

---

## Strong Coupling Corrections in Quantum Thermodynamics

---

Table 1: Notation used throughout the report.

Symbol	Meaning	Mathematical definition
<b>Latin letters</b>		
$H_S$	System Hamiltonian (local Hamiltonian of $S$ )	
$H_B$	Bath Hamiltonian	
$H^{(i)}$	Total Hamiltonian at discrete step $i$	$H^{(i)} = H_S^{(i)} + V + H_B$
$H_S^{(i)}$	System Hamiltonian at step $i$	
$\tilde{H}_S$	Effective system Hamiltonian fixed by the initial system state via $\rho_S = \omega_\beta(\tilde{H}_S)$	$H_S^{(1)} = \tilde{H}_S$
$\tilde{H}^{(0)}$	Initial effective total Hamiltonian	$\tilde{H}^{(0)} = \tilde{H}_S^{(0)} + H_B$
$H^{(A)}$	Total Hamiltonian at stage A (four-stroke heat-engine protocol)	
$H^{(B)}$	Total Hamiltonian at stage B (four-stroke heat-engine protocol)	
$H^{(C)}$	Total Hamiltonian at stage C (four-stroke heat-engine protocol)	
$H^{(D)}$	Total Hamiltonian at stage D (four-stroke heat-engine protocol)	
$V$	System–bath interaction Hamiltonian	

---

*Continued on next page*

---

---

Symbol	Meaning	Mathematical definition
$Z$	Minimizing Hamiltonian of the residual free-energy functional	
$X$	Minimizing Hamiltonian of the irreversible free-energy functional	
$E_S$	Energy of the system	$E_S(\rho) = \text{Tr}(\rho H_S)$
$E_B$	Energy of the bath	$E_B(\rho) = \text{Tr}(\rho H_B)$
$E_{B_h}$	Energy of the hot bath	
$E_{B_c}$	Energy of the cold bath	
$W$	Total work throughout the protocol	
$W^{(i)}$	Work during step $i$	
$W_{B \rightarrow C}$	Work performed from stage $B$ to $C$ of the heat engine protocol	
$W_{D \rightarrow A}$	Work performed from stage $D$ to $A$ of the heat engine protocol	
$N$	Number of quenches/steps in the protocol	
<b>Constants</b>		
$\beta$	Inverse temperature	$\beta = 1/(k_B T)$
<b>States and distributions</b>		
$\rho$	Density operator (state)	$\rho \geq 0, \text{Tr} \rho = 1$
$\rho_S$	Initial state of the system	$\rho_S = \text{Tr}_B(\rho^{(0)})$
$\rho_B$	Reduced state of the bath	$\rho_B = \text{Tr}_S(\rho)$
$\rho^{(0)}$	Initial joint state before the protocol starts	$\rho^{(0)} = \rho_S \otimes \omega_\beta(H_B)$
<i>Continued on next page</i>		

---

Symbol	Meaning	Mathematical definition
$\omega_\beta(H)$	Gibbs state at inverse temperature $\beta$ with respect to Hamiltonian $H$	$\omega_\beta(H) = \frac{e^{-\beta H}}{\text{Tr}(e^{-\beta H})}$
$\omega_{\beta_h}(H)$	Gibbs state at inverse temperature $\beta_h$	$\omega_{\beta_h}(H) = \frac{e^{-\beta_h H}}{\text{Tr}(e^{-\beta_h H})}$
$\omega_{\beta_c}(H)$	Gibbs state at inverse temperature $\beta_c$	$\omega_{\beta_c}(H) = \frac{e^{-\beta_c H}}{\text{Tr}(e^{-\beta_c H})}$
<b>Thermodynamic functionals</b>		
$F(\rho, H)$	(Non-equilibrium) free energy	$F(\rho, H) = \text{Tr}(\rho H) - T S(\rho)$
$S(\rho)$	von Neumann entropy	$S(\rho) = -\text{Tr}(\rho \log \rho)$
$D(\rho \parallel \sigma) / S(\rho \parallel \sigma)$	Quantum relative entropy	$D(\rho \parallel \sigma) = S(\rho \parallel \sigma) = \text{Tr}[\rho(\log \rho - \log \sigma)]$
$\eta$	Efficiency of heat engine	$\eta = \frac{W_{\text{out}}}{Q_h}$
$\eta_C$	Carnot efficiency	$\eta_C = 1 - \frac{T_c}{T_h}$
$\mathcal{Z}(H, \beta)$	Partition function	$\mathcal{Z}(H, \beta) = \text{Tr}(e^{-\beta H})$
$\Delta F^{(\text{irr})}$	Irreversible free-energy change	$\Delta F^{(\text{irr})} = F(\rho^{(0)}, H^{(1)}) - F(\omega_\beta(H^{(1)}), H^{(1)})$
$\Delta F^{(\text{res})}$	Residual free energy change	$\Delta F^{(\text{res})} := F(\omega_\beta(H^{(N)}), H^{(0)}) - F(\omega_\beta(H^{(0)}), H^{(0)})$
<b>Contextual subscripts/superscripts</b>		
<i>Continued on next page</i>		

Symbol	Meaning	Mathematical definition
$(\cdot)_{H,\beta}$	Evaluation with respect to $(H, \beta)$	$(\cdot)_{H,\beta} = \int_0^1 d\tau e^{\beta\tau H} (\cdot) e^{-\beta\tau H}$

Thermodynamics is the fundamental theory concerned with heat and temperature and their relation to energy and work. In phenomenological thermodynamics, an implicit assumption is that coupling between the working systems and their heat baths are so weak that effects of the interaction can be neglected. As a consequence, the equilibrium states of the working systems are thermal states, in fact thermal states of local Hamiltonians.

Thermodynamics sets uncompromising limits on what any heat engine can do: no process beats Carnot efficiency and the second law forbids free work. In the nanoscale and quantum regimes, where surface-to-volume ratios are large and control is local, that assumption breaks down: when the system-bath interaction is appreciable, the system equilibrates not to a Gibbs state of its local Hamiltonian but to the reduced state of the global Gibbs ensemble of system + bath. Taking this fact seriously forces a re-examination of the second law itself in the strong-coupling regime. From both a conceptual and a practical perspective, this matters. Conceptually, the second law has multiple readings, as a bound on extractable work, as a Clausius inequality for heat, and as the Carnot bound on efficiency. Practically, the existence of a strict efficiency limit is inconvenient when one wants to wring out both maximum efficiency and high power from microscopic engines. A natural strategy, then, is to loosen the standard restrictions and ask: Can stronger coupling help? If yes, how far can we push it without violating the second law? The framework of Perarnau-Llobet et al. answers these questions in a model-independent way, deriving universal strong-coupling corrections to all three readings of the second law.

To make these statements precise, the authors set up a minimal, operational framework that we will use in the next sections.

## 1 Framework

We consider a systems  $S$  and a heat bath  $B$ , with internal Hamiltonians  $H_S$  and  $H_B$ , respectively. They can interact via a possibly strong interaction  $V$ . Thermodynamic protocols then consist of transformations over  $H_S$ , and equilibration processes induced by  $V$ . Specifically, we consider protocols of  $N$  steps, and denote by  $\rho^{(i)}$  and  $H^{(i)}$  the state and the Hamiltonian of  $SB$  in the  $i$ th step, consisting of three elementary operations:

(A) *Turning on and off interaction.*

With this, the authors model the process of bringing  $S$  and  $B$  into contact by a sudden quench of the Hamiltonian,  $H^{(i)} = H_S^{(i)} + H_B + V$ , which leaves the state unchanged at the instant of the switching. Similarly,  $V$  can be turned off at any step of the process. To get the expression

for energy and work, we take the derivative of  $E$ :

$$E = \langle H \rangle = \text{Tr}(\rho H) \rightarrow dE = \text{Tr}(d\rho H) + \text{Tr}(\rho dH). \quad (1)$$

We know that the changes in the Hamiltonian corresponds to work while the changes in the state corresponds to the change in entropy and therefore change in heat. Since the process of switching on/off the Hamiltonian is performed instantaneously,  $d\rho = 0$ , and we have

$$dE = \text{Tr}(\rho dH) \rightarrow \Delta E = \text{Tr}(\rho(H^{(i+1)} - H^{(i)})). \quad (2)$$

In this part of the protocol, the system's Hamiltonian changes from  $H^{(i)} = H_S^{(i)} + H_B$  to  $H^{(i+1)} = H_S^{(i)} + H_B + V$ . Therefore,

$$W_{\text{on}}^{(i)} = \text{Tr}(\rho^{(i)} V) = -W_{\text{off}}^{(i)}. \quad (3)$$

**(B) A quench on  $S$ .**

A fast transformation of  $H_S$  is implemented, so that  $H^{(i)} = H_S^{(i)} + H_B + V$  is changed to  $H^{(i+1)} = H_S^{(i+1)} + H_B + V$ , whereas the state  $\rho^{(i)}$  remains unchanged. The corresponding work extracted reads

$$W^{(i)} = \text{Tr}(\rho_S^{(i)}(H_S^{(i)} - H_S^{(i+1)})) \quad (4)$$

where  $\rho_S^{(i)} := \text{Tr}(\rho^{(i)})$ .

**(C) A thermalization process.**

This operation models the closed free evolution of  $S+B$  when  $V$  is present, i.e., under  $H^{(i+1)}$ . During this step the Hamiltonian is fixed, so any energy exchanged between  $S$  and  $B$  is heat; the total energy of  $S+B$  is conserved, while  $E_S$  and  $E_B$  may change with  $Q = -\Delta E_B$ . This operation has no work cost, as the total energy of  $S+B$  is preserved. When they reach equilibrium, we assume that  $S$  is well described by

$$\rho_S^{(i+1)} = \text{Tr}_B(\omega_\beta(H^{(i+1)})), \quad (5)$$

where  $\omega_\beta(H) = e^{-\beta H} / \text{Tr}(e^{-\beta H})$ .

Similarly, we assume that the boundary between  $S$  and  $B$ , i.e., the support of  $V$ , can also be described by the reduced form of  $\omega_\beta(H^{(i+1)})$ . Both assumptions are reasonable for locally interacting systems. We introduce the notation  $\omega^{(i)} := \omega_\beta(H^{(i)})$ ,  $\omega_S^{(i)} := \omega_\beta(H_S^{(i)})$ . We also set the convention  $\hbar = k_B = 1$ .

A thermodynamic protocol consists of an arbitrary sequence of operations of the types (A)-(C). The total expected work  $W$  gained in the process is the sum of all the contributions of the form (3) and (4). In this framework, the Hamiltonian terms  $V$  and  $H_B$  remain fixed throughout the protocol, reflecting the fact that an experimenter will in many realistic situations not have precise control

over  $B$  and the coupling between  $S$  and  $B$ , at least not beyond the capability of turning it on and off. During transformations of the form (C),  $S$  is assumed to be brought to equilibrium after a sufficiently long time. That is, possible finite-time effects are not included.

With the framework in place, we now ask the central question: how much work can be extracted from a given nonequilibrium state when  $S$  is allowed contact with a single bath under strong coupling?

## 2 Maximal work extraction for arbitrary coupling strengths.

We now study work extraction from an out of equilibrium state of  $S$ . In order to avoid the possibility of extracting work from the energy stored in  $V$ , we consider that  $S$  is initially isolated from  $B$ . The initial Hamiltonian is hence noninteracting,  $H^{(0)} = H_S + H_B$ , and the initial state is uncorrelated,  $\rho^{(0)} = \rho_S \otimes \omega_B(H_B)$ . In this section, the task is to optimize work extraction protocols over all cyclic protocols that can be constructed with the operations (A)(C) described in the last section. Here cyclicity is understood in terms of the Hamiltonian, so that at the end of the protocol consisting of  $N + 1$  steps can in principle consist of any of the operations (A), (B), and (C) of the last section, and  $N$  can be arbitrarily large.

We now focus on protocols where the interaction is turned on and off only once in the protocol (i.e., operation 1 is only implemented twice). Later we will show that this is in fact optimal.

### 2.1 Optimal protocol for maximum work extraction

Any single-contact protocol can be described as:

1. A series of quenches are applied to the local Hamiltonian of  $S$ ,  $H_S \rightarrow \dots \rightarrow H_S^{(1)}$ , and the interaction between  $S$  and  $B$  is turned on. The total Hamiltonian becomes  $H^{(1)} = H_S^{(1)} + V + H_B$ . The expected work gain of this process reads

$$W_1 = \text{Tr} \left( (H_S - H_S^{(1)} - V) \rho^{(0)} \right). \quad (6)$$

2. The Hamiltonian of  $S$  is modified while  $S$  is in contact with  $B$ , amounting to a sequence of quenches  $H_S^{(1)} \rightarrow H_S^{(2)} \rightarrow \dots \rightarrow H_S^{(N)}$ , each followed by an equilibration to the corresponding thermal state, until  $H^{(N)} = H_S^{(N)} + V + H_B$  is reached. Note that we assume that  $H_S^{(N)}$  is independent of  $N$ , so that increasing  $N$  means doing the same Hamiltonian transformation in more steps. During this process the state of  $S$  reads  $\rho_S^{(j)} = \text{Tr}_B(\omega^{(j)})$ ,  $j = 1, \dots, N$ . The total work gain is then

$$W_2 = \sum_{i=1}^{N-1} \text{Tr} \left( \rho_S^{(i)} (H_S^{(i)} - H_S^{(i+1)}) \right). \quad (7)$$

Since each quench is preceded by a thermalization step, the pre-quench state is a Gibbs state for the current Hamiltonian. Thus in the quench-work expression we can set  $\rho_S^{(i)} = \omega_S^{(i)} :=$

$\text{Tr}_B(\omega^{(i)})$  with  $\omega^{(i)} = \omega_\beta(H^{(i)})$ . During the subsequent waiting period, the Hamiltonian is held fixed at  $H^{(i+1)}$ ; only the state evolves, and the system relaxes to  $\omega^{(i+1)}$ . This step carries no work cost (all energy exchanged is heat). Since  $H^{(i+1)} - H^{(i)} = H_S^{(i+1)} - H_S^{(i)}$ ,

$$\implies \text{Tr}(\omega^{(i)}(H^{(i)} - H^{(i+1)})) = \text{Tr}(\omega_S^{(i)}(H_S^{(i)} - H_S^{(i+1)})). \quad (8)$$

Therefore, Eq. (7) becomes

$$W_2 = \sum_{i=1}^{N-1} \text{Tr}(\omega^{(i)}(H^{(i)} - H^{(i+1)})). \quad (9)$$

Now we use the quantum relative entropy, defined by  $S(\rho||\sigma) = \text{Tr}(\rho(\ln(\rho) - \ln(\sigma)))$ , to simplify the expression above. Here the inverse temperature  $\beta$  is fixed by the bath.

$$\begin{aligned} \rightarrow S(\omega^{(i)}||\omega^{(i+1)}) &= \text{Tr}(\omega^{(i)}(\ln(\omega^{(i)}) - \ln(\omega^{(i+1)}))) \\ \ln(\omega^{(i)}) &= -\ln\left(\frac{e^{-\beta H^{(i)}}}{\text{Tr}(e^{-\beta H^{(i)}})}\right) = -\beta H^{(i)} - \ln(\mathcal{Z}^{(i)}) \end{aligned} \quad (10)$$

$$\begin{aligned} \implies S(\omega^{(i)}||\omega^{(i+1)}) &= \text{Tr}(\omega^{(i)}(-\beta H^{(i)} - \ln(\mathcal{Z}^{(i)}) + \beta H^{(i+1)} + \ln(\mathcal{Z}^{(i+1)}))) \\ &= -\beta \text{Tr}(\omega^{(i)}(H^{(i)} - H^{(i+1)})) - \ln(\mathcal{Z}^{(i)}) + \ln(\mathcal{Z}^{(i+1)}) \\ \implies \text{Tr}(\omega^{(i)}(H^{(i)} - H^{(i+1)})) &= T \ln(\mathcal{Z}^{(i+1)}) - T \ln(\mathcal{Z}^{(i)}) - TS(\omega^{(i)}||\omega^{(i+1)}) \end{aligned} \quad (11)$$

