Thermodynamics of precision in quantum nano-machines

Antoine Rignon-Bret,1,2 Giacomo Guarnieri,1 John Goold,1 and Mark T. Mitchison1,∗

1School of Physics, Trinity College Dublin, College Green, Dublin 2, Ireland
2École Normale Supérieure, 45 rue d’Ulm, F-75230 Paris, France

Fluctuations strongly affect the dynamics and functionality of nanoscale thermal machines. Recent developments in stochastic thermodynamics have shown that fluctuations in many far-from-equilibrium systems are constrained by the rate of entropy production via so-called thermodynamic uncertainty relations. These relations imply that increasing the reliability or precision of an engine’s power output comes at a greater thermodynamic cost. Here we study the thermodynamics of precision for small thermal machines in the quantum regime. In particular, we derive exact relations between the power, power fluctuations, and entropy production rate for several models of few-qubit engines (both autonomous and cyclic) that perform work on a quantised load. Depending on the context, we find that quantum coherence can either help or hinder where power fluctuations are concerned. We discuss design principles for reducing such fluctuations in quantum nano-machines, and propose an autonomous three-qubit engine whose power output for a given entropy production is more reliable than would be allowed by any classical Markovian model.

I. INTRODUCTION

Close examination of a small-scale system typically reveals significant fluctuations due to thermal noise. Not only do these fluctuations open a window on otherwise hidden phenomena [1–3], they also exert a decisive influence on the functionality of nanoscale machines — such as atomic [4–6] or molecular [7] motors. On a fundamental level, the occurrence of microscopic fluctuations is inextricably linked to dissipation and the emergence of macroscopic irreversibility, which originates from the same underlying randomness. Yet while the quantitative relation between fluctuations and dissipation has long been understood for equilibrium systems [8], the principles that connect fluctuations to irreversible behaviour far from equilibrium are still in the process of being uncovered [9–13].

Recently, it was discovered [14, 15] that non-equilibrium fluctuations are constrained by dissipation through a rather general class of inequalities known as thermodynamic uncertainty relations (TURs) [16]. Broadly speaking, TURs dictate that the currents which characterise any non-equilibrium scenario, e.g., the heat current powering an engine, must fluctuate by a certain minimum amount that is controlled by the rate of entropy production, such that reduced fluctuations necessitate increased entropy production. This principle has striking consequences for the performance of heat engines governed by classical physics: approaching the ultimate Carnot efficiency at finite average power output is possible only if the power fluctuations diverge [17–19]. TURs also limit the accuracy of autonomous clocks [20–26] and biochemical sensors [27], and can be used to infer difficult-to-measure quantities such as the entropy production of molecular motors [28, 29].

Due to their fundamental and practical importance, a wealth of recent research has aimed at extending TURs beyond their original realms of validity, finding that they apply to classical stochastic systems under remarkably general conditions. Examples include finite observation times [30, 31], discrete-time processes [32, 33], counting observables [34], and feedback protocols [35]. Tighter bounds have been derived [36, 37] and connections have been found with other important concepts of non-equilibrium thermodynamics, including fluctuation theorems [38, 39] and information theory [40–42]. A substantial body of theoretical work has also been devoted to the thermodynamics of precision for quantum systems. In general, the classical TURs can be violated in the presence of quantum coherence, which can boost the reliability of nanoscale thermoelectric generators [43–46]. Quantum generalisations of TURs have been proved for steady-state [47] and cyclic [48] quantum heat engines, for quantum systems under linear-response conditions [49], and for Markovian open quantum systems subjected to continuous weak measurements [50–52]. While these general bounds are useful for understanding fundamental limits, the precise relation between entropy production and non-equilibrium fluctuations must be assessed for each specific system on a case-by-case basis [53–55].

Here, we analyse the thermodynamics of precision for small quantum heat engines that perform work on a load with an infinite-dimensional Hilbert space. Our study is motivated by recent experimental implementations of nanoscale devices whose work output is stored in the vibrations of a mechanical oscillator [5, 6, 56]. At such small scales, energetic fluctuations are unavoidable and may significantly affect performance [57–64]. We show this explicitly by deriving exact equalities connecting the power, its fluctuations, and the rate of entropy production for some basic models of quantum heat engines that have been considered in the literature. We will henceforth refer to these equalities as TUR ratios, since they share the structure and spirit of the TURs.

We begin in Sec. II with the two-qubit autonomous engine introduced by Brunner et al. [60], which forms a minimal template for all quantum absorption machines [22, 65]. We derive TUR ratios describing the power fluctuations of this two-qubit engine and discuss the underlying physical mechanisms that give rise to them. In Sec. III, we use these insights to design a three-qubit autonomous engine that exhibits substantially reduced fluctuations due to quantum-coherent energy transport. However, coherence is not always helpful, as we show in Sec. IV by considering the work output of a cyclic Otto en-

∗ mark.mitchison@tcd.ie
engine with a qubit working medium and a harmonic-oscillator load [5]. In that case, local coherence in the load’s energy eigenbasis lead to increased energetic fluctuations compared to the analogous classical process. Our results reveal general design principles for reducing fluctuations in quantum nanomachines, and contribute towards a deeper understanding of the differences between quantum and classical heat engines. Units where $\hbar = k_B = 1$ are used throughout.

II. MINIMAL AUTONOMOUS TWO-QUBIT ENGINE

A. Two-qubit engine model

Our first example is the minimal model of an autonomous quantum heat engine introduced in Ref. [60], which is depicted schematically in Fig. 1. The machine consists of two qubits with energy splitting $E_1$ and $E_2$, coupled to independent heat reservoirs at the respective temperatures $T_1$ and $T_2$, where $T_2 > T_1$. This two-qubit engine performs work on a load system which is described by an infinite-dimensional ladder of equidistant energy eigenstates separated by an energy $E_v = E_2 - E_1$. The Hamiltonian of the system is given by

$$\hat{H} = \hat{H}_0 + \hat{H}_{\text{int}},$$

where $\hat{H}_0 = \frac{1}{2} \sum_{j=1,2} E_j \sigma_j^z + \hat{W},$ and $\hat{H}_{\text{int}} = g \left( \sigma_1^+ \sigma_2^- \hat{A}^* + \sigma_1^- \sigma_2^+ \hat{A} \right),$ (1) (2)

where $\sigma_j^{x,y,z}$ denotes the standard Pauli operators describing qubit $j = 1, 2$, $\sigma_j^z = (\sigma_j^+ + i \sigma_j^-)/2$ are the spin lowering and raising operators, and we defined the load’s energy operator $\hat{W} = \sum_{n=0}^{\infty} n E_v \hat{n}_w (n)$ and lowering operator $\hat{A} = \sum_{n=\infty}^{0} (n - 1) \hat{n}_w (n)$. Similar tripartite engine models have been proposed in the context of quantum optics [59] and optomechanics [62].

The form of the interaction in Eq. (2) allows energy quanta to flow from the hot qubit to the cold one, but only by simultaneously transferring quanta of energy $E_v = E_2 - E_1$ to the load. This process is more likely than its time-reverse (whereby the load loses energy) so long as entropy is produced in accordance with the second law of thermodynamics. Each quantum of energy transferred to the load leads to entropy changes $\Delta S_1 = E_1/T_1$ and $\Delta S_2 = -E_2/T_2$ associated with the heat exchanged with the cold and hot reservoirs, respectively. Noting the total entropy production of this process by $\chi$, the engine’s operating condition is then given by

$$\chi \equiv \frac{E_1}{T_1} - \frac{E_2}{T_2} \geq 0.$$ (3)

To understand this condition microscopically, we identify a virtual qubit in the composite Hilbert space of the hot and cold qubits [60]. The virtual qubit is defined by the pair of states $|0\rangle_v = |1\rangle_1 |0\rangle_2$ and $|1\rangle_v = |0\rangle_1 |1\rangle_2$, where $|0\rangle_j, |1\rangle_j$ are the eigenstates of $\hat{\sigma}_j^z$. These virtual qubit states are eigenstates of $\hat{H}_0$ with energy splitting $E_v$, which exchange energy with the load via $\hat{H}_{\text{int}}$ (see Fig. 2). In the absence of coupling to the load ($g = 0$), the machine qubits are in equilibrium with their corresponding baths, so that the population of each energy eigenstate obeys the Boltzmann distribution, i.e. $P(|1\rangle_j)/P(|0\rangle_j) = e^{E_j/T_j}$. It follows that the virtual qubit states are populated in the ratio $P(|1\rangle_v)/P(|0\rangle_v) = e^{E_v/T_v} - e^{-E_v/T_v}$, where $T_v = -E_v/\chi$ is the virtual temperature [66]. When $\chi$ is large and positive, the virtual temperature is small and negative and the virtual qubit populations are inverted, i.e. $P(|1\rangle_v) \gg P(|0\rangle_v)$. Turning on a weak interaction $g \neq 0$ couples the virtual qubit to the load by driving transitions of the form $|1\rangle_v |n\rangle_w \leftrightarrow |0\rangle_v |n + 1\rangle_w$. The population inversion of the virtual qubit biases the forward transition in favour of the reverse one, causing the load’s energy to increase over time. In this way, the engine’s operation can be understood as “thermalisation” of the load but with a negative virtual temperature.

