Pareto-optimal cycles for power, efficiency and fluctuations of quantum heat engines using reinforcement learning

Paolo A. Erdman [1], Alberto Rolandi [2], Paolo Abiuso [3], Martí Perarnau-Llobet [2,*] and Frank Noé [6,1,4,5,†]

[1]*Department of Mathematics and Computer Science, Freie Universität Berlin, Arnimallee 6, 14195 Berlin, Germany*
[2]*Department of Applied Physics, University of Geneva, 1211 Geneva, Switzerland*
[3]*ICFO Institut de Ciències Fotòniques, The Barcelona Institute of Science and Technology, 08860 Castelldefels (Barcelona), Spain*
[4]*Department of Physics, Freie Universität Berlin, Arnimallee 6, 14195 Berlin, Germany*
[5]*Department of Chemistry, Rice University, Houston, Texas 77005, USA*
[6]*Microsoft Research AI4Science, Karl-Liebknecht Str. 32, 10178 Berlin, Germany*

The full optimization of a quantum heat engine requires operating at high power, high efficiency, and high stability (i.e., low power fluctuations). However, these three objectives cannot be simultaneously optimized—as indicated by the so-called thermodynamic uncertainty relations—and a systematic approach to finding optimal balances between them including power fluctuations has, as yet, been elusive. Here we propose such a general framework to identify Pareto-optimal cycles for driven quantum heat engines that trade off power, efficiency, and fluctuations. We then employ reinforcement learning to identify the Pareto front of a quantum dot-based engine and find abrupt changes in the form of optimal cycles when switching between optimizing two and three objectives. We further derive analytical results in the fast- and slow-driving regimes that accurately describe different regions of the Pareto front.

Introduction. Stochastic heat engines are devices that convert between heat and work on the nanoscale [1–3]. Steady-state heat engines (SSHEs) perform work against external thermodynamic forces (e.g., a chemical potential difference) after reaching a nonequilibrium steady state [4], whereas periodically driven heat engines (PDHEs) perform work against external driving fields through time-dependent cycles [5]. The performance of heat engines is usually characterized by the output power and efficiency, and their optimization has been thoroughly addressed in literature [6–30]. However, in contrast to their macroscopic counterpart, the performance of quantum and microscopic engines is strongly influenced by power fluctuations. Although early works have started optimizing power fluctuation [31–35], a framework to fully optimize the performance of microscopic heat engines that accounts for power fluctuations is currently lacking; this Letter fills this void.

An ideal engine operates at high power, high efficiency, and low-power fluctuations; however, such quantities usually cannot be optimized simultaneously, but one must seek trade-offs. In SSHEs, a rigorous manifestation of this trade-off is given by the thermodynamic uncertainty relations [36–44].

For "classical" stochastic SSHEs (i.e., in the absence of quantum coherence) operating between two Markovian reservoirs at inverse temperatures β_C (cold) and β_H (hot), they read [38]

$$\xi \equiv \frac{2}{\beta_C} \frac{\langle P \rangle}{\langle \Delta P \rangle} \frac{\eta}{\eta_c - \eta} \leqslant 1, \qquad (1)$$

where $\langle P \rangle$ and $\langle \Delta P \rangle$ are, respectively, the average power and power fluctuations, η is the efficiency, and $\eta_c \equiv 1 - \beta_H/\beta_C$ is the Carnot efficiency. Such thermodynamic uncertainty relations imply, for example, that high efficiency can only be attained at the expense of low-power or high-power fluctuations. The thermodynamic uncertainty relation inequality (1) can be violated with quantum coherence [45–53] and in PDHEs [54–59]. This has motivated various generalized thermodynamic uncertainty relations [60–65], in particular, for time-symmetric driving [40,55] and slowly driven stochastic engines [66,67].

Despite their importance, thermodynamic uncertainty relations provide an incomplete picture of the trade-off: whereas high values of ξ may appear more favorable, this does not give us any information on the individual objectives. Indeed, Refs. [56,66] have shown that high values of ξ can be achieved in the limit of vanishing power, whereas often the goal is to operate at high power or efficiency.

In this Letter, we propose a framework to optimize arbitrary trade-offs among power, efficiency and power fluctuations in arbitrary PDHEs described by Lindblad dynamics [68–71]; this framework enables the use of various optimization techniques, such as the Pontryagin minimum principle [72] or reinforcement learning (RL) [73] to find Pareto-optimal cycles, i.e., those cycles where no objective can be
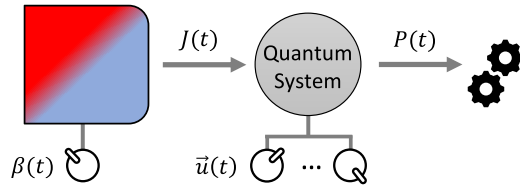
FIG. 1. A quantum system (gray circle) is coupled to a thermal bath (left box) characterized by a controllable inverse temperature $\beta(t)$. The coupling produces a heat flux $J(t)$. Control parameters $\vec{u}(t)$ allow us to control the state of the system and the power $P(t)$ extracted from the system.

further improved without sacrificing another one. We then employ RL to fully optimize a quantum dot- (QD-) based heat engine [74]. We characterize the Pareto front, i.e., the set of values $\{\langle P \rangle, \langle \Delta P \rangle, \eta\}$ corresponding to Pareto-optimal cycles, and evaluate the thermodynamic uncertainty relation ratio ξ on such optimal cycles. Furthermore, we derive analytical results for the Pareto front and ξ in the fast- [6,8,20,23,27,75,76] and slow- [9,26,77–88] driving regimes, i.e., when the period of the cycle is, respectively, much shorter or much longer than the thermalization timescale of the system.

Multiobjective optimization of quantum heat engines. We describe a PDHE as a quantum system coupled to a heat bath whose inverse temperature $\beta(t)$ can be tuned in time between two extremal values β_H and β_C with $\beta_H \leqslant \beta_C$ (Fig. 1). The coupling produces a heat flux $J(t)$ from the bath to the quantum system. The PDHE is further controlled by time-dependent control parameters $\vec{u}(t)$ that allow exchanging work with the system, producing power $P(t)$. A thermodynamic cycle is then described by periodic functions $\beta(t)$ and $\vec{u}(t)$ with period τ. This framework includes standard PDHEs in which the system is sequentially put in contact with two baths [by abruptly changing the values of $\beta(t)$] and cases where $\beta(t)$ varies smoothly in time [66,84,87,89–92]. We assume that the dynamics of the system is described by a Markovian master equation, i.e., that the reduced density matrix ρ_t of the quantum system satisfies

$$\dot{\rho}_t = \mathcal{L}_{\vec{u}(t),\beta(t)}\rho_t, \tag{2}$$

where $\mathcal{L}_{\vec{u}(t),\beta(t)}$ is the Lindbladian describing the evolution of the system [70].