Expressing work in thermodynamic terms enables a direct comparison with classical results. Hence, we use the definition of free energy and von Neumann entropy to further simplify the expression above.

$$E = -\frac{\partial \ln(\mathcal{Z})}{\partial \beta}, \quad S = -\text{Tr}(\rho \ln(\rho)) \quad (12)$$

$$S(\omega) = -\text{Tr}(\omega \ln(\omega)) = \beta \text{Tr}(\omega H) + \underbrace{\text{Tr}(\omega)}_{=1} \ln(\mathcal{Z}) = \beta E(\omega) + \ln(\mathcal{Z}) \quad (13)$$

$$F(\omega, H) = E(\omega) - TS(\omega) = E(\omega) - \underbrace{T\beta}_{=1} E(\omega) - T \ln(\mathcal{Z}) = -T \ln(\mathcal{Z}) \quad (14)$$

$$\stackrel{(11)}{\implies} \text{Tr}(\omega^{(i)}(H^{(i)} - H^{(i+1)})) = F(\omega^{(i)}, H^{(i)}) - F(\omega^{(i+1)}, H^{(i+1)}) - TS(\omega^{(i)}||\omega^{(i+1)}) \quad (15)$$

Using this in Eq. (9), we get

$$\begin{aligned}
 W_2 &= \sum_{i=1}^{N-1} F(\omega^{(i)}, H^{(i)}) - F(\omega^{(i+1)}, H^{(i+1)}) - TS(\omega^{(i)} || \omega^{(i+1)}) \\
 &= F^{(1)} - F^{(2)} + F^{(2)} - F^{(3)} + \dots + F^{(N-1)} - F^{(N)} - T \sum_{i=1}^{N-1} S(\omega^{(i)} || \omega^{(i+1)}) \\
 &= F(\omega^{(1)}, H^{(1)}) - F(\omega^{(N)}, H^{(N)}) - T \sum_{i=1}^{N-1} S(\omega^{(i)} || \omega^{(i+1)}). \tag{16}
 \end{aligned}$$

3. The interaction between  $S$  and  $B$  is turned off, and the Hamiltonian of  $S$  is brought back to the initial form by an arbitrary series of quenches,  $H_S^{(N)} \rightarrow \dots \rightarrow H_S$ . The extracted work is simply

$$W_3 = \text{Tr} \left( (V + H_S^{(N)} - H_S) \omega^{(N)} \right). \tag{17}$$

For quick reference, For orientation, the protocol is sketched in Fig. 2.1 below: starting from an isolated system with Hamiltonian  $H_S$ , the interaction with the bath (blue) is switched on (left red bar), a sequence of small quenches  $H_S^{(i)} \rightarrow H_S^{(i+1)}$  proceeds while the bath equilibrates the system (arrows indicate quenches), and the interaction is finally switched off (right red bar) at  $H_S^{(N)}$ .

These three steps conclude a cyclic Hamiltonian process. Two remarks are now in order. Firstly, the last term of (16) is positive and tends to zero in the limit  $N \rightarrow \infty$  (i.e., for isothermal processes); rigorous proof of the latter claim is provided in the following subsection. That implies

$$W_2 \leq W_2^{(\text{isoth})} = F(\omega^{(1)}, H^{(1)}) - F(\omega^{(N)}, H^{(N)}) \tag{18}$$

where  $W_2^{(\text{isoth})}$  stands for the work gain of an isothermal transformation. Within this framework, an isothermal process is implemented as a sequence of infinitesimal quenches, each followed by an equilibration step. Hence, in the optimal protocol contacts between  $S$  and  $B$  must be in the form of isothermal transformations.

### 2.1.1 Proof that the relative entropy term scales as $O(1/N)$ .

To proceed with the proof, we first reduce the problem using Pinskers inequality, which allows us to bound the relative entropy in terms of the trace norm between consecutive thermal states. Next, we use a Lipschitz inequality for the matrix exponential to bound the difference between  $e^A$  and  $e^B$ , where  $A = -\beta H^{(i)}$  and  $B = -\beta H^{(i+1)}$ , in terms of  $\|A - B\|$ .

We then apply Duhamels formula to estimate the derivative of the partition function, and combine it with the Hölder-Schatten inequality to bound the second term in the difference of thermal states.



These steps together give a bound on the trace norm

$$\|\omega_\beta^{(i)} - \omega_\beta^{(i+1)}\|_1 \leq (\text{constant}) \cdot \|H^{(i+1)} - H^{(i)}\|_\infty.$$

Finally, we show that  $\|H^{(i+1)} - H^{(i)}\|_1 = O(1/N)$ , and hence the relative entropy is  $O(1/N^2)$ , which goes to zero as  $N \rightarrow \infty$ . This proves that the total entropy production,

$$T \sum_{i=1}^{N-1} S(\omega_\beta^{(i)} \parallel \omega_\beta^{(i+1)}),$$

scales as  $O(1/N)$  and vanishes in the reversible limit.

The following definitions will be required later in the proof.

### Trace Norm

For a matrix  $A \in \mathbb{C}^{d \times d}$ , the trace norm is defined as

$$\|A\|_1 := \text{tr} \left( \sqrt{A^\dagger A} \right) = \sum_{i=1}^d s_i(A),$$

where  $s_i(A)$  are the singular values of  $A$ .

-Bhatia, *Matrix Analysis*, Springer, 1997, Section IV.2

### Infinity Norm

For a matrix  $A \in \mathbb{C}^{d \times d}$ , the operator norm is defined as

$$\|A\|_\infty := \sup_{\|x\|_2=1} \|Ax\|_2 = s_1(A),$$

where  $s_1(A)$  is the largest singular value of  $A$ . For Hermitian  $A$ , this equals the largest absolute value among its eigenvalues.

-Bhatia, *Matrix Analysis*, Springer, 1997, Section IV.2

Our first step is to prove Lipschitz inequality, which we will later use in the main argument.

For two density matrices  $A$  and  $B$  with  $A, B \geq 0$ ,  $\text{tr}(A) = \text{tr}(B) = 1$ , and spectra in  $[0, 1]$ , every unitarily-invariant norm  $\|\cdot\|$  satisfies

$$\|e^A - e^B\| \leq e \|A - B\|.$$

-R. Bhatia, *Matrix Analysis*, Springer, 1997, Chapter IX

Define  $F(s) = e^{(1-s)B} e^{sA}$  for  $s \in [0, 1]$ , where

$$A := -\beta H^{(i)}, \quad B := -\beta H^{(i+1)}.$$

clearly  $F(0) = e^B$  and  $F(1) = e^A$ . By differentiating we get,

$$F'(s) = e^{(1-s)B} (A - B) e^{sA}.$$

Therefore,

$$e^A - e^B = F(1) - F(0) = \int_0^1 e^{(1-s)B} (A - B) e^{sA} ds \quad (19)$$

For bounded operators (or finite matrices)  $A, B, C$  let  $\|\cdot\|_p$  denote the Schatten- $p$  norm (so  $p=1$  is the trace/nuclear norm and  $p=\infty$  is the operator norm). The multi-factor Hölder inequality for Schatten norms states that for any  $p_1, p_2, p_3 \in [1, \infty]$  and  $r \in [1, \infty]$  satisfying

$$\frac{1}{r} = \frac{1}{p_1} + \frac{1}{p_2} + \frac{1}{p_3},$$

one has

$$\|ABC\|_r \leq \|A\|_{p_1} \|B\|_{p_2} \|C\|_{p_3} \quad (20)$$

This is the direct extension of the two-factor Hölder-Schatten inequality.

-K. M. R. Audenaert, *Comparisons between quantum state distinguishability measures*, arXiv:1207.1197 (v2)

Choose  $p_1 = p_3 = \infty$  and  $p_2 = 1$ . Because  $1/\infty = 0$ , condition (20) gives  $1/r = 0 + 1 + 0 = 1$ . Plugging these values into (20) yields the inequality

$$\|ABC\|_1 \leq \|A\|_\infty \|B\|_1 \|C\|_\infty.$$

Hence Eq. (19) becomes

$$\|e^A - e^B\|_1 \leq \int_0^1 \|e^{(1-s)B}\|_\infty \|A - B\|_1 \|e^{sA}\|_\infty ds. \quad (21)$$

Because all eigenvalues of  $A$  and  $B$  lie in  $[0, 1]$ ,

$$\|e^{(1-s)B}\|_\infty \leq e^{(1-s)} \leq e, \quad \|e^{sA}\|_\infty \leq e^s \leq e.$$

For any Hermitian matrix  $X$  with spectral decomposition  $X = \sum_i \lambda_i |i\rangle\langle i|$  we have;

$$e^X = \sum_i e^{\lambda_i} |i\rangle\langle i| \implies \|e^X\|_\infty = e^{\lambda_{\max}(X)}. \quad (22)$$

Observing that  $(1-s)B$  has eigenvalues  $(1-s)\lambda_k(B)$ , we obtain (for  $s \in [0, 1]$ )

$$\lambda_{\max}((1-s)B) = (1-s)\lambda_{\max}(B),$$

and since every eigenvalue of  $A$  and  $B$  lies in  $[0, 1]$ ,

$$\lambda_{\max}((1-s)B) = (1-s)\lambda_{\max}(B) \leq 1-s \xrightarrow{\text{using (c)}} \|e^{(1-s)B}\| = e^{\lambda_{\max}((1-s)B)} \leq e^{1-s}$$

and similarly

$$\|e^s A\|_{\infty} = e^{\lambda_{\max}(sA)} \leq e^s.$$

Putting these bounds together gives

$$\begin{aligned} \implies \|e^A - e^B\|_1 &\leq \int_0^1 e \|A - B\|_1 e^{1-s} ds = e \|A - B\|_1, \\ \implies \|e^A - e^B\|_1 &\leq e \|A - B\|_1 = e \beta \|H^{(i+1)} - H^{(i)}\|_1. \end{aligned} \quad (23)$$

### Fan Dominance Theorem

Let  $A, B$  be two  $n \times n$  matrices. If

$$\|A\|_{(k)} \leq \|B\|_{(k)} \quad \text{for } k = 1, 2, \dots, n,$$

then

$$|||A||| \leq |||B||| \quad \text{for all unitarily invariant norms.}$$

-R. Bhatia, *Matrix Analysis*, Springer, 1997, Section IV.2.

Using this theorem, (23) becomes,

$$\rightarrow \boxed{\|e^A - e^B\|_{\infty} \leq e \beta \|H^{(i+1)} - H^{(i)}\|_{\infty}}$$

Next, recall the definitions of the corresponding Gibbs states,

$$\begin{aligned} \omega_{\beta}^{(i)} &= \frac{e^A}{\mathcal{Z}_A}, \quad \omega_{\beta}^{(i+1)} = \frac{e^B}{\mathcal{Z}_B}, \quad \mathcal{Z}_{A/B} := \text{tr } e^{A/B}, \\ \omega_{\beta}^{(i)} - \omega_{\beta}^{(i+1)} &= \frac{e^A}{\mathcal{Z}_A} - \frac{e^B}{\mathcal{Z}_B} = \frac{e^A - e^B}{\mathcal{Z}_A} + e^B \left( \frac{1}{\mathcal{Z}_A} - \frac{1}{\mathcal{Z}_B} \right), \\ \implies \|\omega_{\beta}^{(i)} - \omega_{\beta}^{(i+1)}\|_{\infty} &\leq \underbrace{\frac{\|e^A - e^B\|_{\infty}}{\mathcal{Z}_A}}_{\text{first term}} + \underbrace{\|e^B\|_{\infty} \left| \frac{1}{\mathcal{Z}_A} - \frac{1}{\mathcal{Z}_B} \right|}_{\text{second term}}. \end{aligned} \quad (24)$$

First we find a bound for the first term. In order to do this we need to show that  $\mathcal{Z}_A$  is a positive scalar.

**2.1.1.1 Proof that  $\mathcal{Z} > 1$ :** We know that for any real constant  $c$ ,

$$e^{-\beta(H-c1)} = e^{+\beta c} e^{-\beta H}, \quad \mathcal{Z}(\beta, H-c) = e^{+\beta c} \mathcal{Z}(\beta, H),$$

so the Gibbs state  $\omega_\beta(H)$  is unchanged (the scalar factor  $e^{\beta c}$  cancels between numerator and denominator). Because every Hamiltonian in the paper is bounded below (a consequence of finite dimension), let  $E_0$  be the ground-state energy of  $H$ . Then we may, without loss of generality, replace  $H$  with the shifted Hamiltonian,

$$H' := H - E_0 1,$$

so that  $E'_0 = 0$  and all other eigen-energies satisfy  $E'_j > 0$  for  $j \geq 1$ . Therefore, by the energy-shift identity above,

$$\mathcal{Z}(\beta, H) = \mathcal{Z}(\beta, H').$$

Now we expand the trace in the eigen-basis of  $H'$ ,

$$\mathcal{Z}(\beta, H') = \text{Tr}(e^{-\beta H'}) = \sum_{n=0}^{d-1} e^{-\beta E'_n} = 1 + \sum_{n=1}^{d-1} e^{-\beta E'_n}.$$

Since  $\dim H \geq 2$  in all considered scenarios (system + bath), the sum from  $n = 1$  contains at least one positive term  $e^{-\beta E'_1} > 0$ . Hence,

$$\mathcal{Z}(\beta, H') > 1 \implies \mathcal{Z}(\beta, H) > 1.$$

So for the first term in equation (24), we can use  $Z_A > 1$  and write

$$\begin{aligned} \|e^A - e^B\|_\infty &\leq e\beta \|H^{(i+1)} - H^{(i)}\|_\infty \implies \frac{\|e^A - e^B\|_\infty}{\mathcal{Z}_A} \leq \frac{e\beta}{\mathcal{Z}_A} \|H^{(i+1)} - H^{(i)}\|_\infty \leq e\beta \|H^{(i+1)} - H^{(i)}\|_\infty \\ \implies \frac{\|e^A - e^B\|_\infty}{\mathcal{Z}_A} &\leq e\beta \|H^{(i+1)} - H^{(i)}\|_\infty. \end{aligned} \quad (25)$$

Now for the second term, we define

$$H_s = H^{(i)} + s\Delta H, \quad \Delta H = H^{(i+1)} - H^{(i)} \quad ; \quad \text{for } s \in [0, 1],$$

$$\mathcal{Z}(s) = \text{tr} e^{-\beta H_s} \rightarrow \mathcal{Z}_B - \mathcal{Z}_A = \mathcal{Z}(1) - \mathcal{Z}(0) = \int_0^1 \frac{d\mathcal{Z}}{ds}(s) ds.$$

We now use Duhamel's formula<sup>1</sup> to compute the derivative.

$$\frac{d}{ds} e^{-\beta H(s)} = -\beta \int_0^1 e^{-\beta s H(s)} H'(s) e^{-\beta(1-s)H(s)} ds. \quad (26)$$

<sup>1</sup>See, R. Bhatia, Matrix Analysis, Chapter IX.

This allows us to express  $\frac{d}{ds}e^{-\beta H_s}$  as

$$\frac{d}{ds}e^{-\beta H_s} = -\beta \int_0^1 d\tau e^{-\beta(1-\tau)H_s} \Delta H e^{-\beta\tau H_s}.$$

Then, taking the trace and using its cyclicity, we get

$$\frac{d\mathcal{Z}(s)}{ds} = -\beta \operatorname{Tr}(\Delta H e^{-\beta H_s}). \quad (27)$$

Given the definitions of the Schatten 1-norm and  $\infty$ -norm above, we can now use the Hölder-Schatten inequality.

