity of some protocols not yet encountered.

We then iterate the algorithm for 2×10^4 episodes. The exploration-exploitation dilemma [11] requires a fair amount of exploration, in order to ensure that the agent visits large parts of the RL state space which prevents it from getting stuck in a local minimum from the beginning. Too much exploration, and the agent will not be able to learn. On the other hand, no exploration whatsoever guarantees that the agent will repeat deterministically a given strategy, though it will be unclear whether there exists a better, yet unseen one. In the longer run, we cannot preclude the agent from ending up in a local minimum of reward space. In such cases, we run the algorithm multiple times starting from a random initial condition, and post-select the outcome.

Due to extremely large state space, we employed a novel form of replay to ensure that our RL algorithm could learn from the high fidelity paths it encountered. Our replay algorithm alternates between two different ways of training our RL agent which we call training phases: an "exploratory" training-phase where the RL agent exploits the current Q-function to explore, and a "replay" training-phase where we replay the best encountered protocol. This novel form of replay, to the best of our knowledge, has not been used previously. In the exploratory training-phase, which lasts 40 episodes, the agent takes actions according to a softmax probability distribution based on the instantaneous values of the Q-function. In other words, at each time step, the RL agent looks up the instantaneous values $Q(s, :)$ corresponding to all available actions, and computes a probability for each action: $P(a) \sim \exp(\beta_{\text{RL}} Q(s, a))$. The amount of exploration is set by β_{RL}, with $\beta_{\text{RL}} = 0$ corresponding to random actions and $\beta_{\text{RL}} = \infty$ corresponding to always taking greedy actions with respect to the current Q-function. Here we use an external 'learning' temperature scale, the inverse of which, β_{RL}, is linearly ramped down as the number of episodes progresses. In the replay training-phase, which is also 40 episodes, we replay the best-encountered protocol up to the given episode. Through this procedure, when the next exploratory training-phase begins again, the agent is biased to do variations on top of the best-encountered protocol, effectively improving it, until it reaches a reasonably good fidelity. Two learning curves of

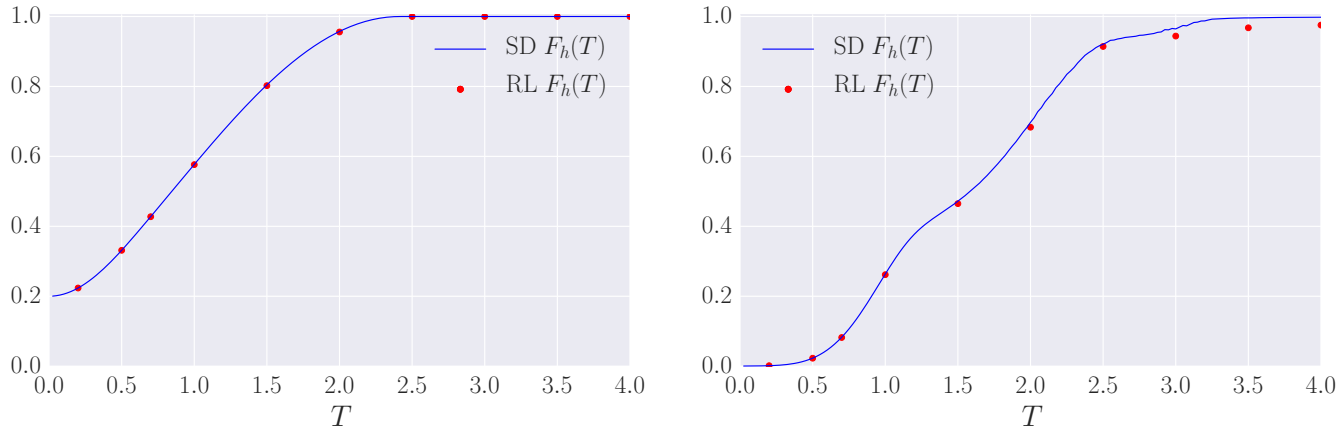


FIG. 6: Comparison between SD (solid lines) and RL (red dots) best fidelities for $L = 1$ (left) and $L = 10$ (right) shows the robustness of the numerical results.

the RL agent are shown in Fig. 5.

Stochastic Descent

To benchmark the results obtained with the RL algorithm, we use a greedy stochastic descent (SD) to sample the fidelity landscape minima with respect to the ramping protocols. We restrict our SD algorithm to exploring bang-bang protocols where $h_x(t) \in \{\pm 4\}$. The algorithm starts from a random protocol and proposes local field updates at a time t chosen uniformly in the interval $[0, T]$. The updates consist in changing the applied field $h_x(t) \to h'_x(t)$ only if this increases the fidelity. The protocol is updated until *all* possible local field updates can only decrease the fidelity. In this regard, the obtained protocol is a local minimum with respect to local field updates. The stochastic descent is repeated multiple times with different initial random protocols. The set of protocols obtained with stochastic descent is used to calculate the glass-like order parameter $q(T)$ (see main text). A Python implementation of the algorithm is available on Github.

A comparison between the ML algorithms and other optimal control algorithms is an interesting topic for future works. Here, we only show that both RL and SD are capable of finding optimal bang-bang protocols for the quantum control problem from the main text, see Fig. 6.

COMPARISON BETWEEN DIFFERENT DRIVING PROTOCOLS FOR THE QUBIT

It is interesting to compare the bang-bang and quasi-continuous driving protocols found by the agent to a simple linear protocol, which we refer to as Landau-Zener (LZ), and the geodesic protocol, which optimizes local fidelity close to the adiabatic limit essentially slowing down near the minimum gap [44]. We find that the RL agent offers significantly better solutions in the overconstrained and glassy phases, where the optimal fidelity is always smaller than unity. The Hamiltonian of the qubit together with the initial and target states read:

$$H(t) = -S^z - h_x(t)S^x, \qquad |\psi_i\rangle \sim (-1/2 - \sqrt{5}/2, 1)^{\mathsf{T}}, \qquad |\psi_*\rangle \sim (1/2 + \sqrt{5}/2, 1)^{\mathsf{T}}, \tag{8}$$

where $|\psi_i\rangle$ and $|\psi_*\rangle$ are the ground state of $H(t)$ for $h_i = -2$ and $h_* = +2$ respectively. Note that for bang-bang protocols, the initial and target states are not eigenstates of the control Hamiltonian since $h_x(t)$ takes on the values ± 4.