B. Thermodynamics of precision for the two-qubit engine

We now turn to the explicit description of the engine’s dynamics. Following Refs. [60, 65], we model the thermalis-
ing effect of the reservoirs by assuming that in each small time interval $dt$, the qubits either evolve coherently under the Hamiltonian $\hat{H}$ or one of them is randomly reset to a local thermal state $\hat{\rho}_j = e^{-\beta_j G_j} / Z_j$, with $\beta_j = 1 / T_j$ and $Z_j = \text{Tr}\left(e^{-\beta_j G_j}\right)$. For each qubit, this resetting is assumed to occur with probability $p$ per unit time and is described by the map $\hat{\rho} \rightarrow \hat{\tau} \otimes \text{Tr}\left(\hat{\rho}\right)$. Taking the limit $dt \rightarrow 0$, the corresponding ensemble dynamics is given by the master equation

$$\frac{d\hat{\rho}}{dt} = -i[\hat{\mathcal{H}}, \hat{\rho}] + \sum_{j=1,2} p \left(\hat{\tau}_j \otimes \text{Tr}\left(\hat{\rho}\right) - \hat{\rho}\right)$$

(4)

On the second line, the reset master equation is expressed explicitly in Lindblad form, where the dissipation super-operator is defined by $\mathcal{D}[\hat{\rho}] = \hat{L} \cdot \hat{\rho} \cdot \hat{L}^\dagger - \frac{1}{2} \{\hat{L}, \hat{L}^\dagger\}$, the gain and decay rates are $\gamma_j = pe^{-\beta_j G_j} / Z_j$, and $\gamma_j = p / 4$ is an effective local dephasing rate. The description of thermalisation in terms of local processes implicitly assumes that the coupling strength $g$ is not much larger than other energy or frequency scales, in particular $E_j, T_j$ and $p$.

The useful output of the two-qubit heat engine is quantified by the energy transferred to the load. This energy transfer is stochastic due to the inevitable fluctuations induced by the coupling to thermal reservoirs. We denote the mean energy of the load by $W = \langle \hat{W} \rangle$ and its variance by $\Delta_W = \langle \hat{W}^2 \rangle - W^2$.

In the long-time limit, the machine reaches a non-equilibrium steady state (NESS) characterised by stationary power $\dot{W} = dW / dt$ and power fluctuations $\Delta_W = d\Delta_W / dt$. Our aim is to relate this energetic output with the rate of entropy production,

$$\dot{\Sigma} = -\beta_1 \hat{Q}_1 - \beta_2 \hat{Q}_2,$$

(6)

where $\hat{Q}_j = \text{Tr}[\hat{\mathcal{H}} \cdot \hat{\rho}]$ is the heat current entering the system from bath $j$ and $\mathcal{D} \cdot \hat{\rho} = p \left(\hat{\tau}_j \otimes \text{Tr}\left(\hat{\rho}\right) - \hat{\rho}\right)$ is the corresponding dissipator. Note that, since the fluctuating energy transfer to the load is identified with work output, we quantify entropy production in terms of the heat flux only and ignore the additional contribution associated with the load’s growing von Neumann entropy [13, 67].

To relate the machine’s fluctuating power output to the entropy production, we exploit the exact solution of the master equation derived in Ref. [60], which is briefly described here and detailed fully in Appendix A. The state of the virtual qubit is characterised by the following two observables:

$$\hat{Z} = |1\rangle \langle 1| - |0\rangle \langle 0| = \frac{1}{2} \left(\hat{\sigma}_-^2 - \hat{\sigma}_+^2\right),$$

(7)

$$\hat{N} = |1\rangle \langle 1| + |0\rangle \langle 0| = \frac{1}{2} \left(1 - \hat{\sigma}_-^2 \hat{\sigma}_+^2\right).$$

(8)

Respectively, these yield the mean bias (population inversion) $\langle \hat{Z} \rangle$ and the mean occupation $\langle \hat{N} \rangle$ of the virtual qubit states. In the absence of coupling to the load, these quantities take the values $\langle \hat{Z} \rangle_{\text{eq}} = \text{Tr}(\hat{\mathcal{H}_1} \otimes \hat{\tau}_2)$ and $\langle \hat{N} \rangle_{\text{eq}} = \text{Tr}(\hat{\mathcal{N}_1} \otimes \hat{\tau}_2)$, where

$$\langle \hat{Z} \rangle_{\text{eq}} = -\tanh(\beta_1 E_1 / 2) \langle \hat{N} \rangle_{\text{eq}},$$

(9)

$$\langle \hat{N} \rangle_{\text{eq}} = \frac{1}{2} \left[1 - \tanh(\beta_1 E_1 / 2) \tanh(\beta_2 E_2 / 2)\right],$$

(10)

which describe a qubit in equilibrium at inverse temperature $\beta_1 = 1 / T_1$, and with a total normalisation $\langle \hat{N} \rangle_{\text{eq}}$ less than unity (because the qubit is virtual). For finite $g$, the virtual qubit drives a coherent current to the load described by the dimensionless operator

$$\hat{C} = i \left(\hat{\sigma}_-^2 \hat{\sigma}_+^2 \hat{A} - \hat{\sigma}_+^2 \hat{\sigma}_-^2 \hat{A}^\dagger\right).$$

(11)

The average power delivered to the load is then $\hat{W} = gE_c \langle \hat{C} \rangle$. In the NESS, all currents are proportional to the bias, $\hat{Q}_j \propto W \propto \langle \hat{C} \rangle \propto \langle \hat{Z} \rangle$, while the load’s energetic fluctuations are determined by both $\langle \hat{Z} \rangle$ and $\langle \hat{N} \rangle$. Explicitly, we obtain the asymptotic solutions

$$W = \Gamma_2 E_c \langle \hat{Z} \rangle_{\text{eq}},$$

(12)

$$\dot{\Sigma} = \Gamma_2 \chi \langle \hat{Z} \rangle_{\text{eq}},$$

(13)

$$\Delta_W = \Gamma_2 E_c^2 \left[\langle \hat{N} \rangle_{\text{eq}} - \frac{2\Gamma_2 (2p^2 + g^2)}{p(p^2 + 2g^2)} \langle \hat{Z} \rangle_{\text{eq}}^2\right],$$

(14)

where $\Gamma_2 = g^2 p / (p^2 + 2g^2)$ is the characteristic rate of energy transfer for the two-qubit engine. Due to the proportionality of the power and heat currents, the engine’s efficiency is given by

$$\eta = \frac{W}{Q_2} = 1 - \frac{E_1}{E_2}.$$

(15)

The Carnot bound, $\eta \leq \eta_C = 1 - T_1 / T_2$, follows directly from condition (3).