We are interested in characterizing the performance of PDHEs computing the average power $\langle P \rangle$, average irreversible entropy production $\langle \Sigma \rangle$, and average power fluctuations $\langle \Delta P \rangle$ in the asymptotic limit cycle [66,89,90], i.e., in the limit of infinite repetitions of the cycle. In such a limit, ρ_t becomes periodic with the same periodicity τ of the control (see the Supplemental Material [93]).

Given the density-matrix ρ_t, the calculation of $\langle P \rangle$ and $\langle \Sigma \rangle$ can be performed using the standard approach first put forward in Ref. [94] (see the Supplemental Material [93] for details). Defining the time-average $\langle O \rangle$ of an arbitrary quantity $O(t)$ as $\langle O \rangle \equiv \tau^{-1} \int_0^\tau O(t)\,dt$, we can calculate $\langle P \rangle$ and $\langle \Sigma \rangle$ by averaging

$$P(t) = -\text{Tr}[\rho_t \dot{H}_{\vec{u}(t)}], \qquad \Sigma(t) = -\text{Tr}[\dot{\rho}_t H_{\vec{u}(t)}]\beta(t). \tag{3}$$

Note that we compute the entropy production $\Sigma(t) \equiv -J(t)\beta(t)$ neglecting the entropy variation ΔS of the quantum system since the periodicity of the state in the limiting cycle implies that $\Delta S = 0$ after each repetition of the cycle.

The average fluctuations $\langle \Delta P \rangle$, however, cannot be expressed as a time average of a function of the state ρ_t since they involve a two-point correlation function. Indeed, from Ref. [82], we can express them as

$$\langle \Delta P \rangle = \lim_{T \to \infty} \frac{1}{T} \int_0^T dt \, \text{Tr}[s_t \dot{H}_{\vec{u}(t)}], \tag{4}$$

where we define

$$s_t \equiv \int_0^t dt' P(t,t')[\dot{H}_{\vec{u}(t')}\rho_{t'}] + \langle w \rangle_t \rho_t + \text{H.c.} \tag{5}$$

In Eq. (5) $P(t,t') \equiv \overleftarrow{T} \exp[\int_{t'}^t dt'' \mathcal{L}_{\vec{u}(t''),\beta(t'')}]$ is the propagator, $\langle w \rangle_t \equiv -\int_0^t dt' \text{Tr}[\rho_{t'}\dot{H}_{\vec{u}(t')}]$ is the total average work extracted between time 0 and t, and H.c. represents the complex conjugate of the right-hand side.

Here, we overcome the difficulty of computing nested integrals and two-point correlation function in Eqs. (4) and (5) by noting that s_t is a traceless Hermitian operator that satisfies

$$\dot{s}_t = \mathcal{L}_{\vec{u}(t),\beta(t)}s_t + \{\rho_t, \dot{H}_{\vec{u}(t)}\} - 2\,\text{Tr}[\rho_t \dot{H}_{\vec{u}(t)}]\rho_t, \tag{6}$$

and becomes periodic with period τ in the limiting cycle (see the Supplemental Material [93]).

Therefore, by considering (ρ_t, s_t) as an "extended state" satisfying the equations of motion in (2) and (6) in the time-interval $[0, \tau]$ with periodic boundary conditions, we can compute $\langle P \rangle$, $\langle \Sigma \rangle$, and $\langle \Delta P \rangle$ as time averages of $P(t)$, $\Sigma(t)$, and $\Delta P(t)$, where $P(t)$ and $\Sigma(t)$ are defined in Eq. (3), and where

$$\Delta P(t) \equiv \text{Tr}[s_t \dot{H}_{\vec{u}(t)}]. \tag{7}$$

Note that these are now linear functionals of the extended state.

To identify Pareto-optimal cycles, we introduce the dimensionless figure of merit

$$\langle F \rangle = a\frac{\langle P \rangle}{P_{\max}} - b\frac{\langle \Delta P \rangle}{\Delta P(P_{\max})} - c\frac{\langle \Sigma \rangle}{\Sigma(P_{\max})}, \tag{8}$$

where $a, b, c \geqslant 0$ are three scalar weights, satisfying $a + b + c = 1$, that determine how much we are interested in each of the three objectives, and $P_{\max}$, $\Delta P(P_{\max})$, and $\Sigma(P_{\max})$ are, respectively, the average power, fluctuations and entropy production of the cycle that maximizes the power. Note, given the relation between entropy production and efficiency, cycles that are Pareto optimal for $\{\langle P \rangle, \langle \Delta P \rangle, \eta\}$, are also Pareto optimal for $\{\langle P \rangle, \langle \Delta P \rangle, \langle \Sigma \rangle\}$ (see the Supplemental Material [93]). The positive sign in front of $\langle P \rangle$ in Eq. (8) ensures that we are maximizing the power, while the negative sign in front of $\langle \Delta P \rangle$ and $\langle \Sigma \rangle$ ensures that we are minimizing power fluctuations and the entropy production. Interestingly, if convex, it has been shown that the full Pareto front can be identified repeating the optimization of $\langle F \rangle$ for many values of a, b, and c [95,96].

Since $\langle F \rangle$ is a linear combination of the average thermodynamic quantities, using Eqs. (3) and (7) we can express $\langle F \rangle$ as a time integral of a function of the extended state (ρ_t, s_t) and

of the controls $\vec{u}(t)$ and $\beta(t)$,

$$\langle F \rangle = \int_0^\tau G[\rho_t, s_t, \vec{u}(t), \beta(t)] dt, \qquad (9)$$

where $G[\rho_t, s_t, \vec{u}(t), \beta(t)]$ is a suitable function. The optimization of $\langle F \rangle$ in this form is precisely the type of problem that can be readily tackled using optimization techniques, such as the Pontryagin minimum principle [72] or RL [73]. In this Letter, we employ the latter.

QD heat engine. In the following, we compute Pareto-optimal cycles in a minimal heat engine consisting of a two-level system coupled to a fermionic bath with flat density of states. This represents a model of a single-level QD [6,10,74]. The Hamiltonian reads

$$H_{u(t)} = u(t) \frac{E_0}{2} \sigma_z, \qquad (10)$$

where $u(t)$ is our single control parameter, E_0 is a fixed energy scale, and σ_z is a Pauli matrix. Denoting with $|1\rangle$ the excited state of $H_{u(t)}$, and defining $p_t \equiv \langle 1 | \rho_t | 1 \rangle$ as the probability of being in the excited state, the Lindblad equation (2) becomes $\dot{p}_t = -\gamma(p_t - \pi_{\vec{u}(t),\beta(t)})$, where γ^{-1} is the thermalization timescale arising from the coupling between system and bath, and $\pi_{\vec{u}(t),\beta(t)} = f[\beta(t) E_0 u(t)]$ is the excited level population of the instantaneous Gibbs state, with $f(x) \equiv (1 + e^x)^{-1}$ [23].