The RL agent is initiated at the field $h(t = 0) = h_{\min} = -4.0$. The RL protocols are constructed from the following set of jumps, δh_x, allowed at each protocol time step δt:

- *bang-bang* protocol: $\delta h_x \in \{0.0, \pm 8.0\}$ which, together with the initial condition, constrains the field to take the values $h_x(t) \in \{\pm 4.0\}$.

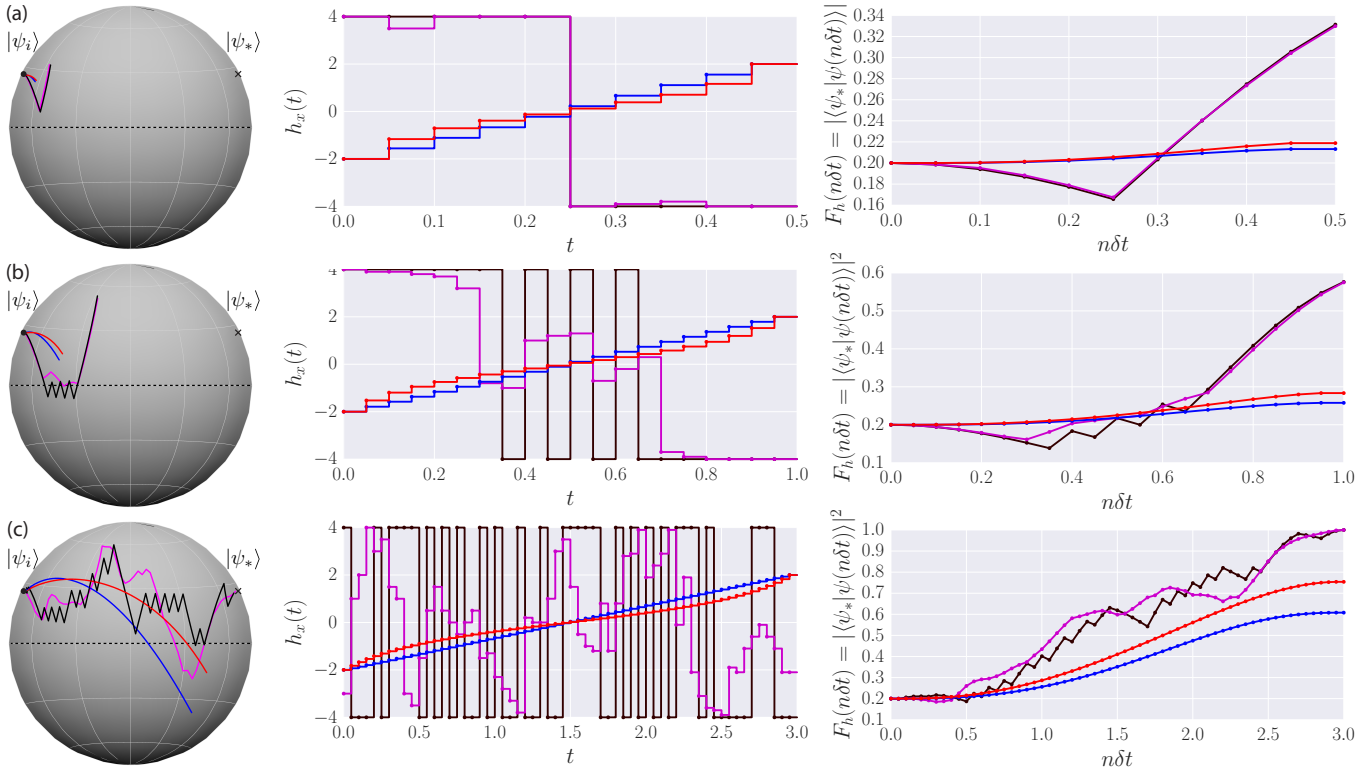


FIG. 7: Comparison between the bang-bang (black) and quasi-continuous (magenta) protocols found by the RL agent, and the Landau-Zener (blue) and geodesic (red) protocols computed from analytical theory in the overconstrained phase for $T = 0.5$ (a), the glassy phase for $T = 1.0$ (b), and the underconstrained phase for $T = 3.0$ (c). The left column shows the representation of the corresponding protocol on the Bloch sphere, the middle one – the protocols themselves, and the right column – the instantaneous fidelity in the target state $\psi_*\rangle$.