It is now straightforward to derive the TUR ratio

$$\frac{\Delta_W}{W^2} \dot{\Sigma} = \chi \left[\coth(\chi / 2) - \frac{2\Gamma_2 (2p^2 + g^2)}{p(p^2 + 2g^2)} \langle \hat{Z} \rangle_{\text{eq}}\right] \geq 2.$$

(16)

This relation encapsulates the trade-off between the precision of work deposition and its associated thermodynamic cost. For any given set of finite temperatures and qubit energies, Eq. (16) gives the necessary conditions to minimise fluctuations in the engine’s power output at fixed entropy production. The optimal operating point in this respect is $g = p$ (see Appendix B), which is effectively an impedance-matching condition between the bath-engine and engine-load couplings.

The inequality (16), which is proved in Appendix B, is of the same form as the classical TUR for steady-state currents [14, 15]. We emphasise, however, that the conditions for the validity of the classical TUR do not hold here due to the presence of quantum coherences. To appreciate the importance of the bound (16) for engine performance, we follow Ref. [30] and rewrite it as

$$\Delta_W \geq \frac{2T_1 \eta}{\eta_C - \eta} \hat{W},$$

(17)

where we have used the steady-state energy balance equation $W = \hat{Q}_1 + \hat{Q}_2$ to express the entropy production rate (6) in terms of the efficiency (15). The above inequality implies that approaching the Carnot efficiency at finite power is possible only by allowing the fluctuations to diverge. For the two-qubit engine, the Carnot point corresponds to $\chi \rightarrow 0$. 
which saturates the bound in Eq. (16). The virtual qubit’s bias \( \langle \hat{Z} \rangle_{eq} \propto \tanh(\chi/2) \) tends to zero in this limit but its normalisation \( \langle \hat{N} \rangle_{eq} \) does not. Therefore, when the two-qubit engine operates at Carnot efficiency, its power output is zero on average but has non-zero fluctuations. We note that this behaviour was already discussed by Brunner et al. [60] in terms of the qualitative notion of the “strength of work”. Our results (16) and (17) give this notion a clear quantitative meaning.

The emergence of a TUR-like bound for this small quantum engine is not surprising, since the load’s behaviour has some similarities with a classical stochastic process. In particular, the variance of the load’s energy grows linearly in time, as expected for diffusive dynamics. Indeed, for \( g \ll p \), the load’s evolution can be accurately approximated by a classical random walk \([21, 22]\), with probabilities per unit time of an upward or downward step related by \( p_\uparrow = e^{\chi}p_\downarrow \). In this limit, the second term in Eq. (16) can be neglected and one recovers the classical random walk, which is the prototypical model first used by Barato and Seifert \([14]\) to illustrate the thermodynamics of precision.

Conversely, at finite coupling \( g \), the fluctuations are reduced relative to the classical random-walk case. To clarify this, it is instructive to rewrite Eq. (16) as

\[
\frac{\dot{W}}{W^2} = \chi \coth(\chi/2) \frac{\langle \hat{N} \rangle - 3\langle \hat{C} \rangle^2}{\langle \hat{N} \rangle_{eq}}. \tag{18}
\]

This form highlights two different ways in which a coherent coupling between the virtual qubit and the load reduces the TUR ratio. First, the coupling depletes the occupation of the virtual qubit below its equilibrium value, \( \langle \hat{N} \rangle < \langle \hat{N} \rangle_{eq} \). Considering the virtual qubit as an effective heat reservoir at temperature \( T_r \), this depletion occurs because the reservoir is not macroscopic and thus experiences strong back-action from its coupling to the system. Second, the coupling reduces the TUR ratio by establishing a finite current \( \langle \hat{C} \rangle \), which is associated with quantum coherences in the energy eigenbasis of the coupled qubit-load system \([68]\). Eq. (18) thus suggests two ways in which small thermal machines might overcome the constraints of the classical TUR: either by harnessing small or non-Markovian heat sources, or by exploiting quantum coherences.

C. Effect of local dephasing

Having established the connection between power, fluctuations and dissipation, it is natural to ask whether a more propitious relationship between these quantities can be arranged. As shown by Eq. (18), the presence of quantum coherence can reduce the power fluctuations for a given entropy production rate. We now show that this effect is enhanced by considering a different dissipation model, in which the local dephasing terms proportional to \( \gamma^r \) in Eq. (4) are neglected. We thus consider the master equation

\[
\frac{d\hat{\rho}}{dt} = -i[\hat{H}, \hat{\rho}] + \sum_{j=1,2} \left( \gamma_j^r \mathcal{D}[\sigma_j^\uparrow] + \gamma_j^l \mathcal{D}[\sigma_j^\downarrow] \right) \hat{\rho}. \tag{19}
\]

A local Lindblad equation of this form can be derived from a time-independent system-reservoir interaction under the assumption of weak coupling \( g \) relative to the temperatures and the local qubit and load energies \([69]\). In order to facilitate comparison with the reset model of Sec. II B, we continue to parametrise the gain and loss rates as \( \gamma^r = pe^{\chi}E_r/2 \).

The dynamics is solved to find the NESS corresponding to Eq. (19) in Appendix A. We find

\[
W = \Gamma_1^r E_r \langle \hat{Z} \rangle_{eq}, \tag{20}
\]

\[
\Sigma = \Gamma_2^r \langle \hat{Z} \rangle_{eq}, \tag{21}
\]

\[
\frac{\Delta \dot{W}}{\dot{W}^2} = \chi \left[ \coth(\chi/2) - \frac{\Gamma_2^r(5p^2 + 4g^2)}{p(p^2 + 4g^2)} \langle \hat{Z} \rangle_{eq}^2 \right]. \tag{22}
\]

The results are very similar to the reset model, Eqs. (12)–(14), but with a modified characteristic rate of energy flux, \( \Gamma_2^r = 2g^2p/(p^2 + 4g^2) \), and somewhat reduced fluctuations. The TUR ratio then follows as

\[
\frac{\Delta \dot{W}}{\dot{W}^2} = \chi \left[ \coth(\chi/2) - \frac{5p^2 + 4g^2}{p(p^2 + 4g^2)} \langle \hat{Z} \rangle_{eq}^2 \right]. \tag{23}
\]

Due to its different dependence on the coefficients \( p \) and \( g \), this quantity can take smaller values than the corresponding TUR ratio (16) for the reset model, given the same temperatures and qubit energies. We show in Appendix C.1 that the bound \( \Delta \dot{W} \Sigma /W^2 \geq 1.982 \) holds, which allows for fluctuations below the classical TUR. However, we have found that such violations are typically extremely small and occur only in a very limited region of the parameter space where \( \beta \approx 1 \).

The TUR ratios for the different 2QE models are compared in Fig. 3.
III. AUTONOMOUS THREE-QUBIT ENGINE

A. Three-qubit engine model

As we have seen, fluctuations in the power output may be reduced in the presence of coherent energy transport, in accordance with previous studies [43–45]. We now use this insight to design an autonomous quantum nano-machine with a more reliable power output than would be classically allowed by the TUR. In particular, we modify the two-qubit engine by adding a third qubit that intermediates the flow of energy from the baths to the load, as depicted in Fig. 4. This shifts the coherent coupling between engine and load further away from the decohering effect of the thermal baths, thereby reducing power fluctuations.

The Hamiltonian of the model is  $\hat{H} = \hat{H}_0 + \hat{V} + \hat{H}_{\text{int}}$, where

$$\hat{H}_0 = \frac{1}{2} \sum_{j=1}^{3} E_j \hat{\sigma}_j^z + \hat{W},$$

$$\hat{V} = k \left( \hat{\sigma}_1^+ \hat{\sigma}_2^- \hat{\sigma}_3^- + \hat{\sigma}_1^- \hat{\sigma}_2^+ \hat{\sigma}_3^+ \right),$$

$$\hat{H}_{\text{int}} = g \left( \hat{\sigma}_3^- \hat{A}_1^+ + \hat{\sigma}_3^+ \hat{A}_1 \right),$$

with $\hat{H}_0$ the free Hamiltonian, $\hat{V}$ the interaction between the three qubits, and $\hat{H}_{\text{int}}$ the engine’s coupling to the load. We take $E_3 = E_2 - E_1 = E_v$ to ensure that all interactions are resonant. As before, we assume that qubits 1 and 2 are locally coupled to thermal baths at temperatures $T_1$ and $T_2$, respectively. For $g = 0$, the additional qubit thus thermalises to the virtual temperature, with bias given by

$$\langle \hat{\sigma}_3^z \rangle_{\text{eq}} = -\tanh(\beta, E_v/2).$$

(27)

This is analogous to Eq. (9) but with unit normalisation (i.e. $\langle \hat{N} \rangle = 1$) because the qubit is physical, not virtual.