Optimal cycles with RL and analytical results. We optimize $\langle F \rangle$ of the QD heat engine using three different tools: RL, analytics in the fast-driving regime, and analytics in the slow-driving regime.

The RL-based method allows us to numerically optimize $\langle F \rangle$ without making any approximations on the dynamics, exploring all possible (time-discretized) time-dependent controls $\beta(t)$ and $u(t)$ subject to the constraints $\beta(t) \in [\beta_H, \beta_C]$ and $u(t) \in [u_{\min}, u_{\max}]$ (thus, beyond fixed structures, such as Otto cycles), and identifying automatically the optimal period. The RL method, based on the soft actor-critic algorithm [97–99] and generalized from [100,101], additionally includes the crucial impact of power fluctuations and identifies Pareto-optimal cycles (see the Supplemental Material [93] for technical details and for benefits of using RL). Machine learning methods have been employed for other quantum thermodynamic [102–104] and quantum control [105–117] tasks.

The fast-driving regime assumes that $\tau \ll \gamma^{-1}$. Interestingly, without any assumption on the driving speed, we show [93] that any trade-off between power and entropy production [$b = 0$ in Eq. (8)] in the QD engine is maximized by Otto cycles in the fast-driving regime, i.e., switching between two values of $\beta(t)$ and $u(t)$ "as fast as possible" [23,27]. We thus expect such "fast-Otto cycles" to be nearly optimal in the high power or efficiency regime. Moreover, we derive analytical expressions to compute and optimize $\{\langle P \rangle, \langle \Delta P \rangle, \langle \Sigma \rangle\}$ efficiently in arbitrary systems in the fast-driving regime [93].

The slow-driving regime corresponds to the opposite limit, i.e., $\tau \gg \gamma^{-1}$. Since entropy production and power fluctuations can be minimized by considering quasistatic cycles (see, e.g., Refs. [56,66]), we expect this regime to be nearly optimal in the low-power regime, i.e., for low values of a in Eq. (8). To make analytical progress in this regime, we maximize Eq. (8) assuming a finite-time Carnot cycle (see
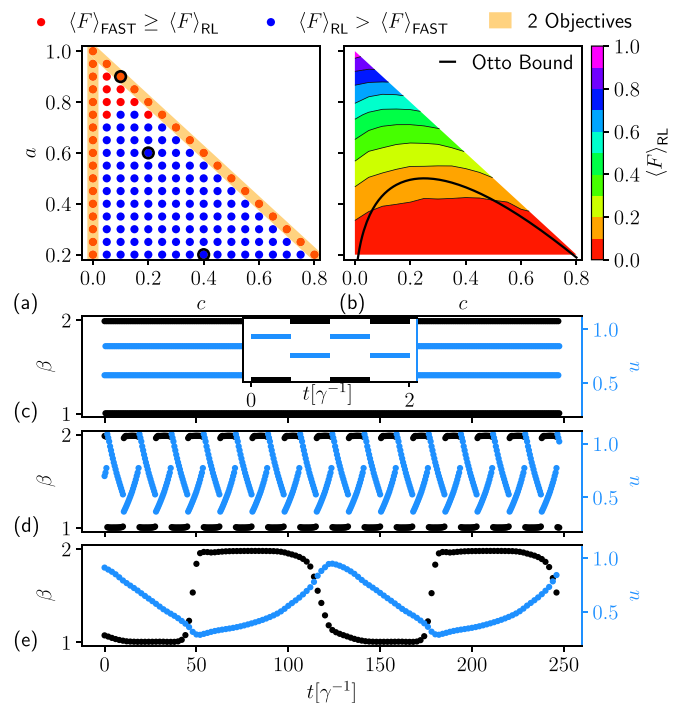

FIG. 2. Optimization of $\langle F \rangle$ at different values of a and c with $b = 1 - a - c$, for a QD-based PDHE. Each dot in panel (a) displays, as a function of c and a, whether $\langle F \rangle_{RL} > \langle F \rangle_{FAST}$ (blue dots) or not (red dots). Points with $a \sim 0$ are not displayed since, in such a regime, optimal cycles become infinitely long (to minimize entropy production and fluctuations) and the RL method does not converge reliably [93]. (b) Contour plot of $\langle F \rangle_{RL}$, as a function of c and a, using the data points of (a). The black line represents the curve below which $\langle F \rangle_{FAST} = 0$. (c)–(e) cycles, described by piecewise constant values of β (black dots) and u (blue dots) as a function of t, identified at the three values of a and c highlighted in black in panel (a) (respectively, from top to bottom). The inset in panel (c) represents a zoom into the corresponding cycle, which is a fast-Otto cycle. Parameters: $\beta_C = 2$, $\beta_H = 1$, $u_{\min} = 0.2$, $u_{\max} = 1.1$, and $E_0 = 2.5$.

the Supplemental Material [93]). The obtained results naturally generalize previous considerations for low-dissipation engines [9,10,13,14,21,22,26] to account for the role of fluctuations (see also Ref. [32]). The main technical tool is the geometric concept of "thermodynamic length" [77,86], which yields the first-order correction in $(\gamma \tau)^{-1}$ from the quasistatic limit.