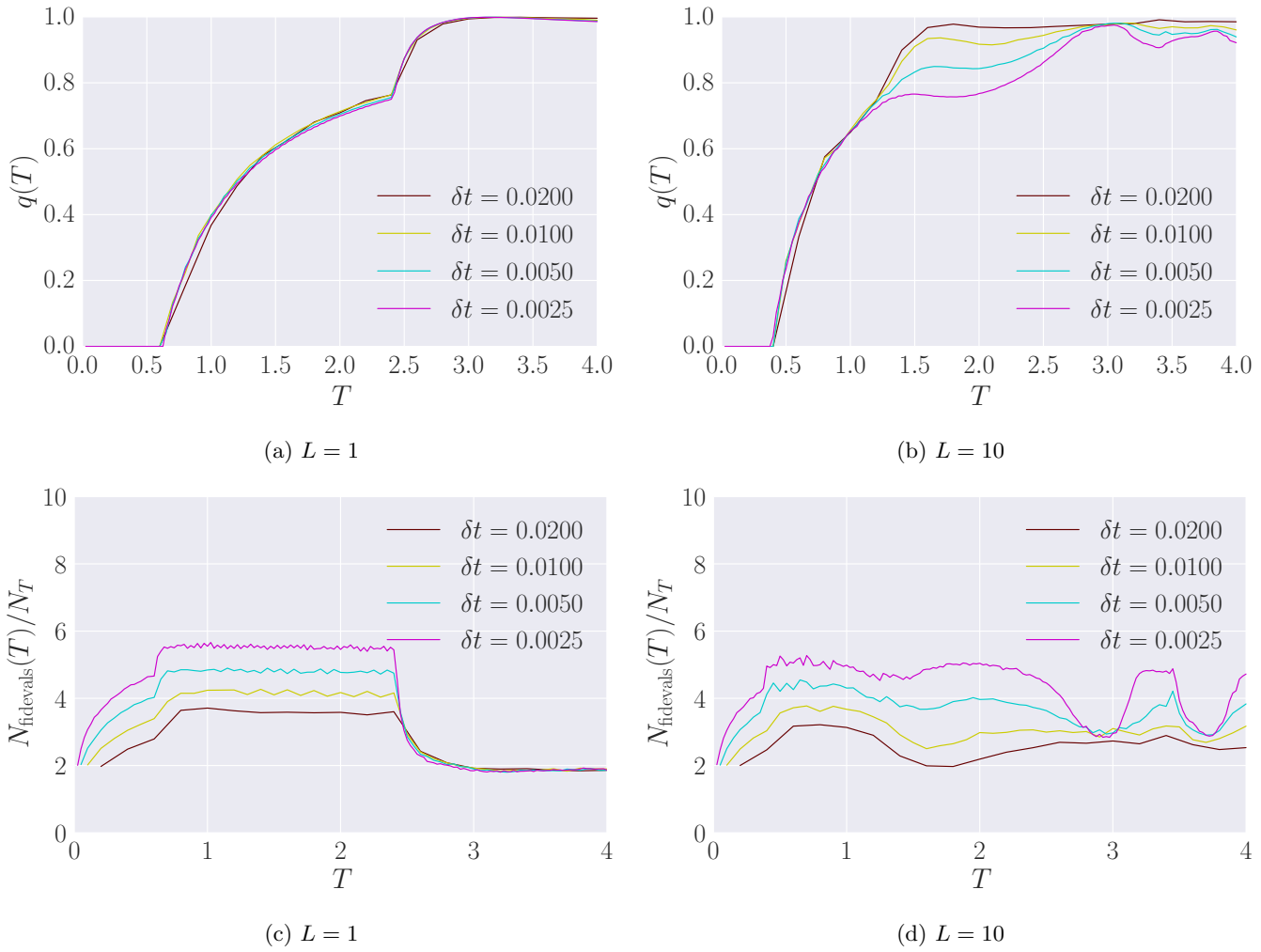
- *quasi-continuous* protocol: $\delta h_x \in \{0.0, \pm 0.1, \pm 0.2, \pm 0.5, \pm 1.0, \pm 2.0, \pm 4.0, \pm 8.0\}$. We restrict the actions available in a state to ensure $h_x(t) \in [-4.0, 4.0]$.

Interestingly, the RL agent figures out that it is always advantageous to first jump to $h_{\max} = +4.0$ before starting the evolution, as a consequence of the positive value of the coefficient in front of S^z.

The analytical adiabatic protocols are required to start and end in the initial and target states, which coincide with the ground states of the Hamiltonians with fields $h_i = -2.0$ and $h_* = 2.0$, respectively. They are defined as follows:

- *Landau-Zener(LZ)* protocol: $h_x(t) = (h_* - h_i)t/T + h_i$
- *geodesic* protocol: $h_x(t) = \tan(at + b)$, where $b = \arctan(h_i)$ and $a = \arctan(h_* - b)/T$.

Figure 7 shows a comparison between these four protocol types for different values of T, corresponding to the three quantum control phases. Due to the instantaneous gap remaining small compared to the total ramp time, the LZ and geodesic protocols are very similar, irrespective of T. The two protocols significantly differ only at large T, where the geodesic protocol significantly outperforms the linear one. An interesting general feature for the short ramp times considered is that the fidelities obtained by the LZ and geodesic protocols are clearly worse than the ones found by the RL agent. This points out the far-from-optimal character of these two approaches, which essentially reward staying close to the instantaneous ground state during time evolution. Looking at the fidelity curves in Fig. 7, we note that, before reaching the optimal fidelity at the end of the ramp for the overconstrained and glassy phases, the instantaneous fidelity drops below its initial value at intermediate times. This suggests that the angle between the initial and target states on the Bloch sphere becomes larger in the process of evolution, before it can be reduced again. Such situation is very reminiscent of counter-diabatic or fast forward driving protocols, where the system can significantly deviate from the instantaneous ground state at intermediate times [9, 29, 45]. Such problems, where the RL agent learns to sacrifice local rewards in view of obtaining a better total reward in the end are of particular interest in RL [11].



(a) $L = 1$ (b) $L = 10$

(c) $L = 1$ (d) $L = 10$

FIG. 8: Finite time step-size, δt, scaling of the order parameter $q(T)$ [top] and the umber of fidelity evaluations per time step, N_{fidevals}/N_T [bottom], to reach a local minimum of the (negative) fidelity landscape.

CRITICAL SCALING ANALYSIS OF THE CONTROL PHASE TRANSITION

In this section, we show convincing evidence for the existence of the phase transitions in the quantum state preparation problem discussed in the main text. We already argued that there is a phase transition in the optimization problem as a function of the protocol time T. Mathematically the problem is formulated as

$$h_x^{\text{optimal}}(t) = \underset{h_x(t):|h_x(t)|\leq 4}{\arg\min} \{-F_h(T)\} = \underset{h_x(t):|h_x(t)|\leq 4}{\arg\max} |\langle\psi_*|\mathcal{T}_t \mathrm{e}^{-i\int_0^T \mathrm{d}t H[h_x(t)]}|\psi_i\rangle|^2, \tag{9}$$

i.e. as finding the optimal driving protocol $h_x^{\text{optimal}}(t)$, which transforms the initial state (the ground state of the Hamiltonian H at $h_x = -2$) into the target state (the ground state of H corresponding to $h_x = 2$) maximizing the fidelity in ramp time T under unitary Schrödinger evolution. We assume that the ramp protocol is bounded, $h_x(t) \in [-4, 4]$, for all times during the ramp. In this section, we restrict the analysis to bang-bang protocols only, for which $h_x(t) \in \{\pm 4\}$. The minimum protocol time step is denoted by δt. There are two different scaling limits in the problem. We define a continuum limit for the problem as $\delta t \to 0$ while keeping the total ramp time $T = const.$ Additionally, there is the usual thermodynamic limit, where we send the system size $L \to \infty$.