### Hölder-Schatten inequality

For any matrices  $A$  and  $B$ ,

$$|\operatorname{tr}(AB)| \leq \|A\|_1 \|B\|_\infty.$$

This inequality is a generalization of the classical Hölder inequality to matrix norms and holds for all Schatten norms  $\|\cdot\|_p, \|\cdot\|_q$  with  $\frac{1}{p} + \frac{1}{q} = 1$ .

-R. Bhatia, *Matrix Analysis*, Springer, 1997, Chapter IV, Theorem IV.2.6

$$\implies \left| \operatorname{tr}[\Delta H e^{-\beta H_s}] \right| \leq \|\Delta H\|_\infty \|e^{-\beta H_s}\|_1 \quad (28)$$

For any Hermitian matrix we have

$$\|e^X\|_1 = \operatorname{Tr} e^X = \sum_i e^{\lambda_i(X)}.$$

Therefore,

$$\begin{aligned} \implies \|e^{-\beta H_s}\|_1 &= \sum_i e^{-\beta \lambda_i(H_s)} \leq d e^{-\beta E_{\min}} \leq d e^{\beta E_{\max}}, \\ \implies \left| \frac{d\mathcal{Z}}{ds}(s) \right| &\leq d\beta \|\Delta H\|_\infty e^{\beta E_{\max}}, \\ \implies |\mathcal{Z}_B - \mathcal{Z}_A| &\leq \int_0^1 d\beta e^{\beta E_{\max}} \|\Delta H\|_\infty ds = d\beta e^{\beta E_{\max}} \|\Delta H\|_\infty. \end{aligned}$$

$e^B$  is positive, so we have  $\|e^B\|_\infty = \text{tr } e^B = \mathcal{Z}_B$ .

$$\begin{aligned}
 &\Rightarrow \|e^B\|_\infty \left| \frac{1}{\mathcal{Z}_A} - \frac{1}{\mathcal{Z}_B} \right| = \frac{\mathcal{Z}_B}{\mathcal{Z}_A \mathcal{Z}_B} |\mathcal{Z}_B - \mathcal{Z}_A| = \frac{|\mathcal{Z}_B - \mathcal{Z}_A|}{\mathcal{Z}_A} \\
 &\Rightarrow \|e^B\|_\infty \left| \frac{1}{\mathcal{Z}_A} - \frac{1}{\mathcal{Z}_B} \right| \leq \frac{d\beta e^{\beta E_{\max}}}{\mathcal{Z}_A} \|H^{(i+1)} - H^{(i)}\|_\infty \leq d\beta e^{\beta E_{\max}} \|H^{(i+1)} - H^{(i)}\|_\infty \\
 &\Rightarrow \|e^B\|_\infty \left| \frac{1}{\mathcal{Z}_A} - \frac{1}{\mathcal{Z}_B} \right| \leq d\beta e^{\beta E_{\max}} \|H^{(i+1)} - H^{(i)}\|_\infty
 \end{aligned} \tag{29}$$

Combining (25) and (29), inequality (24) becomes,

$$\|\omega_\beta^{(i)} - \omega_\beta^{(i+1)}\|_\infty \leq e\beta \|H^{(i+1)} - H^{(i)}\|_\infty + d\beta e^{\beta E_{\max}} \|H^{(i+1)} - H^{(i)}\|_\infty.$$

Since  $E_{\max} \geq 0$ , we know that  $e^{\beta E_{\max}} \geq 1$ . Hence,

$$\begin{aligned}
 e &\leq e e^{\beta E_{\max}} \Rightarrow e + d e^{\beta E_{\max}} \leq (d + e) e^{\beta E_{\max}}, \\
 &\Rightarrow \boxed{\|\omega_\beta^{(i)} - \omega_\beta^{(i+1)}\|_\infty \leq c_1 \beta \|H^{(i+1)} - H^{(i)}\|_\infty \quad ; \quad c_1 = (d + e) e^{\beta E_{\max}}}.
 \end{aligned} \tag{30}$$

The Pinsker inequality can be written as<sup>2</sup>

$$S(\rho \parallel \sigma) \leq \frac{\|\rho - \sigma\|_\infty}{\lambda_{\min}(\sigma)}.$$

Applying this inequality gives

$$S(\omega_\beta^{(i)} \parallel \omega_\beta^{(i+1)}) \leq \frac{\|\omega_\beta^{(i)} - \omega_\beta^{(i+1)}\|_\infty}{\lambda_{\min}(\omega_\beta^{(i+1)})}.$$

We also know

$$\begin{aligned}
 \mathcal{Z}_{i+1} &= \sum_{k=1}^d e^{-\beta E_k^{(i+1)}} \leq d e^{-\beta E_{\min}} =: \mathcal{Z}_{\max} \frac{e^{-\beta E_{\max}}}{\mathcal{Z}_{i+1}} \Rightarrow \geq \frac{e^{-\beta E_{\max}}}{\mathcal{Z}_{\max}}, \\
 \lambda_{\min}(\omega_\beta^{(i+1)}) &= \frac{e^{-\beta E_{\max}}}{\lambda_{i+1}} \geq \frac{e^{-\beta E_{\max}}}{\mathcal{Z}_{\max}} =: \lambda_* > 0.
 \end{aligned}$$

Therefore,

$$\Rightarrow \boxed{S(\omega_\beta^{(i)} \parallel \omega_\beta^{(i+1)}) \leq \frac{c_1 \beta}{\lambda_*} \|H^{(i+1)} - H^{(i)}\|_\infty}. \tag{31}$$

To estimate the norm difference, we express the change in Hamiltonian as an integral over its time

---

<sup>2</sup>O. Bratteli and D. W. Robinson, *Operator Algebras and Quantum Statistical Mechanics*, Vol. 2, Springer, 2002, Eq. (6.2.33).

derivative,

$$\|H^{(i+1)} - H^{(i)}\|_\infty = \left\| \int_{t_i}^{t_{i+1}} \frac{dH_s}{ds} ds \right\|_\infty \leq \int_{t_i}^{t_{i+1}} \left\| \frac{dH_s}{ds} \right\|_\infty ds. \quad (32)$$

To bound this further, we define  $M := \sup_{s \in [0,1]} \left\| \frac{dH_s}{dt} \right\|_\infty$ , which implies

$$\int_{t_i}^{t_{i+1}} \left\| \frac{dH_s}{ds} \right\|_\infty ds \leq \int_{t_i}^{t_{i+1}} M ds = M(t_{i+1} - t_i). \quad (33)$$

Now assuming an equal partition of the total protocol duration  $[0, 1]$  into  $N$  intervals, we have  $t_{i+1} - t_i = \frac{1}{N}$ , and combining (32) and (33) yields

$$\|H^{(i+1)} - H^{(i)}\|_\infty \leq \frac{M}{N} \implies S(\omega_\beta^{(i)} \| \omega_\beta^{(i+1)}) \leq \frac{c_1 \beta}{\lambda_*} \|H^{(i+1)} - H^{(i)}\|_\infty \leq \frac{c_1 \beta}{\lambda_*} \frac{M}{N}. \quad (34)$$

Eq. (34) is the crucial final step; it shows that the spectral difference between consecutive Hamiltonians shrinks as  $O(N^{-1})$  when the process is divided into  $N$  equal steps. This result directly implies that the relative entropy term

$$S(\omega_\beta^{(i)} \| \omega_\beta^{(i+1)}) = O(N^{-1}),$$

also vanishes in the limit  $N \rightarrow \infty$ .

Returning to the optimal protocol for maximum work extraction, we now make a second remark about steps 1 and 3 of the protocol. If the interaction is weak, the energy of turning on and off the interaction can be neglected. Exactly in this case, one can obtain the usual expression of optimal work given by the free energy. This is achieved by choosing  $H_S^{(N)} = H_S$  and  $H_S^{(1)}$  such that  $\omega_\beta(H_S^{(1)}) = \rho_S$ . We compute  $W = W_1 + W_2 + W_3$  under these conditions, in the weak-coupling limit, to obtain the expression for  $W^{\text{weak}}$ .

$$\begin{aligned} W &= W_1 + W_2^{\text{isoth}} + W_3 = \text{Tr} \left( (H_S - H_S^{(1)} - V) \rho^{(0)} \right) + F(\omega^{(1)}, H^{(1)}) - F(\omega^{(N)}, H^{(N)}) \\ &\quad + \text{Tr} \left( (V + H_S^{(N)} - H_S) \omega^{(N)} \right) \\ \xrightarrow[V \rightarrow 0]{H_S^{(N)} = H_S} W^{\text{weak}} &= \text{Tr} \left( H_S \rho^{(0)} \right) - \text{Tr} \left( H_S^{(1)} \rho^{(0)} \right) + F(\omega^{(1)}, H^{(1)}) - F(\omega^{(N)}, H^{(N)}) \end{aligned} \quad (35)$$

In the weak-coupling limit, the interaction is negligible. Therefore,

$$\omega^{(1)} \simeq \omega_\beta(H_S^{(1)}) \otimes \omega_\beta(H_B), \quad \omega^{(N)} \simeq \omega_\beta(H_S^{(N)}) \otimes \omega_\beta(H_B).$$

$$\begin{aligned}
 &\Rightarrow \begin{cases} F(\omega^{(1)}, H^{(1)}) \simeq F(\omega_\beta(H_S^{(1)}), H_S^{(1)}) + F(\omega_\beta(H_B), H_B) = F(\rho_S, H_S^{(1)}) + F(\omega_\beta(H_B), H_B) \\ F(\omega^{(N)}, H^{(N)}) \simeq F(\omega_\beta(H_S^{(N)}), H_S^{(N)}) + F(\omega_\beta(H_B), H_B) = F(\omega_\beta(H_S), H_S) + F(\omega_\beta(H_B), H_B) \end{cases} \\
 &\xrightarrow{(35)} W^{\text{weak}} = \text{Tr}(H_S \rho^{(0)}) - \text{Tr}(H_S^{(1)} \rho^{(0)}) + F(\rho_S, H^{(1)}) + F(\omega_\beta(H_B), H_B) - F(\omega_\beta(H_S), H_S) \\
 &\quad - F(\omega_\beta(H_B), H_B)
 \end{aligned}$$

Using  $\rho^{(0)} = \rho_S \otimes \omega_\beta(H_B)$ , we have

$$\begin{aligned}
 \text{Tr}(H_S \rho^{(0)}) &= \text{Tr}((H_S \otimes I_B)(\rho_S \otimes \omega_\beta(H_B))) = \text{Tr}_S(H_S \rho_S) \underbrace{\text{Tr}_B(\omega_\beta(H_B))}_{=1} = \text{Tr}_S(\rho_S H_S) = \text{Tr}(\rho_S H_S). \\
 \Rightarrow W &= \text{Tr}(H_S \rho_S) - \text{Tr}(H_S^{(1)} \rho_S) + \text{Tr}(\rho_S H_S^{(1)}) - TS(\rho_S) - F(\omega_\beta(H_S), H_S) \\
 &= \text{Tr}(H_S \rho_S) - TS(\rho_S) - F(\omega_\beta(H_S), H_S) = F(\rho_S, H_S) - F(\omega_\beta(H_S), H_S)
 \end{aligned} \tag{36}$$

We add and subtract  $F(\rho^{(0)}, H^{(0)})$  to get

$$\begin{aligned}
 W^{\text{weak}} &= F(\rho^{(0)}, H^{(0)}) - F(\rho_S, H_S) - F(\omega_\beta(H_B), H_B) + F(\rho_S, H_S) - F(\omega_\beta(H_S), H_S) \\
 &= F(\rho^{(0)}, H^{(0)}) - (F(\omega_\beta(H_B), H_B) + F(\omega_\beta(H_S), H_S)) \\
 &= F(\rho^{(0)}, H^{(0)}) - F(\omega^{(0)}, H^{(0)}),
 \end{aligned} \tag{37}$$

where  $\omega^{(0)} := \omega_\beta(H^{(0)})$ . However, if the interaction is nonnegligible, we now show that steps 1 and 3 become a source of irreversibility. Consequently, the optimal protocol consists of a single thermal contact between  $S$  and  $B$ . Further contacts can only decrease the extractable work. This justifies the form of the protocol considered.

Now we express the total work of the optimal protocol in a convenient form. We consider the general case, sum the work contributions from all three steps, and add and subtract  $W^{\text{weak}}$ . We obtain

$$\begin{aligned}
 W &= W_1 + W_2^{\text{isoth}} + W_3 = W^{\text{weak}} + F(\omega^{(0)}, H^{(0)}) - F(\rho^{(0)}, H^{(0)}) + \text{Tr}((H_S - H_S^{(1)} - V)\rho^{(0)}) \\
 &\quad + F(\omega^{(1)}, H^{(1)}) - F(\omega^{(N)}, H^{(N)}) + \text{Tr}((V + H_S^{(N)} - H_S)\omega^{(N)}).
 \end{aligned} \tag{38}$$

To simplify the expression above, we use the following two identities:

$$F(\rho^{(0)}, H^{(1)}) + \text{Tr}((H_S - H_S^{(1)} - V)\rho^{(0)}) = F(\rho^{(0)}, H^{(0)}), \tag{39}$$

$$F(\omega^{(N)}, H^{(N)}) - \text{Tr}((V + H_S^{(N)} - H_S)\omega^{(N)}) = F(\omega^{(N)}, H^{(0)}). \tag{40}$$

First, we derive these identities and then apply them to Eq. (38). Using  $H^{(1)} = H_S^{(1)} + H_B + V$ , we can write

$$\text{Tr}(\rho^{(0)} H^{(1)}) = \text{Tr}(\rho^{(0)} (H_S^{(1)} + H_B + V)). \tag{41}$$



By adding and subtracting  $\text{Tr}(\rho^{(0)}H^{(0)})$ , we get

$$\begin{aligned}\text{Tr}(\rho^{(0)}H^{(1)}) &= \text{Tr}(\rho^{(0)}H^{(0)}) + \text{Tr}(\rho^{(0)}(H_S^{(1)} + H_B + V - H_S - H_B)) \\ &= \text{Tr}(\rho^{(0)}H^{(0)}) + \text{Tr}(\rho^{(0)}(H_S^{(1)} + V - H_S)).\end{aligned}\quad (42)$$

Substituting Eq. (42) into  $F(\rho^{(0)}, H^{(1)}) = \text{Tr}(\rho^{(0)}H^{(1)}) - TS(\rho^{(0)})$  gives

$$\begin{aligned}F(\rho^{(0)}, H^{(1)}) &= \text{Tr}(\rho^{(0)}(H_S^{(1)} + V - H_S)) + \underbrace{\text{Tr}(\rho^{(0)}H^{(0)}) - TS(\rho^{(0)})}_{=F(\rho^{(0)}, H^{(0)})}, \\ \implies &\boxed{F(\rho^{(0)}, H^{(1)}) - \text{Tr}(\rho^{(0)}(H_S - H_S^{(1)} - V)) = F(\rho^{(0)}, H^{(0)})}.\end{aligned}$$

This proves Eq. (39). For Eq. (40), note that

$$\begin{aligned}F(\omega^{(N)}, H^{(N)}) - F(\omega^{(N)}, H^{(0)}) &= \text{Tr}(\omega^{(N)}H^{(N)}) - TS(\omega^{(N)}) - \text{Tr}(\omega^{(N)}H^{(0)}) + TS(\omega^{(N)}), \\ \implies &\boxed{F(\omega^{(N)}, H^{(N)}) - \text{Tr}((V + H_S^{(N)} - H_S)\omega^{(N)}) = F(\omega^{(N)}, H^{(0)})}.\end{aligned}$$