We now present the results. Each point in Fig. 2(a) corresponds to a separate optimization of $\langle F \rangle$ with weights c and a displayed on the x-y axis. Since $b = 1 - a - c$, points lying on the sides of the triangle (highlighted in yellow) correspond to optimizing the trade-off between two objectives, whereas points inside the triangle take all three objectives into account. Denoting the figure of merit optimized with RL and with fast-Otto cycles with $\langle F \rangle_{RL}$ and $\langle F \rangle_{FAST}$, in Fig. 2(a), we show blue (red) dots when $\langle F \rangle_{RL} > \langle F \rangle_{FAST}$ ($\langle F \rangle_{RL} \leqslant \langle F \rangle_{FAST}$), whereas Fig. 2(b) is a contour plot of $\langle F \rangle_{RL}$. As expected, there are red dots when $b = 0$ (along the hypotenuse), but it turns out that fast-Otto cycles are optimal also when $c = 0$. However, as soon as all three weights are finite, the optimal cycles identified with RL change abruptly and
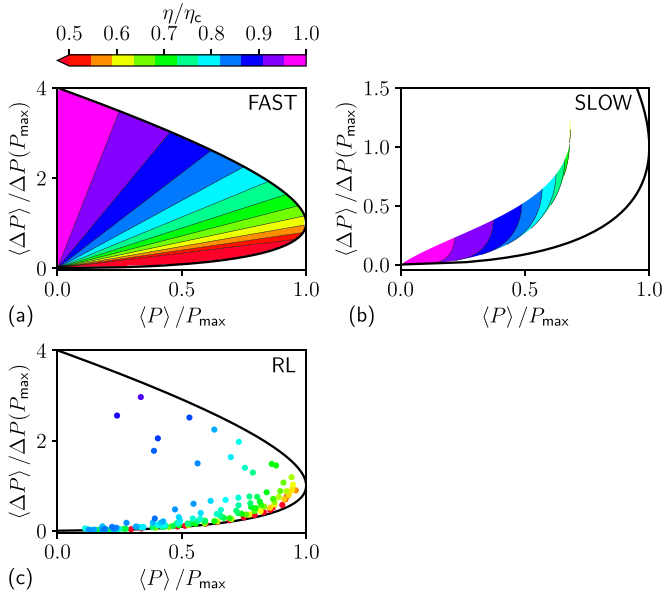
FIG. 3. Pareto front found optimizing $\langle F \rangle$ with fast-Otto cycles in the limit of small temperature differences [panel (a)], optimizing $\langle F \rangle$ in the slow-driving regime [panel (b)], and numerically using RL [panel (c)]. The system parameters are as in Fig. 2. All panels display $\langle \Delta P \rangle / \Delta P(P_{max})$ as a function of $\langle P \rangle / P_{max}$ (x axis) and of η / η_c (color). The black curve represents the outer border of the Pareto front derived analytically in the Supplemental Material [93].

outperform fast-Otto cycles. Furthermore, we note that whereas $\langle F \rangle_{RL}$ is positive for all values of the weights, $\langle F \rangle_{FAST} = 0$ below the black curve shown in Fig. 2(b) (see the Supplemental Material [93] for its analytic expression).

To visualize the changes in protocol space, in Figs. 2(c)–2(e) we show the cycles identified with RL at the three different values of the weights highlighted by a black circle in Fig. 2(a) (respectively, from top to bottom). Since RL identifies piecewise constant controls, the cycle is displayed as dots corresponding to the value of $\beta(t)$ (black dots) and $u(t)$ (blue dots) at each small time step. First, we note that the inverse temperature abruptly switches between β_H and β_C for all values of the weights so that, in this engine, no gain arises when smoothly varying the temperature. As expected, the cycle identified by RL in Fig. 2(c), corresponding to the black point on the hypotenuse in Fig. 2(a), is a fast-Otto cycle (a "zoom" in a short-time interval is shown in the inset). However, moving down in weight space to the black dot at $a = 0.6$ and $c = 0.2$, we see that the corresponding cycle [Fig. 2(d)] now displays a finite period with linear modulations of $u(t)$ at fixed temperatures and a discontinuity of $u(t)$ when switching between β_H and β_C. The cycle in Fig. 2(e), corresponding to the lowest black dot at $a = 0.2$ and $c = 0.4$, displays an extremely long period $\tau \approx 125\gamma^{-1}$, which is far in the slow-driving regime. Optimal cycles, therefore, interpolate between the fast- and the slow-driving regimes as we move in weight space [Fig. 2(a)] from the sides to the lower and central region (i.e., switching from two to three objectives).

In Fig. 3, we display the Pareto front, i.e., we plot the value of P/P_{max}, $\Delta P/\Delta P(P_{max})$, and of the efficiency η/η_c, found maximizing $\langle F \rangle$ for various values of the weights.

Figure 3(a) is derived in the fast-driving regime assuming a small temperature difference, whereas Fig. 3(b) is derived in the slow-driving regime. The RL results, shown in Fig. 3(c), correspond to the points in Fig. 2(a). First, we note that, by definition of the Pareto front, the "outer border" corresponds to points where we only maximize the trade-off between the two objectives $\langle P \rangle$ and $\langle \Delta P \rangle$. Since these points are optimized by fast-Otto cycles, the black border of Fig. 3(a), also shown in Figs. 3(b) and 3(c), is exact and given by (see the Supplemental Material [93] for details)

$$\frac{\langle P \rangle}{P_{max}} = 2\sqrt{\frac{\langle \Delta P \rangle}{\Delta P(P_{max})}} - \frac{\langle \Delta P \rangle}{\Delta P(P_{max})}. \qquad (11)$$

Moreover, in this setup, we can establish an exact mapping between the performance of a SSHE and of our PDHE operated with fast-Otto cycles (see the Supplemental Material [93]). Since SSHEs satisfy Eq. (1), also fast-Otto cycles have $\xi \leqslant 1$. Furthermore, for small temperature differences $\xi = 1$. This allows us to fully determine the internal part of the Pareto front in the fast-driving regime using the thermodynamic uncertainty relations, i.e., $P/P_{max} = [\Delta P/\Delta P(P_{max})](\eta_c - \eta)/\eta$. Indeed, the linear contour lines in Fig. 3(a) stem from the linearity between P and ΔP, the angular coefficient being determined by the efficiency.

Comparing Figs. 3(a) and 3(b), we see where the fast- and slow-driving regimes are optimal. As expected, the slow-driving Pareto front cannot reach the black border, especially in the high-power area where fast-Otto cycles are optimal. However, in the low-power and low fluctuation regime, cycles in the slow driving substantially outperform fast-Otto cycles by delivering a higher efficiency [purple and blue regions in Fig. 3(b)].

Interestingly, the RL points in Fig. 3(c) capture the best features of both regimes. RL can describe the high-power and low fluctuation regime displaying both red and blue/green dots near the lower border. The red dots are fast-Otto cycles that are optimal exactly along the border but deliver a low efficiency. The blue/green dots instead are finite-time cycles that deliver a much higher efficiency by sacrificing a very small amount of power and fluctuations. This dramatic enhancement of the efficiency as we depart from the lower border is another signature of the abrupt change in optimal cycles.