As we already alluded to in the main text, one can think of this optimization problem as a minimization in the (negative) fidelity landscape, determined by the mapping $h_x(t) \mapsto -F_h(T)$, where each protocol is assigned a point in fidelity space – the probability of being in the target state after evolution for a fixed ramp time T. Finding the global minimum of the landscape then corresponds to obtaining the best (optimal) driving protocol for any fixed T.

To obtain the set of local fidelity minima $\{h_x^\alpha(t)|\alpha = 1, \ldots, N_{\text{real}}\}$ of the fidelity landscape at a fixed total ramp time T and protocol step size δt, we apply Stochastic Descent(SD), see above, starting from a random protocol configuration, and introduce random *local* changes to the bang-bang protocol shape until the fidelity can no longer be improved. This method is guaranteed to find a set of representative local (negative) fidelity minima with respect to "1-flip" dynamics, mapping out the bottom of the landscape of $-F_h(T)$. Keeping track of the mean number of fidelity evaluations N_{fidevals} required for this procedure, we obtain a measure for the average time it takes the algorithm to settle in a local minimum. While the order parameter $q(T)$ (see below) was used in the main text as a measure for the static properties of the fidelity landscape, dynamic features are revealed by studying the number of fidelity evaluations N_{fidevals}.

As discussed in the main text, the rich phase diagram of the problem can also be studied by looking at the order parameter function q (closely related to the Edwards-Anderson order parameter for detecting glassy order in spin systems [37]):

$$q(T) = \frac{1}{16 N_T} \overline{\sum_{t=1}^{N_T} \{h_x(t) - \overline{h_x}(t)\}^2}, \qquad \overline{h_x}(t) = \frac{1}{N_{\text{real}}} \sum_{\alpha=1}^{N_{\text{real}}} h_x^\alpha(t). \tag{10}$$

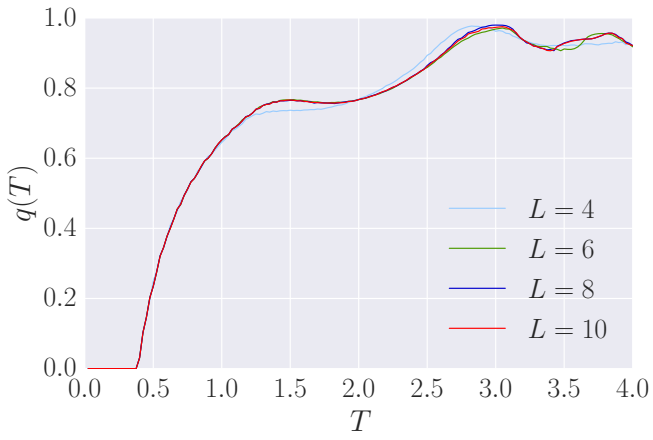
Here, N_T is the total number of protocol time steps of fixed width δt, N_{real} is the total number of random protocol realisations $h_x^\alpha(t)$ probing the minima of the fidelity landscape (see previous paragraph), and the factor $1/16$ serves to normalise the squared bang-bang drive protocol $h_x^2(t)$ within the range $[-1, 1]$.

Single Qubit

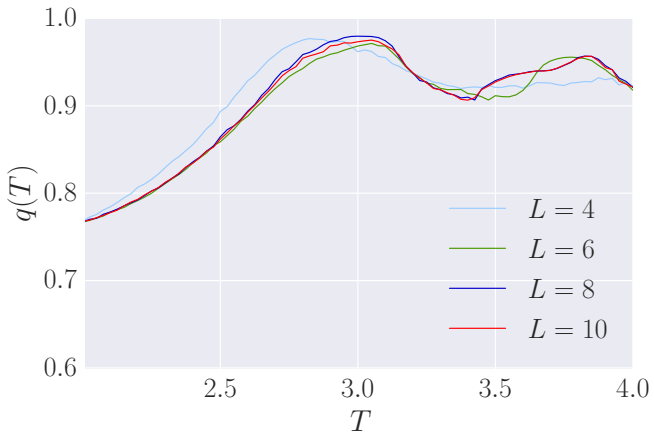
For $T > T_{\text{min}}$, the optimization problem of the single qubit ($L = 1$) is underconstrained, and there exist infinitely many protocols which can prepare the target state with unit fidelity. In analogy with the random k-SAT problem [41], we call this the *underconstrained* phase of the quantum control problem. Intuitively, this comes about due to the large total ramp times available which allow one to correct a set of 'wrong moves' at a later time in order to achieve a better fidelity at the end of the ramp. We have seen that both the Reinforcement Learning (RL) and Stochastic Descent (SD) agents readily and quickly learn to exploit this feature for optimal performance. In this phase, which is not present in the thermodynamic limit $L \to \infty$, there exist analytical results to compute driving protocols of unit fidelity based on the geodesic and counter-diabatic approaches [28, 29, 35]. The driving protocols $h_x^\alpha(t)$, corresponding to the minima of the (negative) fidelity landscape $-F_h(T)$, are completely uncorrelated, resulting in $\overline{h_x}(t) = 0$ and, thus, $q = 1$. As $T \searrow T_{\text{min}}$, the fidelity minima start becoming correlated, reflected in a drop in the value of q. At the time $T = T_{\text{min}}$, a phase transition occurs to a glassy phase with shallow, quasi-degenerate fidelity minima corresponding to many almost equally optimal protocols. Fig. 8a shows that the order parameter q shown in the main text develops a clear non-analyticity in the continuous limit $\delta t \to 0$, which proves the existence of a phase transition in the protocol space. At the same critical time T_{min}, a drastic rapid increase is detected in the number of fidelity evaluations required to map out the fidelity minima, see Fig. 8c.