Substituting Eqs. (39) and (40) into Eq. (38) yields

$$W = W^{\text{weak}} + F(\omega^{(0)}, H^{(0)}) - F(\rho^{(0)}, H^{(1)}) + F(\omega^{(1)}, H^{(1)}) - F(\omega^{(N)}, H^{(0)}) \quad (43)$$

$$= W^{\text{weak}} - \Delta F^{(\text{res})} - \Delta F^{(\text{irr})}, \quad (44)$$

where we define

$$\Delta F^{(\text{res})} := F(\omega_\beta(H^{(N)}), H^{(0)}) - F(\omega_\beta(H^{(0)}), H^{(0)}), \quad (45)$$

$$\Delta F^{(\text{irr})} := F(\rho^{(0)}, H^{(1)}) - F(\omega_\beta(H^{(1)}), H^{(1)}), \quad (46)$$

with  $H^{(1)/(5)} = H_S^{(1)/(N)} + H_B + V$ . To proceed, we now derive the relation that links free-energy differences to relative entropy. Using the definition of free energy,

$$\begin{aligned}F(\rho, H) &= \text{Tr}(\rho H) - TS(\rho) = \text{Tr}(\rho H) - T \text{Tr}(\rho \ln(\rho)), \\ F(\omega_\beta(H), H) &= \text{Tr}(\omega_\beta(H)H) - TS(\omega_\beta(H)) = \text{Tr}(\omega_\beta(H)H) - T \text{Tr}(\omega_\beta(H)(\ln(\omega_\beta(H)))).\end{aligned}$$

Using equation (10),

$$\begin{aligned}\ln(\omega_\beta(H)) &= -\beta H - \ln(\mathcal{Z}), \\ \implies \text{Tr}(\omega_\beta(H) \ln(\omega_\beta(H))) &= \text{Tr}(\omega_\beta(H)(-\beta H - \ln(\mathcal{Z}))) \\ &= -\beta \text{Tr}(\omega_\beta(H)H) - \ln(\mathcal{Z}) \underbrace{\text{Tr}(\omega_\beta(H))}_{=1} \\ &= -\beta \text{Tr}(\omega_\beta(H)H) - \ln(\mathcal{Z})\end{aligned}\quad (47)$$

Substituting this into the definitions of free energy, we obtain

$$\begin{aligned} F(\rho, H) - F(\omega_\beta(H), H) &= \text{tr}(\rho H) - \text{Tr}(\omega_\beta(H)H) + T \text{Tr}(\rho \ln(\rho)) + \beta \times T \text{Tr}(\omega_\beta(H)H) + T \ln(\mathcal{Z}) \\ &= \text{tr}(\rho H) + T \text{Tr}(\rho \ln(\rho)) + T \ln(\mathcal{Z}). \end{aligned} \quad (48)$$

Using  $S(\rho \parallel \sigma) = \text{Tr}[\rho(\ln \rho - \ln \sigma)]$ , we have

$$TS(\rho \parallel \omega_\beta(H)) = T \text{Tr}(\rho(\ln(\rho) - \ln(\omega_\beta(H)))) = T \text{Tr}(\rho \ln(\rho)) + \beta \times T \text{Tr}(\rho H) + T \ln(\mathcal{Z}) \underbrace{\text{Tr}(\rho)}_{=1}. \quad (49)$$

Hence, combining Eq. (48) and (49) yields

$$F(\rho, H) - F(\omega_\beta(H), H) = TS(\rho \parallel \omega_\beta(H)). \quad (50)$$

Therefore, by the nonnegativity of the relative entropy, both  $\Delta F^{(\text{irr})}$  and  $\Delta F^{(\text{res})}$  are nonnegative, with equality iff  $\rho^{(0)} = \omega_\beta(H^{(1)})$  and  $\omega_\beta(H^{(N)}) = \omega_\beta(H^{(0)})$ , respectively. Now we can confidently say that always  $W \leq W^{\text{weak}}$ , so that interactions are detrimental for work extraction. We can safely conclude that  $W \leq W^{\text{weak}}$ , i.e., strong interactions are detrimental to work extraction. Note that  $H^{(0)}$  and  $\rho^{(0)}$  are fixed boundary data of the problem. Therefore, the correction terms in Eqs. (45) and (46) only depend on  $H_S^{(1)}$  and  $H_S^{(N)}$ , namely the specific contact and decoupling Hamiltonians at which  $S$  is coupled to and decoupled from  $B$ . In order to maximize  $W$ , we hence need to minimize  $\Delta F^{(\text{irr})}$  and  $\Delta F^{(\text{res})}$  over  $H_S^{(1)}$  and  $H_S^{(N)}$ , i.e.,

$$W_{\text{max}} = W^{\text{weak}} - \Delta F_{\text{min}}^{(\text{res})} - \Delta F_{\text{min}}^{(\text{irr})}, \quad (51)$$

with

$$\Delta F_{\text{min}}^{(\text{res})} = \min_{H_S^{(1)}} \Delta F_{\text{min}}^{(\text{res})}, \quad (52)$$

$$\Delta F_{\text{min}}^{(\text{irr})} = \min_{H_S^{(N)}} \Delta F_{\text{min}}^{(\text{irr})}. \quad (53)$$

Now we proceed to solve these minimizations, which can be carried out independently. For that, we can use that in a local minimum of a function, its derivative must vanish.

## 2.2 Minimization of $\Delta F^{(\text{irr})}$

Here we address how to minimize the irreversible cost  $\Delta F^{(\text{irr})}$  associated with switching the system-bath interaction on at the start of the protocol. We view this cost as a functional of the initial contact choice  $H_S^{(1)}$  and adopt a simple variational strategy: perturb the chosen Hamiltonian slightly, impose that the first variation of the cost vanishes, and read off the resulting optimality condition. This yields a clear criterion characterizing the best switch-on point and clarifies how strong coupling and correlations modify the familiar weak-coupling picture. Detailed definitions and calculations follow.

Let  $X_S$  be the choice of  $H_S^{(1)}$  yielding  $F_{\min}^{(\text{irr})}$ , and define

$$X(t) := X_S \otimes I_B + t Y_S \otimes I_B + V + I_S \otimes H_B. \quad (54)$$

We set the first derivative to zero. In a Taylor expansion  $X_S(t) = X_S + t Y_S + O(t^2)$ , the derivative at  $t = 0$  is

$$\frac{dX_S(t)}{dt} = Y_S,$$

hence it depends only on  $Y_S$ . Higher-order terms (beyond linear order) do not affect this derivative at  $t = 0$  and can be ignored.

For any  $Y_S$ , it must hold that

$$\left. \frac{d}{dt} \left( F(\rho^{(0)}, X(t)) - F(\omega_\beta(X(t)), X(t)) \right) \right|_{t=0} = 0. \quad (55)$$

That is, if we perturb the solution  $X_S$  by  $t Y_S$ , then the derivative with respect to  $t$  must be zero for all  $Y_S$ . In other words, we are standing in a minimum. Conversely, we use Eq. (55) to find  $X_S$ . We differentiate each term separately and then sum the results. First, the derivative of the first term:

$$\frac{d}{dt} F(\rho^{(0)}, X(t)) = \frac{d}{dt} \left( \text{Tr}(\rho^{(0)} X(t)) \right) - T \frac{d}{dt} \left( S(\rho^{(0)}) \right), \quad (56)$$

$$\frac{d}{dt} \left( \text{Tr}(\rho^{(0)} X(t)) \right) = \text{Tr}(\rho^{(0)} \underbrace{\frac{dX(t)}{dt}}_{=Y_S \otimes I_B}) = \text{Tr}(\rho^{(0)} Y_S \otimes I_B), \quad (57)$$

$$\frac{d}{dt} \left( S(\rho^{(0)}) \right) = 0, \quad (58)$$

$$\implies \frac{d}{dt} F(\rho^{(0)}, X(t)) = \text{Tr}(\rho^{(0)} Y_S \otimes I_B). \quad (59)$$

For the second term,

$$\frac{d}{dt} F(\omega_\beta(X(t)), X(t)) = \text{Tr} \left( \frac{d}{dt} (\omega_\beta(X(t)) X(t)) \right) - \frac{1}{\beta} \frac{d}{dt} S(\omega_\beta(X(t))), \quad (60)$$

$$\begin{aligned} \frac{d}{dt} S(\omega_\beta(X(t))) &= - \frac{d}{dt} (\omega_\beta(X(t)) \ln(\omega_\beta(X(t)))) \\ &= - \text{Tr} \left( \omega_\beta(X(t)) \frac{d}{dt} \ln(\omega_\beta(X(t))) \right) - \text{Tr} \left( \frac{d\omega_\beta(X(t))}{dt} \ln(\omega_\beta(X(t))) \right). \end{aligned} \quad (61)$$

Using  $S(\omega) = -\text{Tr}(\omega \ln \omega)$ ,  $\text{Tr}(\omega_\beta(X(t))) = 1$ , and Eq. (10) (i.e.,  $\ln \omega_\beta(H) = -\beta H - \ln \mathcal{Z}$ ),

$$\frac{d}{dt} S(\omega_\beta(X(t))) = \beta \text{Tr} \left( \omega_\beta(X(t)) \frac{d}{dt} X(t) \right) + \frac{d}{dt} \ln(\mathcal{Z}) - \text{Tr} \left( \frac{d\omega_\beta(X(t))}{dt} \ln(\omega_\beta(X(t))) \right). \quad (62)$$

We previously derived in Eq. (27) (via Duhamels formula and cyclicity) that

$$\frac{d\mathcal{Z}(t)}{dt} = -\beta \operatorname{Tr} \left( \frac{dX(t)}{dt} e^{-\beta X(t)} \right).$$

Hence,

$$\frac{d}{dt} \ln(\mathcal{Z}) = \frac{1}{\mathcal{Z}(t)} \frac{d}{dt} \mathcal{Z}(t) = -\beta \operatorname{Tr} \left( \frac{d}{dt} X(t) \frac{e^{-\beta X(t)}}{\mathcal{Z}(t)} \right) = -\beta \operatorname{Tr} \left( \frac{d}{dt} X(t) \omega_\beta(X(t)) \right), \quad (63)$$

$$\begin{aligned} &\stackrel{(62)}{\longrightarrow} \frac{d}{dt} S(\omega_\beta(X(t))) = \beta \operatorname{Tr} \left( \omega_\beta(X(t)) \frac{d}{dt} X(t) \right) - \beta \operatorname{Tr} \left( \frac{d}{dt} X(t) \omega_\beta(X(t)) \right) \\ &\quad - \operatorname{Tr} \left( \frac{d\omega_\beta(X(t))}{dt} \underbrace{\ln(\omega_\beta(X(t)))}_{=-\beta X(t) - \ln(\mathcal{Z})} \right), \\ &\implies \frac{d}{dt} S(\omega_\beta(X(t))) = -\operatorname{Tr} \left( \frac{d}{dt} \omega_\beta(X(t)) (-\beta X(t) - \ln(\mathcal{Z})) \right). \end{aligned} \quad (64)$$

The derivative of the von Neumann entropy will come in handy later on. So, in the general case for Hamiltonian  $H(t) := H_S \otimes I_B + tY_S \otimes I_B + V + I_S \otimes H_B$ , we have

$$\frac{d}{dt} S(\omega_\beta(H(t))) = -\operatorname{Tr} \left( \frac{d\omega_\beta(H(t))}{dt} \ln(\omega_\beta(H(t))) \right), \quad (65)$$

Plugging this into (60),

$$\begin{aligned} \frac{d}{dt} F(\omega_\beta(X(t)), X(t)) &= \operatorname{Tr} \left( \frac{d}{dt} (\omega_\beta(X(t)) X(t)) \right) - \frac{1}{\beta} \operatorname{Tr} \left( \left( \frac{d}{dt} \omega_\beta(X(t)) \right) (-\beta X(t) - \ln(\mathcal{Z})) \right) \\ &= \operatorname{Tr} \left( \left( \frac{d}{dt} \omega_\beta(X(t)) \right) X(t) \right) + \operatorname{Tr} \left( \frac{d}{dt} (X(t)) \omega_\beta(X(t)) \right) \\ &\quad - \operatorname{Tr} \left( \left( \frac{d}{dt} \omega_\beta(X(t)) \right) X(t) \right) - \frac{1}{\beta} \ln(\mathcal{Z}) \underbrace{\operatorname{Tr} \left( \frac{d}{dt} \omega_\beta(X(t)) \right)}_{=0} \\ &= \operatorname{Tr} \left( \frac{dX(t)}{dt} \omega_\beta(X(t)) \right) = \operatorname{Tr} (\omega_\beta(X(t)) Y_S \otimes I_B). \end{aligned} \quad (66)$$

Using (59) and (66), Eq. (55) gives

$$\begin{aligned} &\left. \operatorname{Tr}(\rho^{(0)} Y_S \otimes I_B) \right|_{t=0} - \left. \operatorname{Tr}(\omega_\beta(X(t)) Y_S \otimes I_B) \right|_{t=0} = 0, \\ &\implies \operatorname{Tr}((\rho^{(0)} - \omega_\beta(X(0))) Y_S \otimes I_B) = 0. \end{aligned} \quad (67)$$

Using

$$\operatorname{Tr}_{SB}(M_{SB} A_S \otimes I_B) = \operatorname{Tr}_S((\operatorname{Tr}_B M_{SB}) A_S),$$

we obtain

$$\begin{aligned}
\text{Tr}_S \left( \text{Tr}_B(\rho^{(0)} - \omega_\beta(X(0))) Y_S \otimes I_B \right) &= \text{Tr}_S \left( \text{Tr}_B(\rho^{(0)} - \omega_\beta(X(0))) Y_S \right) = 0, \\
\implies \text{Tr}_S \left( \underbrace{\text{Tr}_B(\rho^{(0)})}_{=\rho_S^{(0)}} Y_S \right) - \text{Tr}_S \left( \text{Tr}_B(\omega_\beta(X(0))) Y_S \right) &= 0, \\
\implies \text{Tr}_S \left( \rho_S^{(0)} Y_S - \text{Tr}_B(\omega_\beta(X(0))) Y_S \right) &= 0.
\end{aligned} \tag{68}$$

This equality must hold for all  $Y_S$ . Hence,

$$\rho_S^{(0)} = \text{Tr}_B(\omega_\beta(X(0))). \tag{69}$$

Recall that  $X_S$  was the choice of  $H_S^{(1)}$  yielding  $\Delta F_{\min}^{(\text{irr})}$ , hence

$$X(0) = X_S \otimes I_B + I_S \otimes H_B + V = H^{(1)},$$

and therefore,

$$\boxed{\rho_S^{(0)} = \text{Tr}_B(\omega_\beta(H^{(1)}))}. \tag{70}$$

This condition characterizes the minimizer and thus yields  $\Delta F_{\min}^{(\text{irr})}$ .

A natural question is whether this equation has a solution. We now show that there is at least one minimum of  $\Delta F^{(\text{irr})}$ , so that Eq. (70) always provides the desired solution, provided that  $\rho_S$  is a full-rank state.

### 2.2.1 Existence of a local minimum of $\Delta F^{(\text{irr})}$

First, we show that for any full-rank state  $\rho_S$  and non-trivial Hamiltonian  $H_S^{(1)}$ , the term  $\Delta F^{(\text{irr})}$  diverges as the operator norm  $H_S^{(1)}$  grows. More explicitly, parameterize

$$H_S^{(1)} = \lambda \hat{H}_S^{(1)}, \quad \|\hat{H}_S^{(1)}\|_\infty = 1, \quad \lambda = \|H_S^{(1)}\|_\infty,$$

with  $\hat{H}_S^{(1)}$  controlling its direction, and  $\lambda$  its operator norm, and let  $P_S^{(1)}$  denote the ground-state subspace of  $\hat{H}_S^{(1)}$ . Then

$$\lim_{\lambda \rightarrow \infty} \omega_\beta(\lambda \hat{H}_S^{(1)} + V + H_B) := \sigma_{SB}, \tag{71}$$

is supported only with the subspace  $P_S^{(1)} \otimes \mathbb{H}_B$ , where  $\mathbb{H}$  denotes the Hilbert space of the bath. Using (50), we obtain

$$\lim_{\lambda \rightarrow \infty} \Delta F^{(\text{irr})} = \frac{1}{\beta} S(\rho^{(0)} \otimes \omega_\beta(H_B) \| \sigma_{SB}) = \infty, \tag{72}$$

since the relative entropy diverges if the support of the first argument is not contained in that of the second. Here, the system starts in  $\rho_S$ , which is full rank, so it has population in all energy levels, including the excited ones. This means that initially there is a nonzero probability that the system is in an excited state. As  $\lambda \rightarrow \infty$ , the energy gap between the ground and excited state becomes infinite. So,  $\sigma_{SB}$  has zero weight for the excited states. This mismatch causes the relative entropy term to diverge. Thus  $\Delta F^{(\text{irr})}$  has at least one minimum, which can be obtained through (70). If the solution of Eq. (70) is not unique and multiple minima exist, we choose the global minimum.