Violation of thermodynamic uncertainty relation. At last, we analyze the behavior of the thermodynamic uncertainty relation ratio ξ, which represents a relevant quantity combining the three objectives, computing it on Pareto-optimal cycles (recall that $\xi \leqslant 1$ for classical stochastic SSHEs but PDHEs can violate this bound [54–59]). In Fig. 4(a), we show a contour plot of ξ, computed with RL as a function of a and c. Because of the mapping between SSHEs and fast-Otto cycles [93], we have $\xi = 1$ along the sides of the triangle where only two objectives are optimized. However, this mapping breaks down for finite-time cycles, allowing us to observe a strong increase in ξ in the green/purple region in Fig. 4(a). As shown in Fig. 2, this region corresponds to long cycles operated in the slow-driving regime where violations of thermodynamic uncertainty relations had already been reported [56,66]. In Figs. 4(b)–4(d) we show a log-log plot of ξ, respectively, as a function of P/P_{max}, $\Delta P/\Delta P(P_{max})$, and $\Sigma/\Sigma(P_{max})$ with
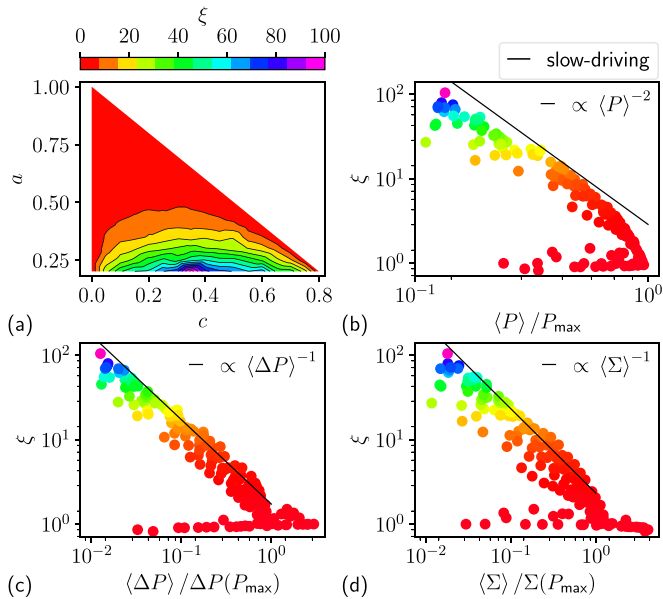
FIG. 4. (a) Contour plot of the SSHE thermodynamic uncertainty relationship ratio ξ as a function of c and a. (b)–(d) log-log plot of ξ, color mapped as in panel (a) as a function of $\langle P \rangle / P_{\max}$, $\langle \Delta P \rangle / \Delta P(P_{\max})$, and $\langle \Sigma \rangle / \Sigma(P_{\max})$, respectively. Every point corresponds to the same RL optimization performed in Fig. 2. The black line is the behavior of ξ derived analytically in the Supplemental Material [93] in the slow-driving regime for small values of $\langle P \rangle$, $\langle \Delta P \rangle$, and $\langle \Sigma \rangle$.

the same color map as in Fig. 4(a). We see that ξ diverges in the limit of low power, low fluctuations, and low entropy production as a power law. Indeed, using the slow-driving approximation, we analytically prove that ξ diverges as $\langle P \rangle^{-2}$, $\langle \Delta P \rangle^{-1}$, and $\langle \Sigma \rangle^{-1}$. Such relations, plotted as black lines, nicely agree with our RL results.

Conclusions. We introduced a general framework to identify Pareto-optimal cycles among power, efficiency, and power

fluctuations in quantum or classical stochastic heat engines, paving the way for their systematic optimization using optimal control techniques, such as the Pontryagin minimum principle [72] or reinforcement learning [73]. As opposed to previous literature reviewed above, we account for the crucial impact of power fluctuations. We then employed RL to optimize a quantum dot-based heat engine, solving its exact finite-time and out-of-equilibrium dynamics, providing us with new physical insights. We observe an abrupt change in Pareto-optimal cycles when switching from the optimization of two objectives where Otto cycles in the fast-driving regime are optimal, to three objectives where the optimal cycles have a finite period. This feature, which shares analogies with the phase transition in protocol space observed in Ref. [96], corresponds to a large enhancement of one of the objectives whereas only slightly decreasing the other ones. Furthermore, we find an exact mapping between Otto cycles in the fast-driving regime and SSHEs, implying that a violation of the thermodynamic uncertainty relation ratio ξ in Eq. (1) requires the optimization of all three objectives. We then find that ξ becomes arbitrarily large in the slow-driving regime. Cycles found with RL display the best features analytically identified in the fast- and slow-driving regimes.

[1] F. Giazotto, T. T. Heikkilä, A. Luukanen, A. M. Savin, and J. P. Pekola, Opportunities for mesoscopics in thermometry and refrigeration: Physics and applications, Rev. Mod. Phys. **78**, 217 (2006).

[2] J. P. Pekola, Towards quantum thermodynamics in electronic circuits, Nat. Phys. **11**, 118 (2015).

[3] S. Vinjanampathy and J. Anders, Quantum thermodynamics, Contemp. Phys. **57**, 545 (2016).

[4] G. Benenti, G. Casati, K. Saito, and R. S.Whitney, Fundamental aspects of steady-state conversion of heat to work at the nanoscale, Phys. Rep. **694**, 1 (2017).

[5] F. Binder, L. Correa, C. Gogolin, J. Anders, and G. Adesso, *Thermodynamics in the Quantum Regime: Fundamental Aspects and New Directions* (Springer International Publishing, Cham, Switzerland, 2019).

[6] E. Geva and R. Kosloff, A quantum-mechanical heat engine operating in finite time. a model consisting of spin-1/2 systems as the working fluid, J. Chem. Phys. **96**, 3054 (1992).

[7] T. Feldmann and R. Kosloff, Performance of discrete heat engines and heat pumps in finite time, Phys. Rev. E **61**, 4774 (2000).

[8] T. Schmiedl and U. Seifert, Efficiency at maximum power: An analytically solvable model for stochastic heat engines, Europhys. Lett. **81**, 20003 (2008).

[9] M. Esposito, R. Kawai, K. Lindenberg, and C. V. den Broeck, Efficiency at Maximum Power of Low-Dissipation Carnot Engines, Phys. Rev. Lett. **105**, 150603 (2010).

[10] M. Esposito, R. Kawai, K. Lindenberg, and C. Van den Broeck, Quantum-dot carnot engine at maximum power, Phys. Rev. E **81**, 041106 (2010).

[11] A. E. Allahverdyan, K. V. Hovhannisyan, A. V. Melkikh, and S. G. Gevorkian, Carnot Cycle at Finite Power: Attainability of Maximal Efficiency, Phys. Rev. Lett. **111**, 050601 (2013).

[12] S. Juergens, F. Haupt, M. Moskalets, and J. Splettstoesser, Thermoelectric performance of a driven double quantum dot, Phys. Rev. B **87**, 245423 (2013).

[13] V. Holubec and A. Ryabov, Efficiency at and near maximum power of low-dissipation heat engines, Phys. Rev. E **92**, 052125 (2015).