For $T_c < T < T_{\text{min}}$ the control problem is in the glassy phase. We showed in the main text by examining the DOS of protocols with respect to local flips of the field, that finding the optimal protocol appears as complicated as finding the ground state of a glass. This is reflected in the observed increase of the average number of fidelity evaluations $N_{\text{fidevals}}$with decreasing δt (c.f Fig. 8c), and a decrease in the learning rate of the RL and SD agents. The local minima protocols $\{h_x^\alpha(t)\}$ are strongly correlated, as can be seen from the finite value of the order parameter $q(T)$ in Fig. 9a. More importantly, for any practical purposes, unit fidelity can no longer be obtained under the given dynamical constraints.

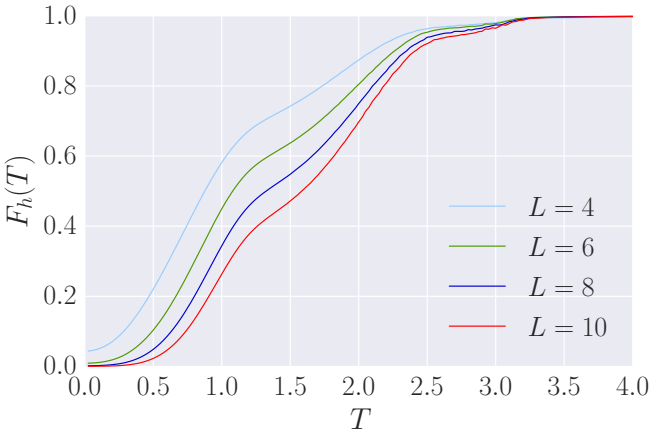
When we reach the second critical point $T = T_c$, another phase transition occurs from the glassy to an overconstrained phase. At $T = T_c$, the order parameter reaches zero, suggesting that the fidelity landscape contains a single minimum. In this phase, i.e. for $T < T_c$, the ramp time is too short to achieve a good fidelity. Nonetheless, in the continuum limit $\delta t \to 0$, there exists a single optimal protocol, although the corresponding maximum fidelity is far from unity. In this overconstrained phase, the optimization problem becomes convex and easy to solve. This is reflected by the observation both the optimal quasi-continuous and bang-band protocols found by the RL agent are nearly identical, cf. Fig. 7. The dynamic character of the phase transition is revealed by a sudden drop in the number of fidelity evaluations N_{fidevals}.



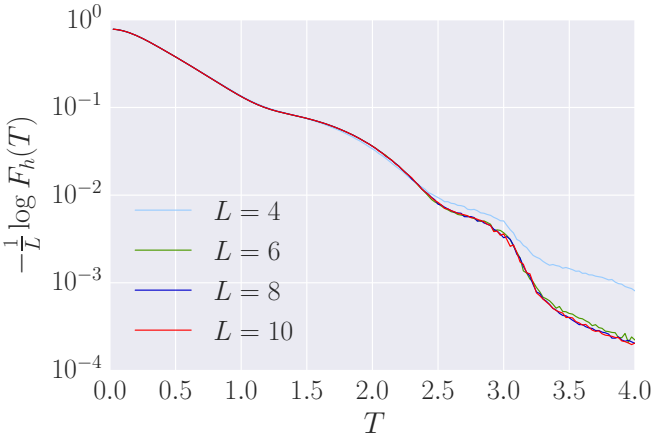
(a) Order parameter $q(T)$ vs. total ramp duration T.



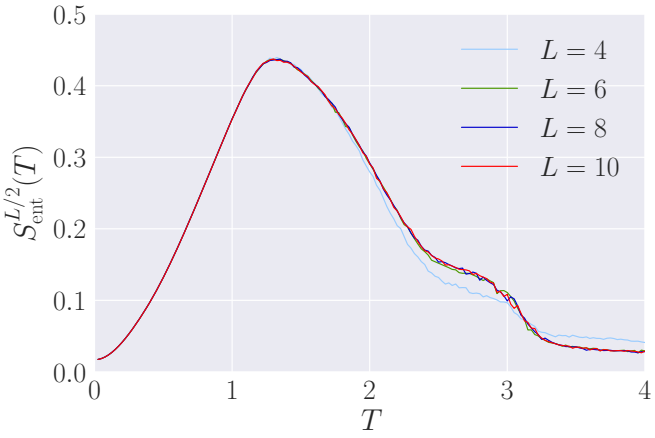
(b) same as (a) with later times zoomed in.



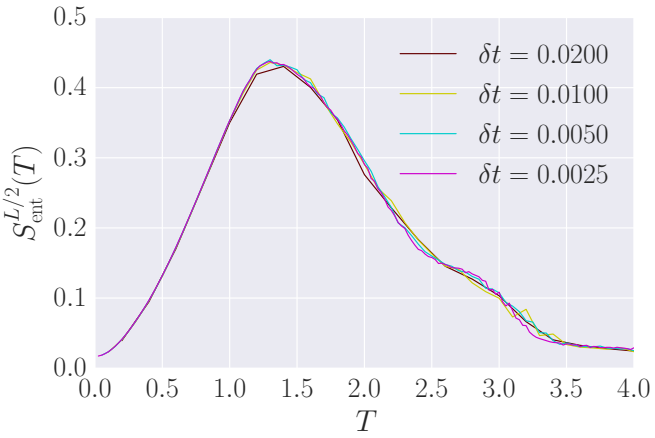
(c) Optimal protocol many-body fidelity: linear scale. Convergence is reached at $L \geq 6$.