### 2.3 Minimization of $\Delta F^{(\text{res})}$

In this section we address how to minimize the residual cost associated with switching the system-bath interaction off at the end of the protocol. We treat this cost as a functional of the final contact choice and use a simple variational argument: perturb the final Hamiltonian slightly, set the first variation of the cost to zero, and read off the resulting optimality condition. We then interpret the condition and discuss when a minimizer exists. Definitions and detailed calculations follow.

As in the previous section, assume that  $Z_S$  solves  $\min_{H_S^{(N)}} \Delta F^{(\text{res})}$  and let  $tY_S$  be a perturbation over this solution. Define

$$Z(t) := Z_S \otimes I_B + tY_S \otimes I_B + V + I_S \otimes H_B. \quad (73)$$

By construction,  $Z(0)$  minimizes  $\Delta F^{(\text{res})}$ , and  $Z(t)$  parametrizes variations around this minimizer. We write  $\mathcal{Z}(t) := \text{Tr}(e^{-\beta Z(t)})$  for the corresponding partition function. (Notational note:  $Z(t)$  is a Hamiltonian, whereas  $\mathcal{Z}(t)$  is its partition function. Do not confuse the two!). Setting the first variation to zero and using Eq. (50),

$$\begin{aligned} F(\omega_\beta(Z(t)), H^{(0)}) - F(\omega_\beta(H^{(0)}), H^{(0)}) &= TS(\omega_\beta(Z(t)) \| \omega_\beta(H^{(0)})) \\ \implies \frac{d}{dt} \Delta F^{(\text{res})} &= \frac{d}{dt} \frac{1}{\beta} S(\omega_\beta(Z(t)) \| \omega_\beta(H^{(0)})) \Big|_{t=0} = 0. \end{aligned} \quad (74)$$

We now compute the derivative of the relative entropy:

$$\begin{aligned} \frac{d}{dt} S(\omega_\beta(Z(t)) \| \omega_\beta(H^{(0)})) &= \frac{d}{dt} \text{Tr} \left( \omega_\beta(Z(t)) \left[ \ln(\omega_\beta(Z(t))) - \ln(\omega_\beta(H^{(0)})) \right] \right) \\ &= \frac{d}{dt} \text{Tr} \left( \omega_\beta(Z(t)) \ln(\omega_\beta(Z(t))) \right) - \frac{d}{dt} \text{Tr} \left( \omega_\beta(Z(t)) \ln(\omega_\beta(H^{(0)})) \right) \\ &= -\frac{d}{dt} S(\omega_\beta(Z(t))) - \text{Tr} \left( \frac{d\omega_\beta(Z(t))}{dt} \ln(\omega_\beta(H^{(0)})) \right) \end{aligned} \quad (75)$$

Using  $\frac{d}{dt}S(\omega) = -\text{Tr}\left(\frac{d\omega}{dt} \ln \omega\right)$  (Eq. (65)),

$$\begin{aligned} \frac{d}{dt}S(\omega_\beta(Z(t))\|\omega_\beta(H^{(0)})) &= \text{Tr}\left(\frac{d\omega_\beta(Z(t))}{dt} \ln(\omega_\beta(Z(t)))\right) - \text{Tr}\left(\frac{d\omega_\beta(Z(t))}{dt} \ln(\omega_\beta(H^{(0)}))\right) \\ &= \text{Tr}\left(\frac{d\omega_\beta(Z(t))}{dt} \left[\ln(\omega_\beta(Z(t))) - \ln(\omega_\beta(H^{(0)}))\right]\right). \end{aligned} \quad (76)$$

By substituting (10) into this, we obtain

$$\begin{aligned} \frac{d}{dt}S(\omega_\beta(Z(t))\|\omega_\beta(H^{(0)})) &= \text{Tr}\left(\frac{d\omega_\beta(Z(t))}{dt} \left[-\beta Z(t) - \ln(\mathcal{Z}(Z(t))) + \beta H^{(0)} + \ln(\mathcal{Z}(H^{(0)}))\right]\right) \\ &= \beta \text{Tr}\left(\frac{d\omega_\beta(Z(t))}{dt} [H^{(0)} - Z(t)]\right) - \underbrace{\ln(\mathcal{Z}(Z(t))) \text{Tr}\left(\frac{d\omega_\beta(Z(t))}{dt}\right)}_{=0} + \underbrace{\ln(\mathcal{Z}(H^{(0)})) \text{Tr}\left(\frac{d\omega_\beta(Z(t))}{dt}\right)}_{=0} \\ \Rightarrow \boxed{\frac{d}{dt}S(\omega_\beta(Z(t))\|\omega_\beta(H^{(0)}))} &= \beta \text{Tr}\left(\frac{d\omega_\beta(Z(t))}{dt} [H^{(0)} - Z(t)]\right). \end{aligned} \quad (77)$$

Hence, from (74),

$$\frac{d}{dt}\Delta F^{(\text{res})} = \beta \text{Tr}\left(\frac{d\omega_\beta(Z(t))}{dt} [H^{(0)} - Z(t)]\right) \Big|_{t=0} = 0. \quad (78)$$

The rate of change of the thermal state with respect to the Hamiltonian is given by the expression

$$\text{Tr}\left(A \frac{d}{dt} \Big|_{t=0} \omega_\beta(H_t)\right) = -\beta \text{cov}_{\omega_\beta(H_0)}(A, H'_0), \quad (79)$$

where generalized covariance, also known as Kubo-Mori inner product or Bogoliubov inner product, is defined as

$$\text{cov}_{\omega_\beta(H)}(A, B) := \text{Tr}(\omega_\beta(H) A_{H,\beta} B) - \text{Tr}(\omega_\beta(H) A) \text{Tr}(\omega_\beta(H) B), \quad (80)$$

where

$$A_{H,\beta} := \int_0^1 d\tau e^{\beta\tau H} A e^{-\beta\tau H}.$$

We prove (79) in the next subsection.

### 2.3.1 Proof of Eq. (79)

Write  $\omega_t = \frac{e^{-\beta H_t}}{\mathcal{Z}_t}$  with  $\mathcal{Z}_t = \text{Tr}(e^{-\beta H_t})$ . Then

$$\frac{d}{dt}\omega_t = \frac{1}{\mathcal{Z}_t} \frac{d}{dt} e^{-\beta H_t} - \frac{e^{-\beta H_t}}{\mathcal{Z}_t^2} \frac{d\mathcal{Z}_t}{dt}. \quad (81)$$

Evaluating at  $t = 0$  and denoting  $\omega_0 = \omega_\beta(H_0)$  gives

$$\left. \frac{d}{dt} \omega_t \right|_{t=0} = \frac{1}{\mathcal{Z}_0} \left. \frac{d}{dt} e^{-\beta H_t} \right|_{t=0} - \frac{e^{-\beta H_0}}{\mathcal{Z}_0^2} \left. \frac{d\mathcal{Z}_t}{dt} \right|_{t=0}. \quad (82)$$

Multiplying by  $A$  and taking the trace we get

$$\text{Tr} \left( A \left. \frac{d}{dt} \omega_t \right|_{t=0} \right) = \frac{1}{\mathcal{Z}_0} \text{Tr} \left( A \left. \frac{d}{dt} e^{-\beta H_t} \right|_{t=0} \right) - \text{Tr}(A\omega_0) \frac{1}{\mathcal{Z}_0} \left. \frac{d\mathcal{Z}_t}{dt} \right|_{t=0}. \quad (83)$$

Using the Duhamel derivative, Eq. (26), and the cyclicity of the trace,

$$\begin{aligned} & \text{Tr} \left( A \left. \frac{d}{dt} \omega_t \right|_{t=0} \right) \\ &= -\beta \text{Tr} \int_0^1 ds \frac{1}{\mathcal{Z}_0} \text{Tr} \left( A e^{-\beta s H_0} H'_0 e^{-\beta(1-s)H_0} \right) + \beta \text{Tr}(A\omega_0) \text{Tr} \left( \omega_0 \int_0^1 ds e^{\beta s H_0} H'_0 e^{-\beta s H_0} \right) \\ &= -\beta \text{Tr}(\omega_0 A_{H_0, \beta} H'_0) + \beta \text{Tr}(A\omega_0) \text{Tr}(\omega_0 H'_0) = -\beta \text{cov}_{\omega_\beta(H_0)}(A, H'_0), \end{aligned} \quad (84)$$

where  $\text{cov}_{\omega_\beta(H)}(A, B) = \text{Tr}(\omega_\beta(H) A_{H, \beta} B) - \text{Tr}(\omega_\beta(H) A) \text{Tr}(\omega_\beta(H) B)$  and

$$A_{H, \beta} := \int_0^1 d\tau e^{\beta \tau H} A e^{-\beta \tau H}. \quad (85)$$

Applying (79) to (78) with  $A = H^{(0)} - Z(0)$  and  $H'_0 = \left. \frac{d}{dt} Z(t) \right|_{t=0} = Y_S \otimes I_B$ , we get

$$\begin{aligned} \frac{d}{dt} \Delta F^{(\text{res})} &= -\beta \text{cov}_{\omega_\beta(Z(0))}(H^{(0)} - Z(0), \left. \frac{d\omega_\beta(Z(t))}{dt} \right|_{t=0}) \\ &= -\beta \text{cov}_{\omega_\beta(Z(0))}(H^{(0)} - Z(0), Y_S) = 0. \end{aligned} \quad (86)$$

Setting  $Z := Z(0)$  and writing out the definition of the generalized covariance (80), we then obtain

$$\begin{aligned} \text{cov}_{\omega_\beta(Z)}(H^{(0)} - Z, Y_S) &= \text{Tr}(\omega_\beta(Z) (H^{(0)} - Z)_{Z, \beta} Y_S) - \text{Tr}(\omega_\beta(Z)(H^{(0)} - Z)) \text{Tr}(\omega_\beta(Z) Y_S) = 0, \\ \implies \text{Tr}(\omega_\beta(Z)(H^{(0)} - Z)) \text{Tr}(\omega_\beta(Z) Y_S) &= \text{Tr}(\omega_\beta(Z) (H^{(0)} - Z)_{Z, \beta} Y_S). \end{aligned} \quad (87)$$

Since this must hold for all  $Y_S$ , and assuming the scalar  $\text{Tr}(\omega_\beta(Z)(H^{(0)} - Z)) \neq 0$ ,

$$\begin{aligned} \implies \text{Tr}(\omega_\beta(Z)(H^{(0)} - Z)) \text{Tr}_S(\text{Tr}_B(\omega_\beta(Z)) Y_S) &= \text{Tr}_S \left( \text{Tr}_B(\omega_\beta(Z) (H^{(0)} - Z)_{Z, \beta}) Y_S \right), \\ \implies \text{Tr}_B(\omega_\beta(Z)) &= \frac{\text{Tr}_B(\omega_\beta(Z) (H^{(0)} - Z)_{Z, \beta})}{\text{Tr}(\omega_\beta(Z)(H^{(0)} - Z))}. \end{aligned} \quad (88)$$

This is the desired implicit solution for  $Z_S$ . If the denominator vanishes, the condition (87) reduces to  $\text{Tr}_B(\omega_\beta(Z) (H^{(0)} - Z)_{Z, \beta}) = 0$ .



Recall that  $Z := Z(0) = (Z_S \otimes I_B) + V + (I_S \otimes H_B)$ ; identifying  $Z_S = H_S^{(N)}$  gives  $Z = H^{(N)}$ . Hence, provided  $\text{Tr}(\omega_\beta(H^{(N)})(H^{(0)} - H^{(N)})) \neq 0$ , the stationarity condition reads

$$\Rightarrow \boxed{\text{Tr}_B(\omega_\beta(H^{(N)})) = \frac{\text{Tr}_B(\omega_\beta(H^{(N)})(H^{(0)} - H^{(N)})_{H^{(N)},\beta})}{\text{Tr}(\omega_\beta(H^{(N)})(H^{(0)} - H^{(N)}))}}. \quad (89)$$

**Mismatch between definition and later formulas.** In the Supplementary Material, the generalized covariance is defined with a single Kubo transform acting on the first slot only,

$$\text{cov}_{\omega_\beta(H)}(A, B) := \text{Tr}(\omega_\beta(H)A_{H,\beta}B) - \text{Tr}(\omega_\beta(H)A) \text{Tr}(\omega_\beta(H)B), \quad A_{H,\beta} := \int_0^1 e^{\beta\tau H} A e^{-\beta\tau H} d\tau.$$

With this definition, any subsequent expansion produces exactly one occurrence of the  $(H^{(N)}, \beta)$ , subscripted (Kubo-transformed) operator. However, the stationarity conditions *as printed* in the two sources read as follows. For ease of comparison we reproduce them here:

$$\text{Tr}_B(\omega_\beta(H^{(N)})) = \frac{\text{Tr}_B(\omega_\beta(H^{(N)})[H_{H^{(N)},\beta}^{(0)} - H^{(N)}])}{\text{Tr}(\omega_\beta(H^{(N)})[H_{H^{(N)},\beta}^{(0)} - H^{(N)}])}, \quad (\text{Main text, Eq. (8)})$$

$$\text{Tr}_B(\omega_\beta(Z)) \text{Tr}(\omega_\beta(Z)[H_{Z,\beta}^{(0)} - Z]) = \text{Tr}_B(\omega_\beta(Z)[H_{Z,\beta}^{(0)} - Z]). \quad (\text{Supplement, Eq. (36)})$$

As written, both formulas display two occurrences of the  $(H^{(N)}, \beta)$  subscript (one in the numerator and one in the denominator). These statements are not simultaneously consistent with the single-slot Kubo transform in the stated definition. At the same time, the derivation leading to Eq. (79) appears internally consistent. We therefore flag this as an issue, and proceed using the equations in the paper and the Supplementary Material.

To finish this section, we show that for residual free energy  $\Delta F^{(\text{res})}$  has at least one local minimum, and hence (89) always provides a solution.