[14] V. Holubec and A. Ryabov, Maximum efficiency of low-dissipation heat engines at arbitrary power, J. Stat. Mech.: Theory Exp. (2016) 073204.

[15] B. Karimi and J. P. Pekola, Otto refrigerator based on a superconducting qubit: Classical and quantum performance, Phys. Rev. B **94**, 184503 (2016).

[16] M. Campisi and R. Fazio, The power of a critical heat engine, Nat. Commun. **7**, 11895 (2016).

[17] K. Yamamoto, O. Entin-Wohlman, A. Aharony, and N. Hatano, Efficiency bounds on thermoelectric transport in magnetic fields: The role of inelastic processes, Phys. Rev. B **94**, 121402(R) (2016).

[18] R. Kosloff and Y. Rezek, The quantum harmonic otto cycle, Entropy **19**, 136 (2017).

[19] V. Patel, V. Savsani, and A. Mudgal, Many-objective thermodynamic optimization of Stirling heat engine, Energy **125**, 629 (2017).

[20] V. Cavina, A. Mari, A. Carlini, and V. Giovannetti, Optimal thermodynamic control in open quantum systems, Phys. Rev. A **98**, 012139 (2018).

[21] P. Abiuso and V. Giovannetti, Non-markov enhancement of maximum power for quantum thermal machines, Phys. Rev. A **99**, 052106 (2019).

[22] Y.-H. Ma, D. Xu, H. Dong, and C.-P. Sun, Universal constraint for efficiency and power of a low-dissipation heat engine, Phys. Rev. E **98**, 042112 (2018).

[23] P. A. Erdman, V. Cavina, R. Fazio, F. Taddei, and V. Giovannetti, Maximum power and corresponding efficiency for two-level heat engines and refrigerators: Optimality of fast cycles, New J. Phys. **21**, 103049 (2019).

[24] P. Menczel, T. Pyhäranta, C. Flindt, and K. Brandner, Two-stroke optimization scheme for mesoscopic refrigerators, Phys. Rev. B **99**, 224306 (2019).

[25] J.-F. Chen, C.-P. Sun, and H. Dong, Boosting the performance of quantum otto heat engines, Phys. Rev. E **100**, 032144 (2019).

[26] P. Abiuso and M. Perarnau-Llobet, Optimal Cycles for Low-Dissipation Heat Engines, Phys. Rev. Lett. **124**, 110606 (2020).

[27] V. Cavina, P. A. Erdman, P. Abiuso, L. Tolomeo, and V. Giovannetti, Maximum-power heat engines and refrigerators in the fast-driving regime, Phys. Rev. A **104**, 032226 (2021).

[28] R. A. Vargas-Hernández, R. T. Q. Chen, K. A. Jung, and P. Brumer, Fully differentiable optimization protocols for non-equilibrium steady states, New J. Phys. **23**, 123006 (2021).

[29] P. Terrén Alonso, P. Abiuso, M. Perarnau-Llobet, and L. Arrachea, Geometric optimization of nonequilibrium adiabatic thermal machines and implementation in a qubit system, PRX Quantum **3**, 010326 (2022).

[30] Z. Ye, F. Cerisola, P. Abiuso, J. Anders, M. Perarnau-Llobet, and V. Holubec, Optimal finite-time heat engines under constrained control Phys. Rev. Res. **4**, 043130 (2022).

[31] V. Holubec, An exactly solvable model of a stochastic heat engine: Optimization of power, power fluctuations and efficiency, J. Stat. Mech.: Theory Exp. (2014) P05022.

[32] H. J. D. Miller and M. Mehboudi, Geometry of Work Fluctuations Versus Efficiency in Microscopic Thermal Machines, Phys. Rev. Lett. **125**, 260602 (2020).

[33] T. Denzler and E. Lutz, Power fluctuations in a finite-time quantum carnot engine, Phys. Rev. Res. **3**, L032041 (2021).

[34] S. Saryal and B. K. Agarwalla, Bounds on fluctuations for finite-time quantum otto cycle, Phys. Rev. E **103**, L060103 (2021).

[35] M. Mehboudi and H. J. D. Miller, Thermodynamic length and work optimisation for gaussian quantum states, Phys. Rev. A **105**, 062434 (2022).

[36] A. C. Barato and U. Seifert, Thermodynamic Uncertainty Relation for Biomolecular Processes, Phys. Rev. Lett. **114**, 158101 (2015).

[37] T. R. Gingrich, J. M. Horowitz, N. Perunov, and J. L. England, Dissipation Bounds All Steady-State Current Fluctuations, Phys. Rev. Lett. **116**, 120601 (2016).

[38] P. Pietzonka and U. Seifert, Universal Trade-Off between Power, Efficiency, and Constancy in Steady-State Heat Engines, Phys. Rev. Lett. **120**, 190602 (2018).

[39] G. Guarnieri, G. T. Landi, S. R. Clark, and J. Goold, Thermodynamics of precision in quantum nonequilibrium steady states, Phys. Rev. Res. **1**, 033021 (2019).

[40] A. M. Timpanaro, G. Guarnieri, J. Goold, and G. T. Landi, Thermodynamic Uncertainty Relations from Exchange Fluctuation Theorems, Phys. Rev. Lett. **123**, 090604 (2019).

[41] G. Falasco, M. Esposito, and J.-C. Delvenne, Unifying thermodynamic uncertainty relations, New J. Phys. **22**, 053046 (2020).

[42] H. M. Friedman, B. K. Agarwalla, O. Shein-Lumbroso, O. Tal, and D. Segal, Thermodynamic uncertainty relation in atomic-scale quantum conductors, Phys. Rev. B **101**, 195423 (2020).

[43] J. M. Horowitz and T. R. Gingrich, Thermodynamic uncertainty relations constrain non-equilibrium fluctuations, Nat. Phys. **16**, 15 (2020).

[44] S. Saryal, M. Gerry, I. Khait, D. Segal, and B. K. Agarwalla, Universal Bounds on Fluctuations in Continuous Thermal Machines, Phys. Rev. Lett. **127**, 190603 (2021).

[45] B. K. Agarwalla and D. Segal, Assessing the validity of the thermodynamic uncertainty relation in quantum systems, Phys. Rev. B **98**, 155438 (2018).

[46] K. Brandner, T. Hanazato, and K. Saito, Thermodynamic Bounds on Precision in Ballistic Multiterminal Transport, Phys. Rev. Lett. **120**, 090601 (2018).

[47] K. Ptaszyński, Coherence-enhanced constancy of a quantum thermoelectric generator, Phys. Rev. B **98**, 085425 (2018).