(d) Optimal protocol many-body fidelity: logarithmic scale. Convergence is reached at $L \geq 6$.



(e) Entanglement entropy of the half chain as a function of the system size L.



(f) Entanglement entropy of the half chain as a function of the protocol step size δt, for $L = 10$.

FIG. 9: Finite system-size L scaling of the order parameter $q(T)$[top], the many-body fidelity $F_h(T)$ [middle] and the entanglement entropy $S_{\text{ent}}^{L/2}$ for protocol step size $\delta t = 0.0025$.

Coupled Qubits

One can also ask the question what happens to the quantum control phases in the thermodynamic limit, $L \to \infty$. To this end, we repeat the calculation for a series of system sizes. Due to the non-integrable character of the many-body problem, we are limited to small system sizes. However, Fig. 9 shows convincing data that we capture the behaviour of the system in the limit $L \to \infty$ for the relatively short ramp times under consideration. Moreover to our surprise the finite size effect almost entirely disappear for $L \geq 6$ for all range of ramp times we are considering. It seems that system is able to find an optimal solution, where the information simply does not propagate outside of a very small region and hence the optimal protocol rapidly becomes completely insensitive to the system size.

Figure 9c-d shows the system size scaling of the negative logarithmic *many-body* fidelity. While at $L = 4$ we do see remnants of finite size effects, starting from $L = 6$ the curves are barely changing. A similar rapid system-size convergence is observed also for the order-parameter $q(T)$ (see Fig. 9a-b) and the entanglement entropy of the half chain (Fig. 9e). The protocol step size dependence of the order parameter $q(T)$, the average fidelity evaluations N_{fidevals}, and the entanglement entropy $S_{\text{ent}}^{L/2}$ of the half-chain are shown in Figs. 8b, 8d and 9f.

USING INSIGHTS FROM MACHINE LEARNING TO CONSTRUCT A VARIATIONAL THEORY FOR THE PHASE DIAGRAM OF THE QUANTUM CONTROL PROBLEM

In the main text we mention the possibility to use the computer agent to synthesise information into a handful effective degrees of freedom of the problem under consideration, and then use those to construct effective theories for the underlying physics. Let us now demonstrate how this works by giving a specific example, which captures the essence of the phase diagram of quantum control both qualitatively and quantitatively.

Single Qubit

We start, stealing some ideas from the computer agent, by carefully studying the optimal driving protocols it finds in the case of the single qubit. Focussing on bang-bang protocols once again, and looking at the optimal drives for the overconstrained and correlated phases, cf. Fig. 7a-b and Movies 1–3, we recognize an interesting pattern: for $T < T_c$, as we explained in the main text, there is only one minimum in the (negative) fidelity landscape, which dictates a particularly simple form for the bang-bang protocol – a single jump at half the total ramp time $T/2$. On the other hand, for $T_c \leq T \leq T_{\min}$, there appears a sequence of multiple bangs around $T/2$, which grows with increasing the ramp time T. By looking at the Bloch sphere representation, see and Fig. 2a and Movies 1–3, we identify this as an attempt to turn off the h_x-field, once the state has been rotated to the equator, so the instantaneous state can be moved in the direction of the target state in the shortest possible distance.

Hence, it is suggestive to try out a three-pulse protocol as an optimal solution, see Fig. 10a: the first (positive) pulse of duration $\tau^{(1)}/2$ brings the state to the equator. Then the h_x-field is turned off for a time $\tilde{\tau}^{(1)} = T - \tau^{(1)}$, after which a negative pulse directs the state off the equator towards the target state. Since the initial value problem is time-reversal symmetric for our choice if initial and target states, the duration of the third pulse must be the same as that of the first one. We thus arrive at a variational protocol, parametrised by $\tau^{(1)}$, see Fig. 10a.

The optimal fidelity is thus approximated by the variational fidelity $\mathcal{F}_h(\tau^{(1)}, T-\tau^{(1)})$ for the trial protocol [Fig. 10a], and can be evaluated analytically in a straightforward manner:

$$\mathcal{F}_h(\tau^{(1)}, T-\tau^{(1)}) = |\langle\psi_*|\mathrm{e}^{-i\frac{\tau^{(1)}}{2}(S^z - h_{\max}S^x)}\mathrm{e}^{-i(T-\tau^{(1)})S^z}\mathrm{e}^{-i\frac{\tau^{(1)}}{2}(S^z + h_{\max}S^x)}|\psi_i\rangle|^2. \tag{11}$$

The exact expression is rather cumbersome and we choose not to show it explicitly. Optimizing the variational fidelity at a fixed ramp time T, we solve the corresponding transcendental equation to find the extremal value $\tau^{(1)}_{\text{best}}$, and the corresponding optimal variational fidelity $\mathcal{F}_h(T)$, shown in Fig. 10b-c. For times $T \leq T_c$, we find $\tau^{(1)} = T$ which corresponds to $\tilde{\tau}^{(1)} = 0$, i.e. a single bang in the optimal protocol. The overconstrained-to-glassy phase transition at T_c is marked by a non-analyticity at $\tau^{(1)}_{\text{best}}(T_c) = T_c \approx 0.618$. This is precisely the minimal time the agent can take, to bring the state to the equator of the Bloch sphere, and it depends on the value of the maximum magnetic field allowed (here $h_{\max} = 4$). Figure 10d shows that, in the overconstrained phase, the fidelity is optimised at the boundary on the variational domain, although $\mathcal{F}_h(\tau^{(1)}, T-\tau^{(1)})$ is a highly nonlinear function of $\tau^{(1)}$ and T.