### 2.3.2 Existence of a finite- $\lambda$ local minimum of $\Delta F^{(\text{res})}$

Unlike  $\Delta F^{(\text{res})}$ ,  $\Delta F^{(\text{res})}$  does not diverge to  $+\infty$  with the operator norm of  $H_S^{(N)}$ , but tends to a constant. Above, the Hamiltonian is parameterized as  $H_S^{(N)} = \lambda \hat{H}_S^{(N)}$ , where  $\lambda$  controls its norm and  $\hat{H}_S^{(N)}$  controls its direction, i.e.  $\|\hat{H}_S^{(N)}\| = 1$ . As in the last section, let  $P_S^{(N)}$  denote the ground state subspace of  $\hat{H}_S^{(N)}$ . In the limit  $\lambda \rightarrow \infty$ , the thermal state  $\omega_\beta^{(N)}(\lambda) = \omega_\beta(\lambda \hat{H}_S^{(N)} + V + H_B)$  tends to the state

$$\lim_{\lambda \rightarrow \infty} \omega_\beta^{(N)}(\lambda) = \sigma_{SB}^{(N)}, \quad (90)$$

which is again supported within the subspace  $P_S^{(N)} \otimes \mathbb{H}_B$ . Since  $\Delta F^{(\text{res})} = F(\omega_\beta^{(N)}(\lambda), H^{(0)}) - F(\omega_\beta(H^{(0)}), H^{(0)})$ , we then have

$$\lim_{\lambda \rightarrow \infty} \Delta F^{(\text{res})} = F(\sigma_{SB}^{(N)}, H^{(0)}) - F(\omega_\beta(H^{(0)}), H^{(0)}). \quad (91)$$

Since  $S(\sigma_{SB}^{(N)}) \geq 0$  (??),

$$F(\sigma_{SB}^{(N)}, H^{(0)}) = \text{Tr}(H^{(0)} \sigma_{SB}^{(N)}) - TS(\sigma_{SB}^{(N)}) \leq \text{Tr}(H^{(0)} \sigma_{SB}^{(N)}). \quad (92)$$

By the operator-norm definition, for any state  $\rho$  and Hermitian  $H$ ,  $\text{Tr}(H\rho) \leq \|H\|$ . Therefore,

$$F(\sigma_{SB}^{(N)}, H^{(0)}) \leq \|H^{(0)}\|. \quad (93)$$

So,

$$\lim_{\lambda \rightarrow \infty} \Delta F^{(\text{res})} \leq \|H^{(0)}\| - F(\omega_\beta(H^{(0)}), H^{(0)}). \quad (94)$$

So along each ray (fixed  $\hat{H}_S^{(N)}$ ) the limit exists and is finite (hence an upper bound for large  $\lambda$ ). In order to show that the minimum of the residual free energy takes place for a finite  $\lambda$ , we show that the asymptotic value above is approached from below, and equivalently the the derivative for large  $\lambda$  tends to 0 from the positive side. The first derivative of the residual energy using  $F(\rho, H) - F(\omega_\beta(H), H) = TS(\rho \| \omega_\beta(H))$  and derivative of relative entropy in Eq. (77), reads

$$\frac{dF(\omega^{(N)}(\lambda), H^{(0)})}{d\lambda} = \text{Tr} \left( \left[ H^{(0)} - H^{(N)}(\lambda) \right] \frac{d\omega^{(N)}(\lambda)}{d\lambda} \right), \quad (95)$$

where  $\omega^{(N)}(\lambda) := \omega_\beta(\lambda \hat{H}_S^{(N)} + V + H_B)$ . We can then use the perturbation formula (10) to obtain

$$\begin{aligned} \frac{dF(\omega^{(N)}(\lambda), H^{(0)})}{d\lambda} &= -\beta \text{cov}_{\omega_\beta(\lambda)} \left( H^{(0)} - H^{(N)}, \hat{H}_S^{(N)} \right) \\ &= \lambda \beta \text{cov}_{\omega_\beta(\lambda)} \left( \hat{H}_S^{(N)}, H_S^{(N)} \right) - \beta \text{cov}_{\omega_\beta(\lambda)} \left( H_S^{(0)} - V, \hat{H}_S^{(N)} \right). \end{aligned} \quad (96)$$

Now we show that in the limit  $\lambda \rightarrow \infty$ , both terms go to zero.

**2.3.2.1 First term** For this term, we have

$$\lim_{\lambda \rightarrow \infty} \text{cov}_{\omega_\beta(\lambda)} \left( \hat{H}_S^{(N)}, H_S^{(N)} \right) = \lim_{\lambda \rightarrow \infty} \text{cov}_{\sigma_{SB}^{(N)}} \left( \hat{H}_S^{(N)}, H_S^{(N)} \right). \quad (97)$$

Write the generalized (KuboMori) covariance

$$\text{cov}_\rho(A, B) := \int_0^1 d\tau \text{Tr}(\rho^{1-\tau} A \rho^\tau B) - \text{Tr}(\rho A) \text{Tr}(\rho B). \quad (98)$$

Let the spectral decomposition of  $\hat{H}_S^{(N)}$  be

$$\hat{H}_S^{(N)} = \sum_k \varepsilon_k |k\rangle\langle k|, \quad \varepsilon_* := \min_k \varepsilon_k.$$

Let  $P_S^{(N)} = \sum_{k: \varepsilon_k = \varepsilon_*} |k\rangle\langle k|$  be the projector onto the ground subspace. On that subspace,

$$\hat{H}_S^{(N)} P_S^{(N)} = \varepsilon_* P_S^{(N)}, \quad H_S^{(N)} P_S^{(N)} = \lambda \hat{H}_S^{(N)} P_S^{(N)} = \lambda \varepsilon_* P_S^{(N)}.$$

Since  $\sigma_{SB}^{(N)}$  has support within  $P_S^{(N)} \otimes H_B$ , the operators

$$A := \hat{H}_S^{(N)} \otimes I_B, \quad B := H_S^{(N)} \otimes I_B,$$

act as scalars on the support:

$$A|_{\text{supp}(\sigma)} = \varepsilon_* I_{\text{supp}}, \quad B|_{\text{supp}(\sigma)} = \lambda \varepsilon_* I_{\text{supp}}.$$

Therefore, for every  $\tau \in [0, 1]$ ,

$$\sigma^{1-\tau} A \sigma^\tau B = (\varepsilon_*) (\lambda \varepsilon_*) \sigma^{1-\tau} I \sigma^\tau = (\lambda \varepsilon_*^2) \sigma,$$

and hence

$$\text{Tr}(\sigma^{1-\tau} A \sigma^\tau B) = \lambda \varepsilon_*^2 \text{Tr}(\sigma) = \lambda \varepsilon_*^2.$$

On the other hand,

$$\text{Tr}(\sigma A) = \varepsilon_*, \quad \text{Tr}(\sigma B) = \lambda \varepsilon_*,$$

so  $\text{Tr}(\sigma A) \text{Tr}(\sigma B) = \lambda \varepsilon_*^2$ . Plugging these into the definition (98) of  $\text{cov}_\sigma(A, B)$ , each integrand is

$$\lambda \varepsilon_*^2 - \lambda \varepsilon_*^2 = 0,$$

and the  $\tau$ -integral is therefore zero:

$$\lim_{\lambda \rightarrow \infty} \text{cov}_{\omega_\beta(\lambda)}(\hat{H}_S^{(N)}, H_S^{(N)}) = 0. \quad (99)$$

**2.3.2.2 Second term** Let  $H(\lambda) = \lambda \hat{H}_S^{(N)} + V + H_B$ ,  $\omega_\beta(\lambda) = \omega_\beta(H(\lambda))$ ,  $A := H_S^{(0)} - V$  and  $B := \hat{H}_S^{(N)}$ . By definition of the generalized covariance (KuboMori inner product),

$$\text{cov}_{\omega_\beta(\lambda)}(A, B) = \text{Tr}(\omega_\beta(\lambda) A_{H(\lambda), \beta} B) - \text{Tr}(\omega_\beta(\lambda) A) \text{Tr}(\omega_\beta(\lambda) B),$$

with  $A_{H,\beta} := \int_0^1 d\tau e^{\beta\tau H} A e^{-\beta\tau H}$ . Use cyclicity of the trace to rewrite the first term:

$$\text{Tr}(\omega_\beta(\lambda) A_{H(\lambda),\beta} B) = \frac{1}{\mathcal{Z}(\lambda)} \int_0^1 d\tau \text{Tr}(e^{-\beta(1-\tau)H(\lambda)} A e^{-\beta\tau H(\lambda)} B).$$

For any  $\tau > 0$ ,

$$e^{-\tau\beta H(\lambda)} B = e^{-\tau\beta(\lambda\hat{H}_S^{(N)} + V + H_B)} \hat{H}_S^{(N)} \xrightarrow{\lambda \rightarrow \infty} 0,$$

so the integrand vanishes for  $\tau \in (0, 1]$ . Hence,

$$\lim_{\lambda \rightarrow \infty} \text{Tr}[\omega_\beta(\lambda) A_{H(\lambda),\beta} B] = 0.$$

For the product term, we use the large- $\lambda$  limit of the Gibbs state

$$\omega_\beta(\lambda) \xrightarrow{\lambda \rightarrow \infty} \sigma_{SB}^{(N)}, \quad \text{with } \text{supp } \sigma_{SB}^{(N)} \subseteq P_S^{(N)} \otimes H_B,$$

where  $P_S^{(N)}$  is the ground space projector of  $\hat{H}_S^{(N)}$ . On this subspace  $B = \hat{H}_S^{(N)}$  acts as 0, hence  $\text{Tr}(\omega_\beta(\lambda) B) \rightarrow \text{Tr}(\sigma_{SB}^{(N)}) = 0$ . Therefore the product  $\text{Tr}(\omega_\beta(\lambda) A) \text{Tr}(\omega_\beta(\lambda) B) \rightarrow 0$  as well. Combining both pieces gives

$$\lim_{\lambda \rightarrow \infty} \text{cov}_{\omega_\beta(\lambda)}(H_S^{(0)} - V, \hat{H}_S^{(N)}) = 0. \quad (100)$$

Together with the previously established limit  $\text{cov}_{\omega_\beta(\lambda)}(\hat{H}_S^{(N)}, H_S^{(N)}) \rightarrow 0$ , this shows that both contributions in Eq. (96) vanish as  $\lambda \rightarrow \infty$ . It remains to show that, for sufficiently large  $\lambda$ , the derivative of the free energy is nonnegative, thereby completing the justification.

We can expand an expression of the form

$$\text{cov}_{\omega_\beta(\lambda)}(A, \hat{H}_S^{(N)}),$$

in  $y = \exp(-\beta\lambda)$  around  $y = 0$ . The above considerations show that the first order vanishes, hence for very large  $\lambda$  we have

$$\text{cov}_{\omega_\beta(\lambda)}(A, \hat{H}_S^{(N)}) = e^{-\beta\lambda} f(A) + O(e^{-2\beta\lambda}). \quad (101)$$

Since

$$\text{cov}_{\omega_\beta(\lambda)}(\hat{H}_S^{(N)}, H_S^{(N)}) \geq 0, \quad (102)$$

for all  $\lambda$ , we have  $f(\hat{H}_S^{(N)}) > 0$ . We can use these results in Eq. (96) and find

$$\frac{dF(\omega^{(N)}(\lambda), H^{(0)})}{d\lambda} = e^{-\beta\lambda} \left( \lambda f(\hat{H}_S^{(N)}) - f(H_S^{(0)} - V) \right) + O(e^{-2\beta\lambda}). \quad (103)$$

Since  $f(\hat{H}_S^{(N)}) > 0$  for large  $\lambda$  the  $f(\hat{H}_S^{(N)})$  dominates, and thus

$$\frac{d}{d\lambda} \Delta F^{(\text{res})}(\lambda) > 0 \quad \text{for all sufficiently large } \lambda. \quad (104)$$

We hence see that the derivative of  $\Delta F^{(\text{res})}$  becomes positive for large enough  $\lambda$  and  $\Delta F^{(\text{res})}$  approaches its limiting value from below for large  $\lambda$ . Since it is a positive, smooth function, this implies the existence of at least one minimum and Eq. (89) always provides a solution to the problem of maximizing work extraction. Again if several solutions exist, the global optimum must be chosen.

In sum, so far we have reduced the problem of finding the optimal protocol for work extraction to the solution of two matrix equations, (70) and (89). We now provide the solution of these equations, and hence an expression for  $W_{\max}$  in (51), at lowest non-vanishing order in the interaction strength. For that, let us replace  $V$  by  $gV$ , and expand relevant quantities over  $g$ .

## 2.4 Expansion in powers of $g$

In this section we develop a systematic expansion in the interaction strength by replacing  $V \rightarrow gV$  and organizing all quantities in powers of  $g$ . We first solve the stationarity conditions to obtain the optimal contact Hamiltonians ( $X$  for switchon and  $Z$  for switchoff) at leading order in  $g$ . We then evaluate the irreversible and residual corrections.

### 2.4.1 First-Order Solution for $X$ and $Z$

Our aim in this section is to derive explicit equations for the effective Hamiltonians  $X(g)$  and  $Z(g)$  (equivalently, for their first-order coefficients  $X_1$  and  $Z_1$ ).

As discussed previously,  $X_S$  is the system Hamiltonian at the moment the interaction is switched on. It is defined by the condition

$$\rho_S(g) = \text{Tr}_B(\omega_\beta(X_S + gV + H_B)), \quad (105)$$

and is precisely the Hamiltonian that minimizes the irreversible free energy loss  $\Delta F^{(\text{irr})}$ . Similarly,  $Z_S$  is the system Hamiltonian at switch-off, defined through

$$\begin{aligned} & \text{Tr}_B(\omega_\beta(Z_S + gV + H_B)) \text{Tr}(\omega_\beta(Z_S + gV + H_B)(H_{Z,\beta}^{(0)} - (Z_S + gV + H_B))) \\ &= \text{Tr}_B(\omega_\beta(Z_S + gV + H_B)(H_{Z,\beta}^{(0)} - (Z_S + gV + H_B))), \end{aligned} \quad (106)$$

and it minimizes the residual free energy  $\Delta F^{(\text{res})}$ . For notational convenience, we define the full Hamiltonians corresponding to the coupling and decoupling stages as

$$X := X_S + gV + H_B \quad , \quad Z := Z_S + gV + H_B. \quad (107)$$

In the strong-coupling regime, because the bath contribution  $gV$  shifts the spectrum of  $S$  and entangles system and bath, the familiar Gibbs form  $\omega_\beta(H_S)$  no longer captures the reduced equilibrium state of  $S$ . A careful protocol must therefore include local Hamiltonians that compensate these shifts and minimize dissipative losses. The optimization problems at the beginning and at the end of the stroke are not identical; Assigning separate symbols  $X_S$  and  $Z_S$  makes this distinction explicit and prevents confusion with the fixed, bare Hamiltonian  $H_S$ .

$X_S(g)$  is the Hamiltonian that cancels the bath's shift at the moment the coupling  $gV$  is switched on;  $Z_S(g)$  performs the analogous role at switch-off. Both operators depend analytically on  $g$  and reduce to the familiar weak-coupling limits when  $g = 0$ . In this section, we derive their first-order terms and show that, after these terms are included, all linear-in- $g$  work and entropy costs disappear, leaving a quadratic correction proportional to the covariance of  $V - \text{Tr}_B[V \omega_\beta(H_B)]$ .

To capture the effect of the coupling, we now expand  $X_S$  and  $Z_S$  in powers of  $g$ . This allows us to determine how the system Hamiltonians deviate from their zeroth order values.

$$X_S = X_S(0) + gX'_S(0) + O(g^2) \quad ; \quad X'_S(0) = \left. \frac{dX_S}{dg} \right|_{g=0} \quad (115)$$

$$Z_S = Z_S(0) + gZ'_S(0) + O(g^2) \quad ; \quad Z'_S(0) = \left. \frac{dZ_S}{dg} \right|_{g=0} \quad (116)$$

Substituting these expansions into the definitions of  $X$  and  $Z$  yields

$$X = \overbrace{X_S(0) + H_B}^{=X_0} + g \overbrace{(V + X'_S(0))}^{=X_1} + O(g^2) = X_0 + gX_1 + O(g^2), \quad (117)$$

$$Z = \underbrace{Z_S(0) + H_B}_{=Z_0} + g \underbrace{(V + Z'_S(0))}_{=Z_1} + O(g^2) = Z_0 + gZ_1 + O(g^2), \quad (118)$$

$$\omega_\beta(X) = \omega_0^{(X)} + g\omega_1^{(X)} + g^2\omega_2^{(X)} + O(g^3) \quad ; \quad \left. \frac{d^i \omega(X)}{dg^i} \right|_{g=0} = \omega_i^{(X)}, \quad (119)$$

$$\omega_\beta(Z) = \omega_0^{(Z)} + g\omega_1^{(Z)} + g^2\omega_2^{(Z)} + O(g^3) \quad ; \quad \left. \frac{d^i \omega(Z)}{dg^i} \right|_{g=0} = \omega_i^{(Z)}, \quad (120)$$

where  $\omega_0^{(Z)} = \omega_\beta(H^{(0)})$  and  $\omega_0^{(X)} = \omega_\beta(\tilde{H}^{(0)})$ .