[48] F. Carollo, R. L. Jack, and J. P. Garrahan, Unraveling the Large Deviation Statistics of Markovian Open Quantum Systems, Phys. Rev. Lett. **122**, 130605 (2019).

[49] J. Liu and D. Segal, Thermodynamic uncertainty relation in quantum thermoelectric junctions, Phys. Rev. E **99**, 062141 (2019).

[50] A. Arash Sand Kalaee, A. Wacker, and P. P. Potts, Violating the thermodynamic uncertainty relation in the three-level maser, Phys. Rev. E **104**, L012103 (2021).

[51] A. Rignon-Bret, G. Guarnieri, J. Goold, and M. T. Mitchison, Thermodynamics of precision in quantum nanomachines, Phys. Rev. E **103**, 012133 (2021).

[52] T. Ehrlich and G. Schaller, Broadband frequency filters with quantum dot chains, Phys. Rev. B **104**, 045424 (2021).

[53] M. Gerry and D. Segal, Absence and recovery of cost-precision tradeoff relations in quantum transport, Phys. Rev. B **105**, 155401 (2022).

[54] A. C. Barato and U. Seifert, Cost and Precision of Brownian Clocks, Phys. Rev. X **6**, 041053 (2016).

[55] K. Proesmans and C. V. den Broeck, Discrete-time thermodynamic uncertainty relation, Europhys. Lett. **119**, 20001 (2017).

[56] V. Holubec and A. Ryabov, Cycling Tames Power Fluctuations near Optimum Efficiency, Phys. Rev. Lett. **121**, 120601 (2018).

[57] L. M. Cangemi, V. Cataudella, G. Benenti, M. Sassetti, and G. De Filippis, Violation of thermodynamics uncertainty relations in a periodically driven work-to-work converter from weak to strong dissipation, Phys. Rev. B **102**, 165418 (2020).

[58] P. Menczel, E. Loisa, K. Brandner, and C. Flindt, Thermodynamic uncertainty relations for coherently driven open quantum systems, J. Phys. A: Math. Theor. **54**, 314002 (2021).

[59] J. Lu, Z. Wang, J. Peng, C. Wang, J.-H. Jiang, and J. Ren, Geometric thermodynamic uncertainty relation in a periodically driven thermoelectric heat engine, Phys. Rev. B **105**, 115428 (2022).

[60] T. Koyuk, U. Seifert, and P. Pietzonka, A generalization of the thermodynamic uncertainty relation to periodically driven systems, J. Phys. A: Math. Theor. **52**, 02LT02 (2019).

[61] T. Koyuk and U. Seifert, Operationally Accessible Bounds on Fluctuations and Entropy Production in Periodically Driven Systems, Phys. Rev. Lett. **122**, 230601 (2019).

[62] T. Koyuk and U. Seifert, Thermodynamic Uncertainty Relation for Time-Dependent Driving, Phys. Rev. Lett. **125**, 260604 (2020).

[63] T. Van Vu and Y. Hasegawa, Thermodynamic uncertainty relations under arbitrary control protocols, Phys. Rev. Res. **2**, 013060 (2020).

[64] E. Potanina, C. Flindt, M. Moskalets, and K. Brandner, Thermodynamic Bounds on Coherent Transport in Periodically Driven Conductors, Phys. Rev. X **11**, 021013 (2021).

[65] Y. Hasegawa, Irreversibility, Loschmidt Echo, and Thermodynamic Uncertainty Relation, Phys. Rev. Lett. **127**, 240602 (2021).

[66] H. J. D. Miller, M. H. Mohammady, M. Perarnau-Llobet, and G. Guarnieri, Thermodynamic Uncertainty Relation in Slowly Driven Quantum Heat Engines, Phys. Rev. Lett. **126**, 210603 (2021).

[67] H. J. D. Miller, M. H. Mohammady, M. Perarnau-Llobet, and G. Guarnieri, Joint statistics of work and entropy production along quantum trajectories, Phys. Rev. E **103**, 052138 (2021).

[68] V. Gorini, A. Kossakowski, and E. C. G. Sudarshan, Completely positive dynamical semigroups of N-level systems, J. Math. Phys. **17**, 821 (1976).

[69] G. Lindblad, On the generators of quantum dynamical semigroups, Commun. Math. Phys. **48**, 119 (1976).

[70] H. Breuer and F. Petruccione, *The Theory of Open Quantum systems* (Oxford University Press, New York, 2002).

[71] M. Yamaguchi, T. Yuge, and T. Ogawa, Markovian quantum master equation beyond adiabatic regime, Phys. Rev. E **95**, 012136 (2017).

[72] D. E. Kirk, *Optimal Control Theory: an Introduction* (Dover, New York, 2004).

[73] R. S. Sutton and A. G. Barto, *Reinforcement Larning: An Introduction* (MIT Press, Cambridge, MA, 2018).

[74] B. Sothmann, R. Sánchez, and A. N. Jordan, Thermoelectric energy harvesting with quantum dots, Nanotechnology **26**, 032001 (2015).

[75] J. P. Pekola, B. Karimi, G. Thomas, and D. V. Averin, Supremacy of incoherent sudden cycles, Phys. Rev. B **100**, 085405 (2019).

[76] L. M. Cangemi, M. Carrega, A. De Candia, V. Cataudella, G. De Filippis, M. Sassetti, and G. Benenti, Optimal energy conversion through antiadiabatic driving breaking time-reversal symmetry, Phys. Rev. Res. **3**, 013237 (2021).

[77] P. Salamon and R. S. Berry, Thermodynamic Length and Dissipated Availability, Phys. Rev. Lett. **51**, 1127 (1983).

[78] J. Wang, J. He, and X. He, Performance analysis of a two-state quantum heat engine working with a single-mode radiation field in a cavity, Phys. Rev. E **84**, 041127 (2011).

[79] J. E. Avron, M. Fraas, G. M. Graf, and P. Grech, Adiabatic theorems for generators of contracting evolutions, Commun. Math. Phys. **314**, 163 (2012).

[80] M. F. Ludovico, F. Battista, F. von Oppen, and L. Arrachea, Adiabatic response and quantum thermoelectrics for ac-driven quantum systems, Phys. Rev. B **93**, 075136 (2016).

[81] V. Cavina, A. Mari, and V. Giovannetti, Slow Dynamics and Thermodynamics of Open Quantum Systems, Phys. Rev. Lett. **119**, 050601 (2017).

[82] H. J. D. Miller, M. Scandi, J. Anders, and M. Perarnau-Llobet, Work Fluctuations in Slow Processes: Quantum Signatures and Optimal Control, Phys. Rev. Lett. **123**, 230603 (2019).

[83] M. Scandi and M. Perarnau-Llobet, Thermodynamic length in open quantum systems, Quantum **3**, 197 (2019).