For $T_c \leq T \leq T_{\min}$, the time $\tau^{(1)}$ is kept fixed (the equator being the only geodesic for a rotation along the $\hat{z}$-axis of the Bloch sphere), while the second pulse time $\tilde{\tau}^{(1)}$ grows linearly, until the minimum time $T_{\min} \approx 2.415$ is eventually

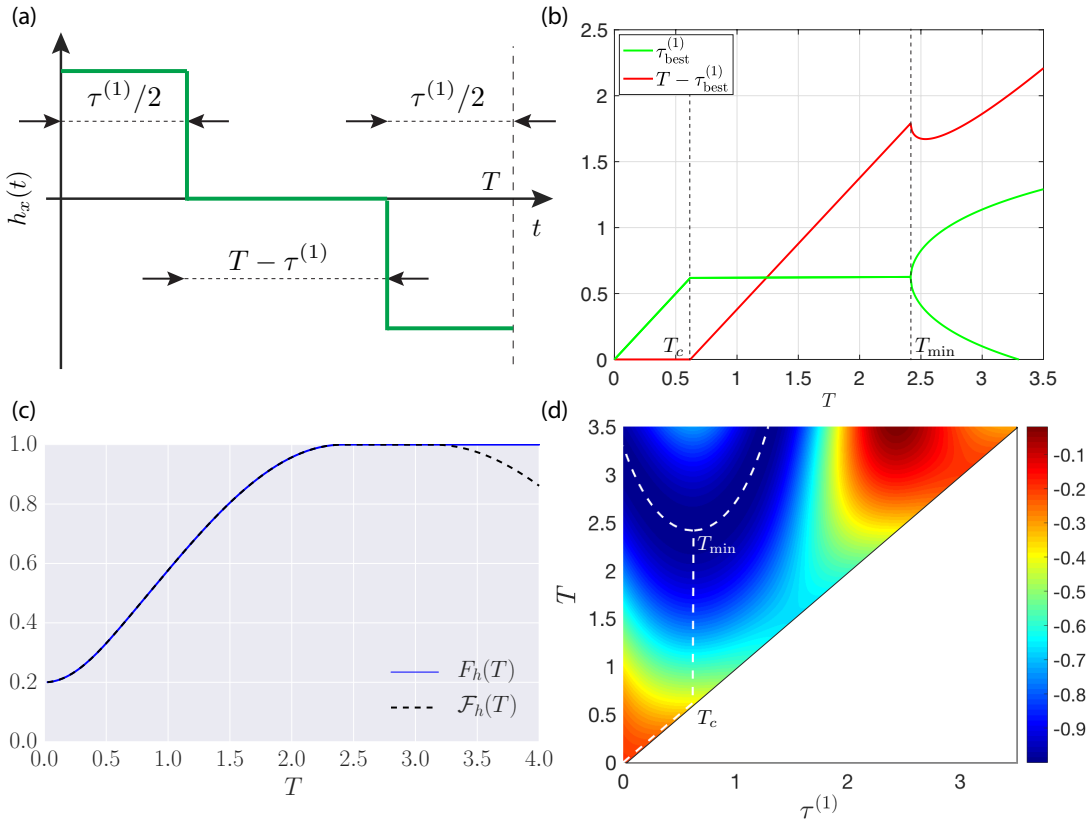


FIG. 10: (a) Three-pulse variational protocol which allows to capture the optimal protocol found by the computer in the overconstrained and the glassy phases of the single qubit problem. (b) $\tau^{(1)}_{\rm best}$ (green), with the non-analytic points of the curve marked by dashed vertical lines corresponding to $T_c \approx 0.618$ and $T_{\rm min} \approx 2.415$. (c) Best fidelity obtained using SD (solid blue) and the variational ansatz (dashed black). (d) The variational (negative) fidelity landscape with the minimum for each T-slice designated by the dashed line which shows the robustness of the variational ansatz against small perturbations.

reached. The minimum time is characterised by a bifurcation in our effective variational theory, as the corresponding variational fidelity landscape develops two maxima, see Fig. 10b,d. Past that point, our simplified ansatz is no longer valid, and the system is in the underconstrained phase. Furthermore, a sophisticated analytic argument based on optimal control theory can give exact expressions for T_c and $T_{\rm min}$ [35].

Coupled Qubits

Let us also discuss the variational theory for the many-body system. Once again, inspired by the structure of the protocols found by the computer, see Movie, we extend the qubit variational protocol, as shown in Fig. 11c. In particular, we add two more pulses to the protocol, retaining its symmetry structure: $h_x(t) \longrightarrow -h_x(T-t)$, whose length is parametrised by a second, independent variational parameter $\tau^{(2)}/2$. Thus, the pulse length where the field is set to vanish, is now given by $\tilde{\tau} = T - \tau^{(1)} - \tau^{(2)}$. This pulses are reminiscent of spin-echo protocols, and appear to be important for entangling and disentangling the state during the evolution.

Notice that this extended variational ansatz includes by definition the simpler ansatz from the single qubit problem discussed above, by setting $\tau^{(2)} = 0$. Let us focus on this simpler case first. The dashed black line in Fig. 11c shows the corresponding 1D variational fidelity. We see that, once again this ansatz captures correctly the critical point T_c separating the overconstrained and the glassy phases. Nevertheless, a comparison with the optimal fidelity [see Fig. 11c] reveals that this variational ansatz breaks down in the glassy phase, although it rapidly converges to the optimal fidelity with decreasing T for $T < T_c$. Going back to Fig. 11b, we note that the value $\tau^{(1)}_{\rm best}$ which maximizes the variational fidelity exhibits a few kinks. However, only the kink at $T = T_c$ captures a physical transition of the