To model the dynamics for finite $g$, we simplify the problem by assuming that the local thermalisation rate is much larger than both the coherent couplings $k$ and $g$. As a result, the rapidly damped qubits behave approximately like memoryless thermal reservoirs on the slow timescale over which energy is transported between the engine and the load. As shown in Appendix C, these qubits can be perturbatively eliminated under the Born-Markov approximation. The result is an effective master equation describing the joint state of qubit 3 and the load, which reads as

$$\frac{d \hat{\rho}}{dt} = -i[\hat{H}_0 + \hat{H}_{\text{int}}, \hat{\rho}] + \gamma^+ D[\hat{\sigma}_3^+] \hat{\rho} + \gamma^- D[\hat{\sigma}_3^-] \hat{\rho},$$

(28)

where $\hat{H}_0 = E_v \hat{\sigma}_3^z / 2 + \hat{W}$ and the gain and decay rates obey $\gamma^+ / \gamma^- = e^t$. It is convenient to parametrise the effective dissipation rate by the parameter $p = \gamma^+ + \gamma^-$, as in Sec. II. Explicit expressions for $\gamma^\pm$ can be found in Appendix C assuming an underlying reset thermalisation model.

B. Thermodynamics of precision for the three-qubit engine

The dynamics under Eq. (28) can be solved analytically to find the power, fluctuations and entropy production in the

![Fig. 4. Schematic of the autonomous three-qubit engine. The energy flow from the baths is intermediated by an additional qubit resonant with the load. This boosts quantum coherence associated with energy transport, allowing for reduced relative power fluctuations.](image)

NESS. As shown in Appendix C, the results are given by

$$\dot{W} = \Gamma_3 E_v \langle \hat{\sigma}_3^z \rangle_{\text{eq}},$$

$$\dot{\Sigma} = \Gamma_3 \chi \langle \hat{\sigma}_3^z \rangle_{\text{eq}},$$

$$\dot{\Lambda}_W = \Gamma_3 E_v^2 \left[ 1 - \frac{6 \Gamma_3 p}{p^2 + 8 g^2} \langle \hat{\sigma}_3^z \rangle_{\text{eq}}^2 \right].$$

(31)

where $\Gamma_3 = 4 g^2 p/(p^2 + 8 g^2)$ is the characteristic energy transfer rate for the three-qubit machine. As for the two-qubit engine, the proportionality between power and heat currents implies the ideal efficiency $\eta = 1 - E_1 / E_2$.

Eqs. (29)–(31) are now easily combined with Eq. (27) to obtain

$$\frac{\dot{\Lambda}_W}{W^2} \Sigma = \chi \coth(\chi/2) \left[ 1 - \frac{6 \Gamma_3 p}{p^2 + 8 g^2} \tanh^2(\chi/2) \right] \geq 1.245 \ldots,$$

(32)

where the lower bound is derived in Appendix B. This inequality allows the TUR ratio to take values significantly lower than the classical Markovian result (cf. Eq. (16)). As we show in Fig. 3, such values are achievable: a judicious choice of the coupling ratio $g/p$ leads to substantially reduced power fluctuations relative to the classical TUR. The significance of this finding can be appreciated by rewriting Ineq. (32) in terms of power and efficiency, as in Ineq. (17). This shows that, for a given average engine performance, the power fluctuations can be reduced to at most 1.245/2 ≈ 62% of those produced by a machine governed by the classical TUR. We note that the quantum steady-state TUR derived in Ref. [47] is always satisfied by Eq. (32).

The above result demonstrates that adding an extra stage to the transport pathway from engine to load can boost quantum coherence and thereby reduce fluctuations in the machine’s output. Similar effects have been reported in the context of thermoelectric devices [43, 44] and autonomous quantum clocks [24]. In the present model, these advantages come at the cost of reduced power output compared to the two-qubit autonomous engine because the optimal coupling regime is $g \ll p$, which by assumption is much smaller than the rate of thermalisation (see Appendix C). However, we stress that
FIG. 5. Schematic depiction of an Otto cycle where a qubit working medium is driven by a harmonic oscillator. The harmonic motion of the oscillator, stylised here as a coiled spring of extension \( x \), modulates the level spacing of the qubit periodically. The qubit couples to hot and cold baths every half-period, thus changing the excited-state occupation probability (grey circles). The difference in heat absorbed from the hot and cold baths is converted into work done on the spring, driving oscillations of increasing amplitude.

this non-essential approximation was made only to simplify the analytical treatment; it is possible that quantum-enhanced reliability can be obtained in strong-coupling regimes of high power, although we leave this question for future work.

IV. SINGLE-SPIN OTTO ENGINE WITH A HARMONIC OSCILLATOR FLYWHEEL

A. Spin-oscillator engine cycle

As our final example, we investigate the thermodynamics of precision for a four-stroke heat engine comprising a single qubit working medium coupled to a harmonic oscillator load. Our model is inspired by the experiment reported in Ref. [5], where the qubit comprises two Zeeman spin levels of a trapped ion and the oscillator corresponds to the ion’s centre-of-mass degree of freedom. The qubit undergoes an Otto cycle driven by the harmonic motion of the oscillator, which therefore acts as a quantised flywheel [63]. Closely related engine models have also been discussed in Refs. [57, 61].

We consider a simple linear coupling between the qubit and oscillator, described by the Hamiltonian

\[
\hat{H} = \frac{\omega_j}{2} \sigma_z + \omega_0 \hat{a}^\dagger \hat{a} + \frac{1}{2} \omega_0 d \hat{\sigma}_z (\hat{a} + \hat{a}^\dagger),
\]  (33)

where \( \hat{\sigma}_z \) is a Pauli operator describing the qubit, \( \hat{a} \) and \( \hat{a}^\dagger \) are canonical ladder operators for the harmonic oscillator, and \( d \) is a dimensionless parameter that quantifies the qubit-oscillator coupling strength. A semi-classical depiction of the engine cycle is shown in Fig. 5. The qubit is coupled alternately to cold and hot baths at temperatures \( T_1 \) and \( T_2 \) every half period, \( \delta t/2 = \pi/\omega_0 \). In between these isochoric strokes, the qubit is decoupled from the baths and the system evolves freely under the Hamiltonian (33). The changing population of the qubit during the isochores generates an effective periodic driving force on the oscillator, proportional on average to \( \langle \hat{\sigma}_z \rangle \). This drives increasingly large oscillations of the flywheel, which corresponds to the useful work output of the engine. In turn, the resulting harmonic motion of the oscillator modulates the effective energy splitting of the qubit by an amount proportional to the displacement \( (\hat{a} + \hat{a}^\dagger) \).

We focus on the regime of large qubit frequency (\( \omega_j \gg \omega_0 \)), weak qubit-oscillator coupling (\( d \ll 1 \)), and correspondingly small oscillator displacements. We also assume for simplicity that the isochores have a negligible duration and result in perfect thermalisation of the qubit. The heat absorbed during the cold and hot isochores is thus given approximately by the change in the qubit’s mean energy, i.e.

\[
Q_1 = (p_1 - p_2) \left( \omega_j - \omega_0 d (\hat{a} + \hat{a}^\dagger) \right),
\]  (34)

\[
Q_2 = (p_2 - p_1) \left( \omega_j + \omega_0 d (\hat{a} + \hat{a}^\dagger) \right),
\]  (35)

where \( p_j = e^{-\beta_j \omega_j/2} \sum_j \) is the qubit’s excited-state population at temperature \( T_j = 1/\beta_j \), while \( (\hat{a} + \hat{a}^\dagger) \) is the maximum displacement amplitude. Hence, the mean energy transferred to the flywheel per cycle is approximately

\[
W_{\text{cyc}} = Q_1 + Q_2 = 2\omega_0 d (p_2 - p_1) (\hat{a} + \hat{a}^\dagger).
\]  (36)

For the purposes of the present study, we consider this energy to be the engine’s useful work output. We note that, since the displacement increases on each cycle, the engine’s power output increases over time and the cycle is not closed.