Using Eq. (119), we have

$$\text{Tr}_B(\omega_\beta(X)) = \text{Tr}_B(\omega_0^{(X)}) + g \text{Tr}_B(\omega_1^{(X)}) + O(g^2). \quad (121)$$

Initially, the interaction is off, so the system's initial state  $\rho_S$  cannot depend on  $g$ . This means

$$\implies \text{Tr}_B(\omega_1^{(X)}) = \text{Tr}_B \left( \left. \frac{d\omega_\beta(X(g))}{dg} \right|_{g=0} \right) = 0. \quad (122)$$

To evaluate the derivative above, we expand  $\omega_\beta(X(g))$ ,

$$\begin{aligned} \frac{d\omega_\beta(X(g))}{dg} &= \frac{1}{\mathcal{Z}} \frac{d}{dg} e^{-\beta X(g)} - \frac{1}{\mathcal{Z}^2} \frac{d\mathcal{Z}}{dg} e^{-\beta X(g)}, \\ &= \frac{-\beta}{\mathcal{Z}} \int_0^1 ds e^{-\beta s X(g)} X'(g) e^{-\beta(1-s)X(g)} - \frac{e^{-\beta X(g)}}{\mathcal{Z}} (-\beta \text{Tr}(X_1 \omega_\beta(X(g)))) . \end{aligned}$$

To simplify notation, set

$$\int_0^1 ds e^{-\beta s X(g)} X'(g) \underbrace{\frac{e^{-\beta X(g)}}{\mathcal{Z}} e^{+\beta s X(g)}}_{=\omega_0^{(X)}} \Big|_{g=0} := -\beta \left( X_1 \omega_0^{(X)} \right)_{X_0} .$$

Finally, combining this with Eq. (122) yields

$$-\beta \left( X_1 \omega_0^{(X)} \right)_{X_0} + \beta \omega_0^{(X)} \text{Tr} \left( X_1 \omega_0^{(X)} \right) = 0 \rightarrow \left( X_1 \omega_0^{(X)} \right)_{X_0} = \omega_0^{(X)} \text{Tr} \left( X_1 \omega_0^{(X)} \right) . \quad (123)$$

Before the interaction is switched on, system and bath are uncorrelated. So,

$$\omega_0^{(X)} = \omega_\beta(\tilde{H}^{(0)}) = \omega_\beta(\tilde{H}_S^{(0)}) \otimes \omega_\beta(H_B) .$$

Using this, Eq. (123) becomes

$$\text{Tr} \left( X_1 \omega_0^{(X)} \right) \text{Tr}_B \left( \omega_0^{(X)} \right) = \text{Tr}_B \left( \left( X_1 \omega_0^{(X)} \right)_{X_0} \right) \implies \text{Tr} \left( X_1 \omega_0^{(X)} \right) \mathbb{1}_S = \text{Tr}_B \left( X_1 \omega_\beta(H_B) \right)_{X_S^{(0)}} . \quad (124)$$

Similar to switch-on process, the decoupling of the bath and the system is assumed to occur instantaneously. Consequently, the final state  $\rho_S^{(N)}$  after decoupling coincides with the state of the system immediately before the interaction is switched off, and therefore it should not depend on  $g$ . Applying the same steps as in the derivation leading up to Eq. (124) for  $\rho_S^{(N)} = \text{Tr}_B(\omega_\beta(H^{(N)}))$ , we obtain

$$\text{Tr} \left( Z_1 \omega_0^{(Z)} \right) \mathbb{1}_S = \text{Tr}_B \left( Z_1 \omega_\beta(H_B) \right)_{Z_S^{(0)}} . \quad (125)$$

We know that  $X_1 = X'_S(0) + V$  and define  $V' = -\text{Tr}(V \omega_\beta(H_B))$ . We now substitute these relations.

$$\text{Tr} \left( (X'_S(0) + V) \omega_0^{(X)} \right) \mathbb{1}_S = \text{Tr}_B \left( (X'_S(0) + V) \omega_\beta(H_B) \right)_{X_S^{(0)}}$$

$$\begin{aligned}
\text{RHS} &= \text{Tr}_B (X'_S(0) \omega_\beta(H_B))_{X_S^{(0)}} + \text{Tr}_B (V \omega_\beta^{(H_B)})_{X_S^{(0)}} = X'_S(0) + \text{Tr}_B (V \omega_\beta(H_B)) \\
&= (X'_S(0))_{X_S^{(0)}} - (V')_{X_S^{(0)}} \\
\text{LHS} &= \text{Tr}_S \left( \text{Tr}_B (X'_S(0) \omega_0^{(X)}) \right) + \text{Tr}_S \left( \text{Tr}_B (V \omega_0^{(X)}) \right) = \text{Tr}_S (X'_S(0) \omega_\beta(X_S(0))) \\
&\quad - \text{Tr}_S (V' \omega_\beta(X_S(0))) \\
&\implies \text{Tr}_S (X'_S(0) \omega_\beta(X_S(0))) \mathbf{1}_S - (X'_S)_{X_S^{(0)}} = \text{Tr}_S (V' \omega_\beta(X_S(0))) - (V')_{X_S^{(0)}} \tag{126}
\end{aligned}$$

The relations obtained so far can be reorganized as linear equations for  $X_1$  and  $Z_1$ . To solve them compactly, and thus close the derivation of the equations for  $X$  and  $Z$ , we introduce the super-operators  $L$  and  $R$ .

$$L(\cdot) := \text{Tr}_S(\omega_\beta(X_S(0)) \cdot) \mathbf{1}_S - (\cdot)_{X_S^{(0)}} \tag{127}$$

An analogous super-operator for the final step is

$$R(\cdot) := \text{Tr}_S(\omega_\beta(Z_S(0)) \cdot) \mathbf{1}_S - (\cdot)_{Z_S^{(0)}}. \tag{128}$$

We now show that Both  $L$  and  $R$  are linear. For example,

$$\begin{aligned}
L(\alpha A + \beta B) &= \text{Tr}(\omega_\beta(X_S(0))(\alpha A + \beta B)) \mathbf{1}_S - (\alpha A + \beta B)_{X_S(0)} \\
&= \alpha L(A) + \beta L(B),
\end{aligned}$$

and the same argument applies to  $R$ . Using this, Eqs. (126) and (125) become

$$L(X'_S(0)) = L(V'), \tag{129}$$

$$R(Z'_S(0)) = R(V'). \tag{130}$$

Assuming  $L(M) = 0$ , where  $M$  denotes the null operator on  $L(\cdot)$ , and using linearity of  $L$ , we obtain

$$\begin{aligned}
&\xrightarrow{(129)} X'_S(0) + M = V' = -\text{Tr}(V \omega_\beta(H_B)) + O(g^2), \\
&\xrightarrow{(115)} X_S = \tilde{H}_S - g(\text{Tr}_B(V \omega_\beta(H_B)) - M) + O(g^2). \tag{131}
\end{aligned}$$

Similarly for  $M'$  and  $R(\cdot)$ ,

$$\begin{aligned}
&\xrightarrow{(130)} Z'_S(0) = V' + M' = -\text{Tr}(V \omega_\beta(H_B)) + M', \\
&\xrightarrow{(116)} Z_S = H_S - g(\text{Tr}_B(V \omega_\beta(H_B)) - M') + O(g^2). \tag{132}
\end{aligned}$$

Since the operator  $M$  does not affect the extracted work we may set  $M = 0$ . Recalling that  $H_S^{(1)} = X_S$



and  $H_S^{(N)} = Z_S$ , we finally obtain the expressions for the system Hamiltonians:

$$\boxed{H_S^{(1)} = \tilde{H}_S - g \operatorname{Tr}_B(\mathbf{1}_S \otimes \omega_\beta(H_B) V) + O(g^2)}, \quad (133)$$

$$\boxed{H_S^{(N)} = H_S - g \operatorname{Tr}_B(\mathbf{1}_S \otimes \omega_\beta(H_B) V) + O(g^2)}. \quad (134)$$

### 2.4.2 Free-Energy Corrections

In this section we want to compute how expansions (117) to (120) enter the irreversible and residual free energies. For this, we use the following relation

$$F(\rho, H) - F(\omega_\beta(H), H) = T S(\rho \| \omega_\beta(H)), \quad (135)$$

with the relative entropy defined as,

$$S(\rho \| \sigma) := \operatorname{Tr}(\rho (\log \rho - \log \sigma)).$$

Therefore,

$$\Delta F_{\min}^{(\text{irr})} = T S(\omega_\beta(\tilde{H}^{(0)}) \| \omega_\beta(X)) \implies \frac{d}{dg} \Delta F_{\min}^{(\text{irr})} = T \frac{d}{dg} S(\omega_\beta(\tilde{H}^{(0)}) \| \omega_\beta(X)), \quad (136)$$

$$\Delta F_{\min}^{(\text{res})} = T S(\omega_\beta(Z) \| \omega_\beta(H^{(0)})) \implies \frac{d}{dg} \Delta F_{\min}^{(\text{res})} = T \frac{d}{dg} S(\omega_\beta(Z) \| \omega_\beta(H^{(0)})). \quad (137)$$

Since the two derivatives are handled in the same manner, we compute only one of them and later apply the results to the other. To evaluate these free-energy corrections, we need the derivative of the relative entropy.

For clarity, we introduce a dummy parameter  $t$ , perform the calculation in terms of  $t$ , and later substitute  $t \rightarrow g$ . With  $\omega_t := \omega_\beta(H_t)$ ,

$$\begin{aligned} S(\omega_0 \| \omega_t)' &= \frac{d}{dt} \operatorname{Tr}(\omega_0 (\log(\omega_0) - \log(\omega_t))) = -\frac{d}{dt} \operatorname{Tr}(\omega_0 \log(\omega_t)) = -\operatorname{Tr}\left(\omega_0 \frac{d}{dt} \log(\omega_t)\right), \\ \log(\omega_t) &= -\beta H_t - \log(\mathcal{Z}(t)) \implies S(\omega_0 \| \omega_t)' = -\operatorname{Tr}\left(\omega_0 \left(-\beta H_t' - \frac{\mathcal{Z}'(t)}{\mathcal{Z}(t)}\right)\right), \\ \implies S(\omega_0 \| \omega_t)' &= \beta \operatorname{Tr}(\omega_0 H_t') + \underbrace{\log(\mathcal{Z}(t))'}_1 \operatorname{Tr}(\omega_0). \end{aligned} \quad (138)$$

Using Duhamel formula one finds,

$$\left. \frac{d}{dt} e^{-\beta H_t} \right|_{t=0} = -\beta \int_0^1 ds e^{-\beta s H_0} H'_0 e^{-\beta(1-s)H_0} \quad ; \quad H'_0 = \left. \frac{dH}{dt} \right|_{t=0}, \quad (139)$$

$$\begin{aligned} \implies \left. \frac{d}{dt} \mathcal{Z} \right|_{t=0} &= -\beta \int_0^1 ds \text{Tr} \left( e^{-\beta s H_0} H'_0 e^{-\beta(1-s)H_0} \right), \\ \implies \left. \frac{d}{dt} \mathcal{Z} \right|_{t=0} &= -\beta \text{Tr} \left( H'_0 e^{-\beta H_0} \right) \implies \log(\mathcal{Z}(t))' = -\beta \text{Tr} (H'_t \omega_t), \end{aligned} \quad (140)$$

$$\xrightarrow{(139), (140)} S(\omega_0 || \omega_t)' = \beta \text{Tr}((\omega_0 - \omega_t) H'_t). \quad (141)$$

By substituting (141) we get

$$\frac{d}{dg} \Delta F_{\min}^{(\text{irr})} = \text{Tr} \left( (\omega_\beta(\tilde{H}^{(0)}) - \omega_\beta(X(g))) X_1 \right) \quad (142)$$

Using Eq. (119), we have

$$\frac{d}{dg} \Delta F_{\min}^{(\text{irr})} = \text{Tr} \left( (\omega_0^{(X)} - (\omega_0^{(X)} + g \omega_1^{(X)})) X_1 \right) = -g \text{Tr}(\omega_1^{(X)} X_1) + O(g^2), \quad (143)$$

$$\left. \frac{d^2}{dg^2} \Delta F_{\min}^{(\text{irr})} \right|_{g=0} = -\text{Tr}(X_1 \omega_1^{(X)}). \quad (144)$$

By differentiating  $S(\omega_\beta(Z) || \omega_\beta(H^{(0)}))$  with respect to  $g$ , and using expansion (120) together with the same steps as the irreversible case, we obtain

$$\frac{d\Delta F_{\min}^{(\text{res})}}{dg} = -g \text{Tr} \left( Z_1 \omega_1^{(Z)} \right) + O(g^2), \quad (145)$$

$$\left. \frac{d^2 \Delta F_{\min}^{(\text{res})}}{dg^2} \right|_{g=0} = -\text{Tr} \left( Z_1 \omega_1^{(Z)} \right). \quad (146)$$

Applying the generalized covariance identity (Eq. (79)) to Eqs. (143) and (145), we obtain

$$\text{Tr} \left( X_1 \frac{d\omega_\beta(X(g))}{dg} \right) = -\beta \text{cov}_{\omega_\beta(\tilde{H}^{(0)})}(X_1, X_1), \quad (147)$$

$$\text{Tr} \left( Z_1 \frac{d\omega_\beta(Z(g))}{dg} \right) = -\beta \text{cov}_{\omega_\beta(H^{(0)})}(Z_1, Z_1). \quad (148)$$

Now we substitute Eqs. (144) and (146) in the expansions of  $F_{\min}^{(\text{irr})}$  and  $F_{\min}^{(\text{res})}$ ,

$$\Delta F_{\min}^{(\text{irr})} = \frac{\beta g^2}{2} \text{cov}_{\omega_\beta(\tilde{H}^{(0)})}(V - \text{Tr}_B[V \omega_\beta(H_B)], V - \text{Tr}_B[V \omega_\beta(H_B)]) + O(g^3), \quad (149)$$

$$\Delta F_{\min}^{(\text{res})} = \frac{\beta g^2}{2} \text{cov}_{\omega_\beta(H^{(0)})}(V - \text{Tr}_B[V \omega_\beta(H_B)], V - \text{Tr}_B[V \omega_\beta(H_B)]) + O(g^3). \quad (150)$$

**Inconsistency: Nonzero Linear Contributions to  $\Delta F$ .** The equations above reproduce the papers stated conditions. However, a separate issue arises when we insert the perturbative solutions for  $X$  and  $Z$  into the freeenergy expansions. Writing, schematically,

$$\Delta F^{(\text{irr})}(g) = \Delta F_0^{(\text{irr})} + g \Delta F_1^{(\text{irr})} + g^2 \Delta F_2^{(\text{irr})} + \dots, \quad (151)$$

$$\Delta F^{(\text{res})}(g) = \Delta F_0^{(\text{res})} + g \Delta F_1^{(\text{res})} + g^2 \Delta F_2^{(\text{res})} + \dots, \quad (152)$$

our substitutions yield nonzero linear terms, i.e.,  $\Delta F_1^{(\text{irr})} \neq 0$  and/or  $\Delta F_1^{(\text{res})} \neq 0$  in general. This appears to contradict the claim that both costs are purely  $O(g^2)$ , as asserted in Eqs. (149) and (150). In other words, while the paper states that the leading corrections are quadratic in  $g$ , the explicit first-order contributions do not cancel without further conditions. Therefore, to reproduce Eqs. (149)-(150) one would need additional assumptions (not stated in the main text or the Supplement) that eliminate the  $O(g)$  terms, e.g., constraints that center the interaction or impose symmetries relations that force the linear coefficient to vanish. To keep pace with the papers presentation, we proceed by adopting the printed result and retaining only the  $O(g^2)$  terms in the free energies, while noting that a clarification of the missing assumptions is required for full consistency.