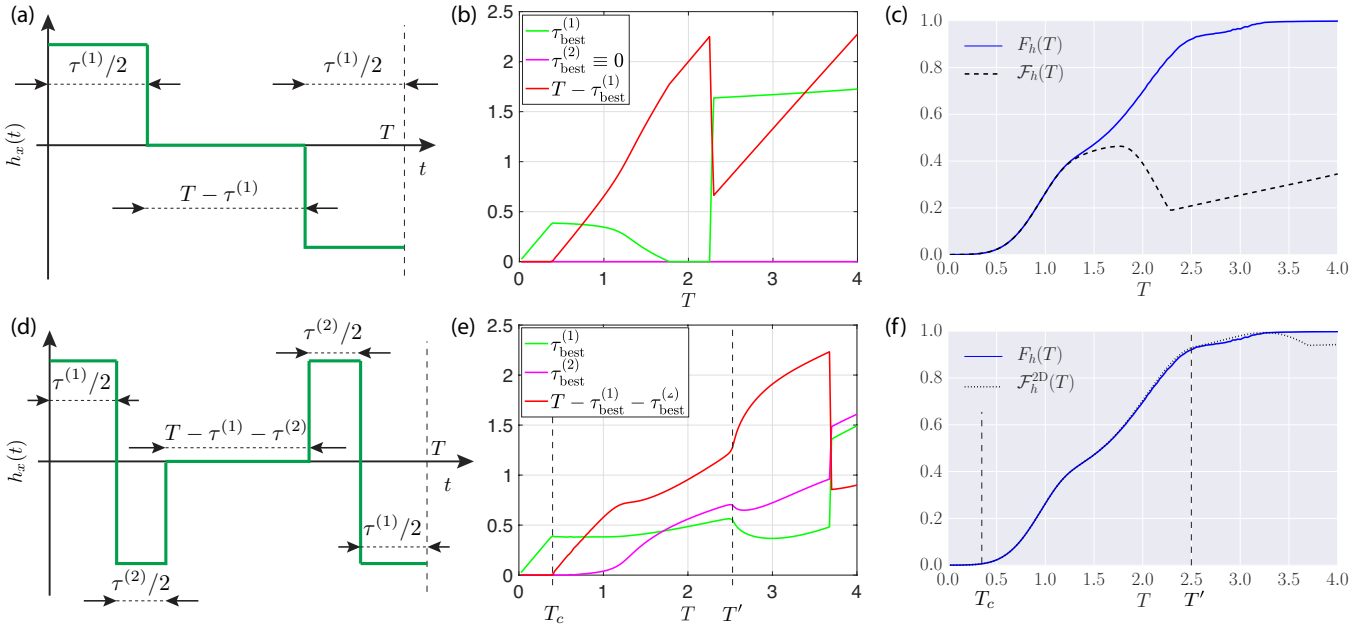


FIG. 11: (a) Three-pulse variational protocol which allows to capture the optimal protocol found by the computer in the overconstrained phase but fails the glassy phase of the nonintegrable many-body problem. This ansatz captures the non-analytic point at $T_c \approx 0.4$ but fails in the glassy phase. (b) The pulse durations $\tau^{(1)}_{\rm best}$ (green) and $\tau^{(2)}_{\rm best}$ (magenta), for highest fidelity variational protocol of length T of the type shown in (a). The fidelity of the variational protocols exhibit a physical non-analyticity at $T_c \approx 0.4$ and unphysical kinks outside the validity of the ansatz. (c) 1D maximal variational fidelity (dashed back) compared to best numerical protocol (solid blue). (d) Five-pulse variational protocol which allows to capture the optimal protocol found by the computer in the overconstrained phase and parts of the glassy phase of the nonintegrable many-body problem. (e) The pulse durations $\tau^{(1)}_{\rm best}$ (green) and $\tau^{(2)}_{\rm best}$ (magenta) for for best variational protocol of length T of the type shown in (d). These variational protocols exhibit physical non-analyticities at $T_c \approx 0.4$ and $T' \approx 2.5$ (vertical dashed lines) (f) 2D maximal variational fidelity (dashed-dotted back) compared to best numerical protocol (solid blue).

original control problem, while the others appear as artifacts of the simplified variational theory.

Let us now turn on the second variational parameter $\tau^{(2)}$, and consider the full two-dimensional variational problem:

$$\mathcal{F}^{\rm 2D}_h(\tau^{(1)},\tau^{(2)},T-\tau^{(1)}-\tau^{(2)}) = \tag{12}$$
$$|\langle\psi_*|\mathrm{e}^{-i\frac{\tau^{(1)}}{2}(S^z-h_{\max}S^x)}\mathrm{e}^{-i\frac{\tau^{(2)}}{2}(S^z+h_{\max}S^x)}\mathrm{e}^{-i(T-\tau^{(1)})S^z}\mathrm{e}^{-i\frac{\tau^{(2)}}{2}(S^z-h_{\max}S^x)}\mathrm{e}^{-i\frac{\tau^{(1)}}{2}(S^z+h_{\max}S^x)}|\psi_i\rangle|^2.$$

Figure 11 shows the best variational fidelity $\mathcal{F}^{\rm 2D}_h$ [f] and the corresponding values of $\tau^{(1)}_{\rm best}$ and $\tau^{(2)}_{\rm best}$ [e] for this maximum fidelity variational protocol [d]. There are two important points here: (i) Fig. 11f shows that the 2D variational fidelity seemingly reproduces the optimal fidelity on a much longer scale, i.e. for all protocol durations $T \lesssim 3.3$. (ii) the 2D variational ansatz reduces to the 1D one in the overconstrained phase $T \leq T_c$. In particular, both pulse lengths $\tau^{(1)}_{\rm best}$ and $\tau^{(2)}_{\rm best}$ exhibit a non-analyticity at $T = T_c$, but also at $T' \approx 2.5$. Interestingly, the 2D variational ansatz captures the optimal fidelity on both sides of T' which suggests that there is likely yet another transition within the glass phase, hence the different shading in the many-body phase diagram [Fig. 3, main text]. Similar to the 1D variational problem, here we also find artefact transitions (non-analytic behavior in $\tau^{(i)}_{\rm max}$) outside of the validity of the approximation.

In summary, we have shown how, by carefully studying the driving protocols found by the computer agent, one can obtain ideas for effective theories which capture the essence of the underlying physics. This is similar to using a ϕ^4-theory to describe the physics of the Ising phase transition. The key difference is that the present problem is out-of-equilibrium, where no general theory of statistical mechanics exists so far. We hope that, an underlying pattern between such effective theories can be revealed with time, which might help shape the principles of a theory of statistical physics away from equilibrium.