B. Thermodynamics of precision for a random walk in phase space

Thermal fluctuations of the qubit lead to a stochastic transfer of energy to the flywheel. We now relate the corresponding fluctuations of the engine’s work output to the entropy production. Under the assumption of fast isochores, the flywheel dynamics can be modelled as a random walk in phase space [5], as depicted in Fig. 6. Depending on whether the qubit occupies its ground or excited state, the oscillator experiences a force in the positive or negative direction. If the state of

![Image](6)

**FIG. 6.** The flywheel dynamics can be modelled as a discrete-time random walk in phase space, where each value of \( \alpha \) represents a coherent state of the oscillator. The state of the qubit can be visualised as a spin pointing up or down, whose orientation determines the direction of the force on the oscillator (inset). If the spin flips its orientation after half the cycle, the effective driving force is resonant with the oscillatory motion, thus changing the displacement in phase space by \( \pm 2d \). If the spin’s orientation does not change, the effective force is constant and the displacement is unaltered after a full cycle.
the qubit changes after one half-period, the force is resonant with the motion and a displacement of \( \pm 2d \) results after a full period. The probabilities for a positive or negative displacement to occur are respectively given by \( p_+ = p_2(1 - p_1) \) and \( p_- = p_1(1 - p_2) \), while with probability \( p_0 = 1 - p_+ - p_- \) there is no displacement. The ratio of the forward and backward rates is given by \( p_+/p_- = e^t \), where, by analogy with Eq. (3), we define the bias parameter

\[
\chi = (\beta_1 - \beta_2)\alpha_2 \geq 0 .
\]

Assuming the oscillator is initialised in its ground state, after \( N \) engine cycles (i.e., \( N \) oscillation periods) it is in a coherent state \( |\alpha_N\rangle \), where \( \alpha_N \) is a real random variable representing the distance travelled after \( N \) steps of the random walk. Normal-ordered quantum expectation values may thus be computed as \( \langle \hat{a}^\dagger \hat{a}^\dagger \hat{a} \hat{a} \rangle = \mathbb{E}[e^{i\alpha_N}] \), where \( \mathbb{E}[\bullet] \) represents an average over random-walk trajectories. Such averages can be found systematically by taking derivatives of the moment generating function \( G_N(s) = \mathbb{E}[e^{s\alpha_N}] \) at \( s = 0 \). Since each step of the random walk is independent and identically distributed, we write \( G_N(s) = [G_1(s)]^N \), where the generating function for one cycle is [5]

\[
G_1(s) = 1 + 2 \sinh(ds) \left( p_+ e^{ds} - p_- e^{-ds} \right) .
\]

The total energy change of the flywheel after \( N \) cycles is \( \Delta W = \omega_0 \hat{\vartheta} \) for large \( N \), and \( \omega_0 \) is the natural angular frequency of the flywheel.

The change in energy over a single cycle is therefore \( \Delta W_{\text{cyc}} = W_{N+1} - W_N \approx 4\omega_0 d(p_+ - p_-)\langle \hat{\vartheta} \rangle \), in exact agreement with Eq. (36). The fluctuations of the flywheel's energy are given by \( \Delta W_{\text{fl}} = \omega_0^2 \hat{\vartheta}^2 - W_{N-1}^2 \). Explicitly, we find that

\[
\Delta W_{\text{fl}} = 64\omega_0^2 d^2(p_+ - p_-)^2 \left[ p_+ + p_- - (p_+ - p_-)^2 \right] N^3 + O(N^5) .
\]

This quadratic increase in energy with \( N \) follows from a linear growth of displacement, i.e., \( \langle \hat{\vartheta} \rangle = 2d(p_+ - p_-)N \), as expected for a biased random walk over coherent states. We note that the change in energy over a single cycle is therefore \( \Delta W_{\text{cyc}} = W_{N+1} - W_N \approx 4\omega_0 d(p_+ - p_-)\langle \hat{\vartheta} \rangle \), in exact agreement with Eq. (36). The fluctuations of the flywheel's energy are given by \( \Delta W_{\text{fl}} = \omega_0^2 \hat{\vartheta}^2 - W_{N-1}^2 \). Explicitly, we find that

\[
\Delta W_{\text{fl}} = 64\omega_0^2 d^2(p_+ - p_-)^2 \left[ p_+ + p_- - (p_+ - p_-)^2 \right] N^3 + O(N^5) .
\]

The entropy production after \( N \) cycles is given by \( \Sigma_N = -N(\beta_1 Q_1 + \beta_2 Q_2) \). Using Eqs. (34) and (35), this can be approximated to leading order by

\[
\Sigma_N \approx (\beta_1 - \beta_2)(p_+ - p_-)\omega_0 N,
\]

since the bare Zeeman energy \( \omega_0 \) is assumed to dominate the contribution from the oscillator displacement.

To obtain an expression connecting power fluctuations and entropy production rate, we convert Eqs. (39)–(41) into coarse-grained rates of change over an engine cycle, e.g. \( \dot{W} = (W_{N+1} - W_N)/\delta t \), with \( \delta t = 2\pi/\omega_0 \). We thus obtain the TUR ratio for the flywheel at leading order for large \( N \): \[
\frac{\dot{W}}{W^2} = 3\chi \left[ \coth(\chi/2) - (p_+ - p_-) \right] + O(N^{-1}) .
\]

This differs in two important respects from the analogous expression \( \chi \coth(\chi/2) \), which holds for the classical, continuous-time random walk [14]. The second term inside the square brackets of Eq. (42), which reduces the relative fluctuations, appears due to the discrete-time nature of the evolution under consideration. This term vanishes in the limit of a continuous-time process where \( p_+ \sim \delta t \to 0 \). This illustrates the fact that a continuous-time random walk exhibits enhanced fluctuations because of the random timing of the steps [33]. The other important feature in Eq. (42) is the overall factor of three multiplying the expression. The physical origin of this factor is the fact that the random walk takes place over coherent states in phase space, rather than energy eigenstates in Fock space. Even a pure coherent state \( |\alpha\rangle \) has a finite energy uncertainty given by \( \langle \alpha | \hat{\vartheta}^2 | \alpha \rangle - \langle \alpha | \hat{\vartheta} | \alpha \rangle^2 = |\alpha|^2 \). The growth of energetic fluctuations is thus faster than would be obtained for a random walk between Fock states [70].

V. DISCUSSION

When considering heat engines at the smallest scales, the fluctuations in output power may be as important as the average performance. Here we have derived explicit expressions for these fluctuations, relating them to the average rate of entropy production for several quantum heat engine models of current interest. Our results exemplify the entropic penalty associated with a reliable power output. In particular, just as for many classical heat engines [18, 19, 48, 71–73], approaching the Carnot efficiency (i.e. zero entropy production) at finite power comes at the cost of catastrophically large fluctuations. This conclusion holds for both autonomous (Secs. II and III) and cyclic (Sec. IV) few-qubit engines.

Interestingly, however, quantum mechanics opens the possibility of reducing an engine’s relative power fluctuations below the level allowed by classical stochastic thermodynamics. In particular, Eq. (18) clarifies two different ways in which autonomous few-qubit machines can surpass the classical TUR. Either one can reduce the occupation \( |\hat{N}| \) of the virtual qubit, or one can boost the coherent current \( \langle \hat{N} \rangle \). This conclusion should apply quite generally to more complicated multi-level machines, which can also be analysed in terms of virtual qubits [60, 74, 75]. We demonstrated this principle explicitly by considering two variants of the two-qubit engine in Secs. II C and III. Both of these exhibit smaller power fluctuations because the effect of decoherence is reduced, either as a consequence of the thermalisation model (Sec. II C) or the geometry of the engine itself (Sec. III). This suggests a general design principle for reducing fluctuations in autonomous thermal machines by boosting coherent transport, e.g. via the introduction of additional stages in the transport pathway between engine and load [24, 43]. We note that, according to Eq. (18), it may also be possible to tame power fluctuations by designing the engine so that the virtual qubit’s occupation is reduced, yet we have not found explicit models where this is possible.