[84] K. Brandner and K. Saito, Thermodynamic Geometry of Microscopic Heat Engines, Phys. Rev. Lett. **124**, 040602 (2020).

[85] B. Bhandari, P. T. Alonso, F. Taddei, F. von Oppen, R. Fazio, and L. Arrachea, Geometric properties of adiabatic quantum thermal machines, Phys. Rev. B **102**, 155407 (2020).

[86] P. Abiuso, H. J. D. Miller, M. Perarnau-Llobet, and M. Scandi, Geometric optimisation of quantum thermodynamic processes, Entropy **22**, 1076 (2020).

[87] A. G. Frim and M. R. DeWeese, Geometric Bound on the Efficiency of Irreversible Thermodynamic Cycles, Phys. Rev. Lett. **128**, 230601 (2022).

[88] J. Eglinton and K. Brandner, Geometric bounds on the power of adiabatic thermal machines, Phys. Rev. E **105**, L052102 (2022).

[89] K. Brandner, K. Saito, and U. Seifert, Thermodynamics of Micro- and Nano-Systems Driven by Periodic Temperature Variations, Phys. Rev. X **5**, 031019 (2015).

[90] K. Brandner and U. Seifert, Periodic thermodynamics of open quantum systems, Phys. Rev. E **93**, 062134 (2016).

[91] O. Movilla Miangolarra, R. Fu, A. Taghvaei, Y. Chen, and T. T. Georgiou, Underdamped stochastic thermodynamic engines in contact with a heat bath with arbitrary temperature profile, Phys. Rev. E **103**, 062103 (2021).

[92] O. Movilla Miangolarra, A. Taghvaei, R. Fu, Y. Chen, and T. T. Georgiou, Energy harvesting from anisotropic fluctuations, Phys. Rev. E **104**, 044101 (2021).

[93] See Supplemental Material at http://link.aps.org/supplemental/ 10.1103/PhysRevResearch.5.L022017 for details on the the

general framework, the reinforcement learning method, and the fast and slow driving regimes.

[94] R. Alicki, The quantum open system as a model of the heat engine, J. Phys. A **12**, L103 (1979).

[95] L. F. Seoane and R. Solé, Multiobjective optimization and phase transitions, in *Proceedings of ECCS 2014*, edited by S. Battiston, F. De Pellegrini, G. Caldarelli, and E. Merelli (Springer, Cham, 2016), pp. 259-270.

[96] A. P. Solon and J. M. Horowitz, Phase Transition in Protocols Minimizing Work Fluctuations, Phys. Rev. Lett. **120**, 180605 (2018).

[97] T. Haarnoja, A. Zhou, K. Hartikainen, G. Tucker, S. Ha, J. Tan, V. Kumar, H. Zhu, A. Gupta, P. Abbeel, and S. Levine, Soft actor-critic algorithms and applications, arXiv:1812.05905.

[98] O. Delalleau, M. Peter, E. Alonso, A. Logut, Discrete and continuous action representation for practical RL in video games, arXiv:1912.11077.

[99] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor, *International Conference on Machine Learning*, Vol. 80 (PMLR, Birmingham, UK, 2018), p. 1861.

[100] P. A. Erdman and F. Noé, Identifying optimal cycles in quantum thermal machines with reinforcement-learning, npj Quantum Inf **8**, 1 (2022).

[101] P. A. Erdman and F. Noé, Driving black-box quantum thermal machines with optimal power/efficiency trade-offs using reinforcement learning, arXiv:2204.04785.

[102] P. Sgroi, G. M. Palma, and M. Paternostro, Reinforcement Learning Approach to Nonequilibrium Quantum Thermodynamics, Phys. Rev. Lett. **126**, 020601 (2021).

[103] I. Khait, J. Carrasquilla, and D. Segal, Optimal control of quantum thermal machines using machine learning, Phys. Rev. Res. **4**, L012029 (2022).

[104] Y. Ashida and T. Sagawa, Learning the best nanoscale heat engines through evolving network topology, Commun. Phys. **4**, 45 (2021).

[105] M. Bukov, A. G. R. Day, D. Sels, P. Weinberg, A. Polkovnikov, and P. Mehta, Reinforcement Learning in Different Phases of Quantum Control, Phys. Rev. X **8**, 031086 (2018).

[106] T. Fösel, P. Tighineanu, T. Weiss, and F. Marquardt, Reinforcement Learning with Neural Networks for Quantum Feedback, Phys. Rev. X **8**, 031084 (2018).

[107] Z. An and D. Zhou, Deep reinforcement learning for quantum gate control, Europhys. Lett. **126**, 60002 (2019).

[108] M. Y. Niu, S. Boixo, V. N. Smelyanskiy, and H. Neven, Universal quantum control through deep reinforcement learning, npj Quantum Inf **5**, 33 (2019).

[109] X.-M. Zhang, Z. Wei, R. Asad, X.-C. Yang, and X. Wang, When does reinforcement learning stand out in quantum control? A comparative study on state preparation, npj Quantum Inf **5**, 85 (2019).

[110] M. Dalgaard, F. Motzoi, J. J. Sørensen, and J. Sherson, Global optimization of quantum dynamics with alphazero deep exploration, npj Quantum Inf **6**, 6 (2020).

[111] J. Mackeprang, D. B. R. Dasari, and J. Wrachtrup, A reinforcement learning approach for quantum state engineering, Quantum Mach. Intell. **2**, 5 (2020).

[112] F. Schäfer, M. Kloc, C. Bruder, and N. Lörch, A differentiable programming method for quantum control, Mach. Learn.: Sci. Technol. **1**, 035009 (2020).

[113] R. Sweke, M. S. Kesselring, E. P. L. van Nieuwenburg, and J. Eisert, Reinforcement learning decoders for fault-tolerant quantum computation, Mach. Learn.: Sci. Technol. **2**, 025005 (2021).

[114] F. Schäfer, P. Sekatski, M. Koppenhöfer, C. Bruder, and M. Kloc, Control of stochastic quantum dynamics by differentiable programming, Mach. Learn.: Sci. Technol. **2**, 035004 (2021).

[115] R. Porotti, A. Essig, B. Huard, and F. Marquardt, Deep reinforcement learning for quantum state preparation with weak nonlinear measurements, Quantum **6**, 747 (2022).

[116] F. Metz and M. Bukov, Self-correcting quantum many-body control using reinforcement learning with tensor networks, arXiv:2201.11790.

[117] L. Coopmans, S. Campbell, G. De Chiara, and A. Kiely, Optimal control in disordered quantum systems, Phys. Rev. Res. **4**, 043138 (2022).