Nevertheless, quantum coherence can lead to drawbacks as well as advantages. We showed in Sec. IV that an engine
which drives transitions between coherent states of the load suffers from greater fluctuations. This is simply because coherent superpositions of energy eigenstates have an intrinsic energy uncertainty above and beyond the thermodynamic uncertainty introduced by coupling to heat reservoirs. We note, however, that such coherences do have some potential to be extracted as useful work with an appropriate protocol [76]. A careful consideration of this problem [64] would lead us to evaluate the ergotropy [77] or the non-equilibrium free energy [78] of the load, as opposed to its mere energy statistics, a question that we leave for future work.

From a more foundational perspective, our findings highlight the importance of fluctuations in certifying the non-classical behaviour of quantum thermal machines [79–84]. In contrast to an engine’s average performance [85], power fluctuations that violate a classical TUR cannot be emulated by any classical (Markovian) stochastic model and thus constitute an unambiguous quantum advantage. We note that the use of fluctuations to tease out the quantum character of dynamical processes has a distinguished history, e.g. in the field of quantum optics [86, 87]. We therefore hope that our work will stimulate further research on the general characterisation and effective suppression of fluctuations in quantum thermal machines.

ACKNOWLEDGMENTS

We are grateful to Stephen R. Clark and Paul Skrzypczyk for enlightening discussions. We acknowledge funding from the European Research Council Starting Grant ODYSSEY (G. A. 758403). JG is supported by a SFI-Royal Society University Research Fellowship.

Appendix A: Solution for the two-qubit engine

1. Reset model

For completeness, in this Appendix we detail the solution of the two-qubit engine modelled by the reset master equation. The results are equivalent to those already obtained by Brunner et al. [60], but it is convenient to recast them here in our notation. Starting from the master equation [Eq. (4)] written as $d\rho/ dt = - i[\hat{H}, \rho] + \sum_{j=1,2} D_j \rho$, the dynamics of any operator $\hat{O}$ in the Heisenberg picture is determined by the equation of motion $d\hat{O}/ dt = i[\hat{H}, \hat{O}] + \sum_{j=1,2} D_j \hat{O}$, where the adjoint dissipator is defined implicitly by $\text{Tr}[\hat{D}_j^\dagger \hat{O}] = \text{Tr}[\hat{\rho} \hat{D}_j^\dagger \hat{O}]$; see Ref. [88] for details.

The Heisenberg equation for the load Hamiltonian is simply $d\hat{W}/ dt = g E, \hat{C}$, where the current operator is defined in Eq. (11). The state of the engine qubits is determined by the bias [Eq. (7)] and occupation [Eq. (8)] of the virtual qubit, as well as the total spin $S = \frac{1}{2}(\hat{\sigma}_+^2 + \hat{\sigma}_-^2)$ (equivalent to the bias of an “anti-virtual qubit” [60]). Together with the current, these quantities obey the coupled equations
\[
\frac{d\hat{C}}{dt} = 2g\hat{Z} - 2p\hat{C}, \quad (A1)
\]
\[
\frac{d\hat{Z}}{dt} = p(\langle \hat{Z} \rangle_{eq} - \hat{Z}) - 2g\hat{C}, \quad (A2)
\]
\[
\frac{d\hat{N}}{dt} = p(1 - 2\hat{N} + \langle \hat{Z} \rangle_{eq}\hat{Z} - 2\langle \hat{S} \rangle_{eq} \hat{S}), \quad (A3)
\]
\[
\frac{d\hat{S}}{dt} = p(\langle \hat{S} \rangle_{eq} - \hat{S}), \quad (A4)
\]
where the equilibrium averages denoted by, for example, $\langle \hat{Z} \rangle_{eq} = \text{Tr}(\hat{Z}\tau_1 \otimes \tau_2)$, can be evaluated using the formula $\langle \hat{O} \rangle_{eq} = - \tanh(E_j/2T_j)$. The quasi-stationary state is found by setting the mean value of the above derivatives above to zero, yielding the solutions $\langle \hat{S} \rangle = \langle \hat{S} \rangle_{eq}$ and
\[
\langle \hat{C} \rangle = \frac{\langle \hat{Z} \rangle_{eq}}{p^2 + 2g^2}, \quad (A5)
\]
\[
\langle \hat{Z} \rangle = \frac{\langle \hat{Z} \rangle_{eq}}{p^2 + 2g^2}, \quad (A6)
\]
\[
\langle \hat{N} \rangle = \langle \hat{N} \rangle_{eq} - \frac{g^2}{p^2 + 2g^2} \langle \hat{Z} \rangle_{eq}^2. \quad (A7)
\]
The mean rate of energy transfer to the load is then given by $W = E_v \langle \hat{Z} \rangle_{eq}$, with $E_v = \frac{g^2}{p^2 + 2g^2}$, which is equivalent to Eq. (12).

To find the fluctuations, we consider the equation of motion for $\hat{W}^2$, which is $d\hat{W}^2/ dt = g E_v \hat{K}$, with $\hat{K} = [\hat{W}, \hat{C}]$. We also define the operator $\hat{\Omega} = 2\hat{W}$ in the Schrödinger picture, which in the Heisenberg picture is coupled to $\hat{K}$ via the equations
\[
\frac{d\hat{K}}{dt} = 2g \left(2\hat{\Omega} + E_v \hat{N}\right) - 2p\hat{K}, \quad (A8)
\]
\[
\frac{d\hat{\Omega}}{dt} = p \left(\langle \hat{Z} \rangle_{eq}\hat{W} - \hat{\Omega}\right) - g\hat{K}. \quad (A9)
\]
After eliminating $\hat{\Omega}$ from the equations and focussing on asymptotically long times where the solutions in Eqs. (A5)–(A7) hold, we obtain
\[
\left[d^2/ dt^2 + 3p d/ dt + 2p^2 + 4g^2\right] \langle \hat{K} \rangle = 2g p E_v \left(2E_v \langle \hat{Z} \rangle_{eq}^2 + \langle \hat{N} \rangle\right). \quad (A10)
\]
This describes a damped harmonic oscillator under a driving force given by the right-hand side of the equation. Considering times long enough for transient oscillations to decay to zero, i.e. $pt \gg 1$, the solution is easily found by using the ansatz $\langle \hat{K} \rangle = a + bt$ and solving for the constants $a$ and $b$. This procedure ultimately yields
\[
\frac{d}{dt} \langle \hat{W}^2 \rangle = \frac{d}{dt} \langle \hat{W} \rangle^2 + \Gamma E_v \left[2g \left(\langle \hat{Z} \rangle_{eq} - \frac{2g^2 (g^2 + 2p^2)}{(p^2 + 2g^2)^2} \langle \hat{Z} \rangle_{eq}^2\right)\right], \quad (A11)
\]
which is equivalent to Eq. (14).

Finally, we need the energy currents entering the system from each bath, $j = 1, 2$, defined by $Q_j = \langle \hat{D}_j^\dagger \hat{H} \rangle$. A simple
calculation reveals that
\[ \dot{Q}_j = p E_j \left( \langle \dot{\sigma}_j \rangle_{eq} - \langle \dot{\sigma}_j \rangle - p \langle \tilde{H}_{\text{int}} \rangle \right). \]  
(A12)

The Heisenberg equation for \( \tilde{H}_{\text{int}} \) is \( d \tilde{H}_{\text{int}}/dt = -2p \tilde{H}_{\text{int}} \), whose stationary solution is \( \langle \tilde{H}_{\text{int}} \rangle = 0 \). Eq. (A12) therefore recovers the local energy current defined by Brunner et al. [60]. (Note that the mean interaction energy vanishes only because of the assumption of resonant interactions, i.e. \( [H_0, \tilde{H}_{\text{int}}] = 0 \), which in the language of collisional models means that the qubit resets do not perform work on average [89–91].) It is then straightforward to verify from the Heisenberg equations for \( \langle \dot{\sigma}_j \rangle \) that the stationary heat currents are
\[ \dot{Q}_1 = -g E_1 \langle \dot{C} \rangle, \quad \dot{Q}_2 = g E_2 \langle \dot{C} \rangle. \]  
(A13)

2. Local Lindblad equation

We now carry out the same calculation for the local Lindblad dissipation model defined by Eq. (19), which differs from the reset model of Eq. (4) by the absence of the dephasing terms proportional to \( \gamma \). These terms affect only off-diagonal operators in the computational basis. Therefore, only the equations of motion for \( \dot{C} \) and \( \dot{K} \) are altered, being given by
\[ \frac{d \dot{C}}{dt} = 2g \dot{Z} - p \dot{C}, \]  
(A14)
\[ \frac{d \dot{K}}{dt} = 2g \left( \frac{2}{3} \dot{\Omega} + E \dot{N} \right) - p \dot{K}. \]  
(A15)

We see that the effective decoherence rate is cut in half. All other equations of motion and manipulations proceed as in the previous section. We find the solutions
\[ \langle \dot{C} \rangle = \frac{2g p}{p^2 + 4g^2} \langle \dot{Z} \rangle_{eq}, \]  
(A16)
\[ \langle \dot{Z} \rangle = \frac{p^2}{p^2 + 4g^2} \langle \dot{Z} \rangle_{eq}, \]  
(A17)
\[ \langle \dot{N} \rangle = \langle \dot{N} \rangle_{eq} - \frac{2g^2}{p^2 + 4g^2} \langle \dot{Z} \rangle_{eq}^2, \]  
(A18)

from which we obtain the power \( \dot{W} = g E_r \langle \dot{C} \rangle = \Gamma_r E_r \langle \dot{Z} \rangle_{eq} \), with \( \Gamma_r = 2g^2 p/(p^2 + 4g^2) \). The heat currents are given by Eq. (13) together with Eq. (16).

The above solutions are then used to find an effective equation of motion for \( \langle \dot{K} \rangle \), which reads as
\[ \left[ \frac{d^2}{dt^2} + 2p \frac{d}{dt} + p^2 + 4g^2 \right] \langle \dot{K} \rangle = 2g p E_r \left( 2\Gamma_r \langle \dot{Z} \rangle_{eq}^2 + \langle \dot{N} \rangle \right). \]  
(A19)

Postulating an asymptotic solution of the form \( \langle \dot{K} \rangle = a + bt \), we eventually find
\[ \dot{\Lambda}_W = \Gamma_r E_r \left[ \langle \dot{N} \rangle_{eq} - \frac{2g^2 (5p^2 + 4g^2)}{p^2 + 4g^2} \langle \dot{Z} \rangle_{eq}^2 \right], \]  
(A20)

which is equivalent to Eq. (22).

Appendix B: Bounds for the TUR ratio

In this Appendix, we derive bounds for the TUR ratios given in the main text. Let us begin with the two-qubit engine described by a reset thermalisation model in Eq. (4), whose TUR ratio is given in Eq. (16). We first note from Eqs. (9) and (10) that \( \langle \dot{Z} \rangle_{eq} \geq \frac{1}{2} \tan \chi/2 \), where equality is reached only in the limit of vanishing bias and work output, i.e. \( \beta_j E_j \rightarrow 0 \). However, this simple inequality allows us to bound the TUR ratio as
\[ \frac{\dot{\Lambda}_W}{W_2} \geq \chi \coth(\chi/2) \left[ 1 - \frac{1}{2} f(g/p) \tanh^2(\chi/2) \right]. \]  
(B1)

where we defined the function
\[ f(r) = \frac{2r^2 (2 + r^2)}{(1 + 2r^2)^2}. \]  
(B2)

The above function is positive and has a maximum equal to \( \max_r f(r) = 2/3 \), which is attained at \( r = 1 \). It follows that
\[ \frac{\dot{\Lambda}_W}{W_2} \geq \chi \coth(\chi/2) \left[ 1 - \frac{3}{4} \tanh^2(\chi/2) \right] \geq 2, \]  
(B3)

where the final equality is saturated only in the limit \( \chi \rightarrow 0 \).

For the local Lindblad equation (19), we can write an expression analogous to Ineq. (B1) but with a different function
\[ f(r) = \frac{2r^2 (5 + 4r^2)}{(1 + 4r^2)^2}, \]  
(B4)

whose maximum value \( \max_r f(r) = 25/32 \) is attained at \( r = \sqrt{5/12} \). This leads to the bound
\[ \frac{\dot{\Lambda}_W}{W_2} \geq \chi \coth(\chi/2) \left[ 1 - \frac{25}{64} \tanh^2(\chi/2) \right] \geq 1.982 \ldots, \]  
(B5)

which can be verified numerically. In practice, we find that any violations of the classical bound are very small and occur in a restricted region of the parameter space.

Finally, we move to the three-qubit engine model of Sec. III. Here, we have an exact equality given by Eq. (32), which is of the form of the right-hand side of Ineq. (B3) with
\[ f(r) = \frac{24r^2}{(1 + 8r^2)^2}. \]  
(B6)

This function is upper-bounded by \( \max_r f(r) = 3/4 \), which occurs when \( r = 1/2 \sqrt{2} \). As a result, we obtain the bound
\[ \frac{\dot{\Lambda}_W}{W_2} \geq \chi \coth(\chi/2) \left[ 1 - \frac{3}{4} \tanh^2(\chi/2) \right] \geq 1.245 \ldots. \]  
(B7)

Appendix C: Solution for the three-qubit engine

1. Master equation derivation

In this Appendix, we detail the solution of the three-qubit engine model in the limit of fast thermalisation (i.e. weak coherent coupling). We begin by sketching the derivation of the
effective master equation. Let \( \hat{\mathcal{R}}(t) \) denote the total density matrix of the three qubits and the load. We consider a reset thermalisation model for the qubits, described by a dissipator \( \mathcal{D} \mathcal{R} = \sum_{j=1}^3 \gamma_j \mathcal{R}_j \otimes \mathcal{T}_j (\mathcal{R} - \mathcal{R}) \) and thermalisation rate \( \gamma' \). We also define the Hamiltonian superoperators \( \mathcal{H}_\text{int}, \mathcal{H}_\text{load} \) and \( \mathcal{V} \) in terms of the commutators with the corresponding Hamiltonians in Eqs. (24)–(26), e.g. \( \mathcal{H}_\text{int} \mathcal{R} = -i[\mathcal{H}, \mathcal{R}] \). The master equation can then be written in the form \( \frac{d\mathcal{R}}{dt} = (\mathcal{L}_0 + \mathcal{V} + \mathcal{H}_\text{int}) \mathcal{R} \), where \( \mathcal{L}_0 = \mathcal{H}_0^\dagger + \mathcal{D} \).

The assumption of fast thermalisation means that \( \mathcal{L}_0 \gg \mathcal{V}, \mathcal{H}_\text{int} \), so that the damped qubits mostly remain in thermal equilibrium and are only weakly perturbed by the interaction terms. This idea can be formalised using standard projection-operator techniques [88, 92]. We define a projector by

\[
\mathcal{P} \hat{\mathcal{R}} = \hat{\mathcal{P}}_1 \otimes \hat{\mathcal{P}}_2 \otimes \mathcal{T}_{12} (\mathcal{R}),
\]

which satisfies the easily verified properties and relations

\[
[\mathcal{P}, \mathcal{L}_0] = 0, \quad [\mathcal{P}, \mathcal{H}_\text{int}] = 0, \quad \mathcal{P} \mathcal{V} \mathcal{P} = 0.
\]

We also define the orthogonal projector \( \mathcal{Q} = 1 - \mathcal{P} \), such that \( \mathcal{P}^2 = \mathcal{P}, \mathcal{Q}^2 = \mathcal{Q} \), and \( \mathcal{Q} \mathcal{P} = \mathcal{P} \mathcal{Q} = 0 \). Let us move to an interaction picture via the transformation \( \hat{\mathcal{R}}(t) = e^{-\mathcal{L}_0 t} \hat{\mathcal{R}}(t) \), where the dynamics is described by the master equation

\[
\frac{d\hat{\mathcal{R}}}{dt} = \hat{\mathcal{V}}(t) \hat{\mathcal{R}}(t) + \hat{\mathcal{H}}\text{int}(t) \hat{\mathcal{R}}(t).
\]

Above, tildes denote superoperators in the interaction picture, e.g. \( \hat{\mathcal{V}}(t) = e^{-\mathcal{L}_0 t} \mathcal{V} e^{\mathcal{L}_0 t} \). Applying projectors to both sides of the master equation, inserting appropriate factors of \( 1 = \mathcal{P} + \mathcal{Q} \), and using the properties (C2), we find that

\[
\frac{d}{dt} \mathcal{P} \hat{\mathcal{R}} = \mathcal{P} \hat{\mathcal{V}}(t) \mathcal{P} \hat{\mathcal{R}}(t) + \hat{\mathcal{H}}\text{int}(t) \mathcal{P} \hat{\mathcal{R}}(t),
\]

\[
\frac{d}{dt} \mathcal{Q} \hat{\mathcal{R}} = \mathcal{Q} \hat{\mathcal{V}}(t) \mathcal{Q} \hat{\mathcal{R}}(t) + \mathcal{Q} \hat{\mathcal{V}}(t) \mathcal{P} \hat{\mathcal{R}}(t) + \hat{\mathcal{H}}\text{int}(t) \mathcal{Q} \hat{\mathcal{R}}(t).
\]

The solution of Eq. (C5) with initial condition \( \mathcal{Q} \hat{\mathcal{R}}(0) = 0 \) is

\[
\mathcal{Q} \hat{\mathcal{R}}(t) = \int_0^t dt' \mathcal{G}(t, t') \mathcal{V}(t') \mathcal{P} \hat{\mathcal{R}}(t'),
\]

\[
\mathcal{G}(t, t') = T \exp \left[ \int_{t'}^t dt'' \mathcal{Q} \left( \mathcal{V}(t'') + \mathcal{H}_\text{int}(t'') \right) \right],
\]

where we have used Eq. (C2) to write \( \mathcal{Q} \mathcal{V} \mathcal{P} = (1 - \mathcal{P}) \mathcal{V} \mathcal{P} = \mathcal{V} \mathcal{P} \). Plugging the above solution back into Eq. (C4) yields

\[
\frac{d}{dt} \mathcal{P} \hat{\mathcal{R}} = \mathcal{H}_\text{int}(t) \mathcal{P} \hat{\mathcal{R}}(t) + \int_0^t dt' \mathcal{P} \hat{\mathcal{V}}(t') \mathcal{G}(t, t') \mathcal{P} \hat{\mathcal{R}}(t'),
\]

\[
\approx \mathcal{H}_\text{int}(t) \mathcal{P} \hat{\mathcal{R}}(t) + \int_0^t dt' \mathcal{P} \hat{\mathcal{V}}(t') \mathcal{V}(t - t') \mathcal{P} \hat{\mathcal{R}}(t - t'),
\]

\[
\approx \mathcal{H}_\text{int}(t) \mathcal{P} \hat{\mathcal{R}}(t) + \int_0^\infty dt' \mathcal{P} \hat{\mathcal{V}}(t - t') \mathcal{P} \hat{\mathcal{R}}(t).
\]

On the second line, we have expanded \( \mathcal{G}(t, t') \) to lowest order in the small quantities \( \mathcal{V} \) and \( \mathcal{H}_\text{int} \), and shifted the integration variable as \( t' \to t - t' \). On the third line, we have invoked the Markov approximation by assuming that the integrand decays on a timescale much shorter than the characteristic evolution timescale of \( \hat{\mathcal{R}}(t) \). Since the former timescale is given by \( (\gamma')^{-1} \) and the latter is determined by the inverse of \( k \) and \( g \), this approximation is consistent with our starting assumption that \( \gamma' \gg g, k \).

Finally, we trace over qubits 1 and 2 and transform back to the Schrödinger picture to obtain the master equation (28), with \( \hat{\mathcal{P}}(t) = \mathcal{T}_{12} [\hat{\mathcal{R}}(t)] \) and the rates given by

\[
\gamma^+ = \frac{k^2 e^{g/2}}{p^2 Z', Z_2^2}.
\]

The self-consistency of the above derivation requires that \( \gamma^+ + \gamma^- = p \ll \gamma' \), hence the energy transfer dynamics described by Eq. (28) is necessarily much slower than the underlying thermalisation processes. We note that other choices for the dissipator \( \mathcal{D} \) that obey \( \mathcal{D} \hat{\mathcal{P}}_1 \otimes \hat{\mathcal{P}}_2 = 0 \) would lead to a master equation of the same form, but with different expressions for the rates.

### 2. Asymptotic solution

The solution proceeds straightforwardly according to the methods of Appendix A. The mean energy of the load follows from the equation of motion \( \frac{d\hat{W}}{dt} = g \mathcal{E}_V(\hat{C}) \), with the current operator now given by

\[
\hat{\mathcal{C}} = ig (\hat{\sigma}_3 \hat{A} - \hat{\sigma}_3 \hat{A}^\dagger).
\]

The current and qubit bias obey the coupled differential equations

\[
\frac{d\hat{\mathcal{C}}}{dt} = 2g \hat{\sigma}_3^z - \frac{p}{2} \hat{\mathcal{C}},
\]

\[
\frac{d\hat{\sigma}_3^z}{dt} = p \left( (\hat{\sigma}_3^z)_{\text{eq}} - \hat{\sigma}_3^z \right) - 2g \hat{\mathcal{C}},
\]

whose steady-state solution is

\[
\langle \hat{\mathcal{C}} \rangle = \frac{4g p}{p^2 + 8g^2} \langle \hat{\sigma}_3^z \rangle_{\text{eq}},
\]

\[
\langle \hat{\sigma}_3^z \rangle_{\text{eq}} = \frac{p^2}{p^2 + 8g^2} \langle \hat{\sigma}_3^z \rangle_{\text{eq}},
\]

from which Eq. (29) follows.

The power fluctuations follow from \( \frac{d\hat{W}^2}{dt} = g \mathcal{E}_V(\hat{\mathcal{C}}) \), with \( \widehat{\mathcal{C}} = (\hat{W}, \hat{\mathcal{C}}) \). Defining the Schrödinger-picture operator \( \hat{\Omega} = \hat{\sigma}_3 \hat{W} \), the relevant Heisenberg equations read as

\[
\frac{d\hat{\Omega}}{dt} = 2g (2\hat{\Omega} - \mathcal{E}_V) - \frac{p}{2} \hat{\Omega},
\]

\[
\frac{d\hat{\mathcal{E}}_V}{dt} = p \left( (\hat{\sigma}_3^z)_{\text{eq}} \hat{W} - \hat{\Omega} \right) - g \hat{\mathcal{E}}_V.
\]

Eliminating \( \hat{\Omega} \) and taking the long-time limit, we obtain

\[
\left[ \frac{d^2}{dt^2} + \frac{3p}{2} \frac{d}{dt} + \frac{p^2}{2} + 4g^2 \right] \langle \hat{\mathcal{E}}_V \rangle = 2gp \mathcal{E}_V (2\Gamma t Z^2_{\text{eq}} + 1).
\]
Seeking an asymptotic solution of the form $\langle \hat{K} \rangle = a + bt$, we deduce Eq. (31).

Finally, the energy currents can be computed using the master equation describing the three qubits together with the load, as given in Appendix C1. Considering the equation of motion for $\langle \hat{\sigma}_j^z \rangle$ separately for $j = 1, 2, 3$, and setting all derivatives to zero, one can show that a relation of the same form as Eq. (A13) holds, but with $\hat{C}$ given in Eq. (C10). Combining this with the solution (C13), Eq. (30) follows.
This is easily shown by interpreting the generating function (38) as describing a single step in Fock space, which would lead to an expression precisely three times smaller than Eq. (42).