Next, we turn to the remarks made in the paper, expanding and clarifying their implications.

(I) Optimizing the local Hamiltonians wipes out every linear-in- $g$  term, so the leading strong coupling cost appears at order  $g^2$ . In other words, The expansion of  $\Delta F_{\min}^{(\text{irr})/(\text{res})}$  at order  $O(g^2)$  depends only on  $H^{(1)/(N)}$  at order  $O(g)$ .

(II) We found the work to be  $W(g) = W_{\text{weak}} - \Delta F_{\min}^{(\text{irr})}(g) - \Delta F_{\min}^{(\text{res})}(g)$ . The first order correction to  $W$  vanishes for any  $H_S^{(1)} = \tilde{H}_S + O(g)$ ,  $H_S^{(N)} = H_S + O(g)$ . This follows from the fact that the penalty terms  $\Delta F_{\min}^{(\text{irr})}(g)$  are differentiable function of  $g$  and have a minimum at  $g = 0$ .

(III) In the first-order step we subtract the baths average influence, writing,

$$X_1 = V - \text{Tr}_B[V \omega_\beta(H_B)] + M,$$

and similarly for  $Z_1$ . This removes the constant term the bath would otherwise add, so every energy shift that scales like  $g$  is canceled and no work is lost at linear order. The first non-zero penalty therefore exists at  $g^2$ . After this term is stripped away, the interaction we keep,  $V - \text{Tr}_B[V \omega_\beta(H_B)]$ , has zero average over the bath's equilibrium state. In other words, if you sample the bath many times, the push it exerts on the system through this centered interaction cancels out on average. What matters now are the fluctuations around that zero mean, and the strength of those kicks is measured by the covariance  $\text{cov}_{\omega_\beta(H^{(0)})}(V - \text{Tr}_B[V \omega_\beta(H_B)], V - \text{Tr}_B[V \omega_\beta(H_B)])$ . This covariance tells us how widely the interaction energy is spread when the bath is in equilibrium.

(IV) The generalized covariance  $\text{cov}_{\omega_\beta(H)}(A, B)$  shows how much an observable changes when

you make a tiny tweak to the Hamiltonian.

In summary, in Sec. 2 we showed that the leading corrections to the maximum extractable work at strong coupling appear at order  $g^2$ , and are fully determined by the thermal fluctuations of the interaction after subtracting its bath average. Moreover, we proved that in the limit of infinitely strong coupling ( $g \rightarrow \infty$ ), both the irreversible and residual free-energy penalties diverge. This means that if the local Hamiltonians are kept fixed, strong coupling to the bath can completely destroy the possibility of extracting work. Thus, optimal work extraction in the strong-coupling regime not only requires controlling the interaction strength but also carefully adapting the local system Hamiltonians to cancel shifting effects and suppress unwanted energy fluctuations.

### 3 Heat dissipation

We now turn to the heat dissipation analysis, following the explanations provided in the Supplementary Material<sup>3</sup>. In thermodynamics, it is often the case that the optimal protocols for one task turns out to be also optimal for others. Here we apply this logic to show that the previous results can be readily applied in order to minimize heat dissipation and maximize efficiency of heat engines strongly coupled to baths. The goal of this section is to relate the total heat dissipated into the bath during the protocol to the change in entropy. To do so we consider an initially non-interacting Hamiltonian,  $H^{(0)} = H_S^{(0)} + H_B$ , an initial state  $\rho^{(0)} = \rho_S \otimes \omega_B(H_B)$  and the final state  $H^{(N)}$  (with  $H_S^{(N)}$  fixed and independent of  $N$ ). Based on the Supplementary Material, the protocol is broken into three steps, each with a specific role. The three steps are:

1. **A series of quenches are applied to the local Hamiltonian of  $S$ ,  $H_S^{(0)} \rightarrow \dots \rightarrow H_S^{(1)}$ . After that, the interaction between  $S$  and  $B$  is turned on.**

Here

$$H^{(0)} = H_S^{(0)} + H_B \quad , \quad H^{(1)} = H_S^{(1)} + V + H_B.$$

Because the state is unaffected by an instantaneous quench, the work invested (or extracted) during the whole sequence is simply the difference of energies evaluated in the unchanged state  $\rho^{(0)}$ ,

$$\Delta U_1 = \text{Tr}(\rho^{(0)} (H^{(1)} - H^{(0)})) = W_1 \implies W_1 = \text{Tr}((H_S^{(0)} - H_S^{(1)} - V) \rho^{(0)}). \quad (153)$$

2. **While the system  $S$  stays in thermal contact with the bath  $B$ , we gradually tune its Hamiltonian until it reaches  $H_S^{(N)}$ ; the complete Hamiltonian then becomes  $H^{(N)} = H_S^{(N)} + V + H_B$ . This amounts to a series of quenches  $H_S^{(1)} \rightarrow \dots \rightarrow H_S^{(N)}$ , each followed by an equilibrium to the corresponding thermal state.**

---

<sup>3</sup>See, Supplementary Material of Perarnau-Llobet et al. for details

We know that in the limit  $N \rightarrow \infty$  this setup reproduces an isothermal, quasi-static path.

$$\begin{aligned}
W_2 &= \sum_{i=1}^{N-1} \text{Tr}(\rho_S^{(i)} (H_S^{(i)} - H_S^{(i+1)})) \quad , \quad \rho_S^{(i)} := \text{Tr}_B \omega^{(i)} \\
\implies W_2 &= F(\omega^{(1)}, H^{(1)}) - F(\omega^{(N)}, H^{(N)}) - T \sum_{i=1}^{N-1} S(\omega^{(i)} \parallel \omega^{(i+1)}) \quad (154)
\end{aligned}$$

As we showed in the last report, the relative entropy goes to zero as  $N \rightarrow \infty$  (Eq. (34)),

$$\implies W_2^{\text{isoth}} = F(\omega^{(1)}, H^{(1)}) - F(\omega^{(N)}, H^{(N)}). \quad (155)$$

Using the first law of thermodynamics we have

$$\Delta U_2 = Q_2 - W_2 \implies Q_2 = \Delta U_2 + W_2, \quad (156)$$

$$\Delta U_2 = \text{Tr}(H^{(N)} \omega^{(N)}) - \text{Tr}(H^{(1)} \omega^{(1)}). \quad (157)$$

Substituting  $W_2$  and Using  $F = U - TS$  gives

$$\begin{aligned}
Q_2 &= \text{Tr}(H^{(N)} \omega^{(N)}) - \text{Tr}(H^{(1)} \omega^{(1)}) + \text{Tr}(H^{(1)} \omega^{(1)}) - TS(\omega^{(1)}) \\
&\quad - \text{Tr}(H^{(N)} \omega^{(N)}) + TS(\omega^{(N)}) - T \sum_{i=1}^{N-1} S(\omega^{(i)} \parallel \omega^{(i+1)}) \quad (158)
\end{aligned}$$

$$= T\Delta S_2 - T \sum_{i=1}^{N-1} S(\omega^{(i)} \parallel \omega^{(i+1)}), \quad (159)$$

where  $\Delta S_2 = S(\omega^{(N)}) - S(\omega^{(1)})$ . As the number of steps  $N$  grows, The second term in (159) goes to zero and the Clausius heat associated with the net entropy increase,  $T\Delta S_2$ , will be the only term left.

3. **The interaction between  $S$  and  $B$  is turned off, so that the final Hamiltonian is  $H^{(N+1)} = H_S^{(N)} + H_B$ .**

The work extracted from the system is

$$W_3 = \text{Tr}\left(\left(H^{(N)} - H^{(N+1)}\right) \omega^{(N)}\right) = \text{Tr}\left(V \omega^{(N)}\right). \quad (160)$$

This protocol is essentially the same as the optimal protocol for work extraction but without the last step, which ensured cyclicity.

The final Hamiltonian  $H_S^{(N)}$  is fixed, and hence so is the final state of  $S$ ,

$$\rho_S^{(N)} = \text{Tr}_B\left(\omega_\beta\left(H^{(N)}\right)\right). \quad (161)$$

As a consequence, the entropy change is fixed (since the initial state is also fixed),

$$\Delta S = S(\rho_S^{(N)}) - S(\rho_S^{(0)}). \quad (162)$$

In order to define the average heat, we use the first law of thermodynamics,

$$Q = \Delta E_S + W \quad \text{for the system,} \quad (163)$$

$$Q = -\Delta E_B \quad \text{for the bath.} \quad (164)$$

Note that in Eq. (163),  $W$  denotes the work performed on the system, which explains the positive sign in the expression.

At the beginning and the end of the process, the system  $S$  and the bath  $B$  are assumed to be non-interacting, meaning there is no coupling term  $V$  between them. This simplifies the analysis because in this scenario, the energy change is solely due to the systems internal energy and the bath's internal energy.  $Q$  represents the total heat exchanged with the bath. Since the system is interacting with the bath and the energy of the system changes, the energy lost by the system will be absorbed by the bath.  $-Q$  is the dissipated heat. It is the energy that flows from the system to the bath.

Since the  $\Delta E_S$  is fixed (because the initial and the final Hamiltonians of the system are fixed), it naturally follows that minimizing heat dissipation corresponds to maximizing extractable work. Therefore based on the protocol found for the maximum work extraction, the second step of the heat dissipation protocol corresponds to an isothermal process. The total work of steps 1-3 of the protocol is

$$\begin{aligned} W &= W_1 + W_2 + W_3 = \text{Tr}((H_S^{(0)} - H_S^{(1)} - V)\rho^{(0)}) + F(\omega^{(1)}, H^{(1)}) - F(\omega^{(N)}, H^{(N)}) + \text{Tr}(V\omega^{(N)}) \\ \text{Tr}[(H_S^{(0)} - H_S^{(1)} - V)\rho^{(0)}] &= \text{Tr}(H_S^{(0)}\rho^{(0)}) - \text{Tr}((H_S^{(1)} + V)\rho^{(0)}) + \text{Tr}(H_B\rho^{(0)}) - \text{Tr}(H_B\rho^{(0)}) \\ &\quad + TS(\rho^{(0)}) - TS(\rho^{(0)}) = F(\rho^{(0)}, H^{(0)}) - F(\rho^{(0)}, H^{(1)}) \\ \implies \text{Tr}((H_S^{(0)} - H_S^{(1)} - V)\rho^{(0)}) &= F(\rho^{(0)}, H^{(0)}) - F(\rho^{(0)}, H^{(1)}) \end{aligned} \quad (165)$$

$$\begin{aligned} \implies W &= F(\rho^{(0)}, H^{(0)}) - \underbrace{F(\rho^{(0)}, H^{(1)}) + F(\omega^{(1)}, H^{(1)}) - F(\omega^{(N)}, H^{(N)})}_{-\Delta F^{(\text{irr})}} + \text{Tr}(V\omega^{(N)}) \\ \implies W &= F(\rho^{(0)}, H^{(0)}) - F(\omega^{(N)}, H^{(N)}) + \text{Tr}(V\omega^{(N)}) - \Delta F^{(\text{irr})} \end{aligned} \quad (166)$$

Note that,

$$\begin{aligned} F(\omega^{(N)}, H^{(N)}) + \text{Tr}(V\omega^{(N)}) &= \text{Tr}((H_S^{(N)} + H_B + V)\omega^{(N)}) - TS(\omega^{(N)}) + \text{Tr}(V\omega^{(N)}) \\ &= F(\omega^{(N)}, H^{(0)}) - \text{Tr}((H_S - H_S^{(N)})\omega^{(N)}) + 2\text{Tr}(V\omega^{(N)}), \\ \implies W &= F(\rho^{(0)}, H^{(0)}) - F(\omega^{(N)}, H^{(0)}) - \Delta F^{(\text{irr})} + \text{Tr}((H_S - H_S^{(N)})\omega^{(N)}). \end{aligned} \quad (167)$$

**Implicit boundary condition behind Eq. (168).** The Supplementary Material states

$$W = F(\rho^{(0)}, H^{(0)}) - F(\omega^{(N)}, H^{(0)}) - \Delta F^{(\text{irr})}. \quad (168)$$

This formula is valid only if one imposes the boundary condition  $H_S^{(N)} = H_S$ , i.e., the bare system Hamiltonian is the same at the beginning and at the end of the protocol. At first glance this seems at odds with the papers emphasis on a non-cyclic protocol; in fact, there is no logical conflict: the process can be non-cyclic for the composite  $S+B$  (because the interaction is switched on/off and correlations are created/erased) while still closing the local loop for  $S$  by taking  $H_S^{(N)} = H_S$ . The problem is that this boundary condition is not stated explicitly in either the main text or the Supplement. Consequently, our calculations for the more general case  $H_S^{(N)} \neq H_S$  do not reproduce Eq. (168). To remain consistent with the papers derivations, we therefore adopt  $H_S^{(N)} = H_S$  in what follows and proceed using Eq. (168).

Using the first law of thermodynamics, and recalling that  $\rho_0 = \rho_S \otimes \omega_B(H_B)$ ,  $\rho_S^{(N)} = \text{Tr}_B(\omega^{(N)})$  and  $\rho_B^{(N)} = \text{Tr}_S(\omega^{(N)})$ , we obtain

$$\begin{aligned} \Delta E_S &= \text{Tr}(H_S^{(N)} \rho_S^{(N)}) - \text{Tr}(H_S \rho_S^{(0)}) = \text{Tr}(H_S^{(N)} \omega^{(N)}) - \text{Tr}(H_S \rho^{(0)}), \\ \implies Q &= \text{Tr}(H_S^{(N)} \omega^{(N)}) - \text{Tr}(H_S \rho^{(0)}) + W \\ &= \text{Tr}(H_S \omega^{(N)}) - \text{Tr}(H_S^{(0)} \rho^{(0)}) + \text{Tr}((H_S + H_B) \rho^{(0)}) - TS(\rho^{(0)}) - \text{Tr}((H_S + H_B) \omega^{(N)}) \\ &\quad + TS(\omega^{(N)}) - \Delta F^{(\text{irr})} \\ &= \text{Tr}(H_B \omega_B(H_B)) - \text{Tr}(H_B \rho_B^{(N)}) - TS(\rho_S^{(0)} \otimes \omega_B(H_B)) + TS(\omega^{(N)}) - \Delta F^{(\text{irr})}. \end{aligned} \quad (169)$$

Adding and subtracting  $TS(\rho_B^{(N)})$ ,

$$\begin{aligned} Q &= \text{Tr}(H_B \omega_B(H_B)) - TS(\omega_B(H_B)) - \text{Tr}(H_B \rho_B^{(N)}) + TS(\rho_B^{(N)}) \\ &\quad + T(S(\omega^{(N)}) - S(\rho_B^{(N)}) - S(\rho_S^{(0)})) - \Delta F^{(\text{irr})} \\ &= F(\omega_B(H_B), H_B) - F(\rho_B^{(N)}, H_B) + T(S(\omega^{(N)}) - S(\rho_B^{(N)}) - S(\rho_S^{(0)})) - \Delta F^{(\text{irr})} \end{aligned} \quad (170)$$

We now make use of mutual information to write,

$$S(\omega^{(N)}) = S(\rho_S^{(N)}) + S(\rho_B^{(N)}) - I(\omega^{(N)}; S : B),$$

and using  $F(\rho, H) - F(\omega_B(H), H) = TS(\rho || \omega_B(H))$ , we finally obtain

$$Q = T(S(\rho_S^{(N)}) - S(\rho_S^{(0)})) - T(S(\rho_B^{(N)} || \omega_B(H_B)) + I(\omega^{(N)}; S : B)) - \Delta F^{(\text{irr})}. \quad (171)$$

Using Eq. (162) and  $\Delta F_B^{(res)} = T \left( S(\rho_B^{(N)} \| \omega_\beta(H_B)) \right)$ , this becomes

$$Q = T\Delta S - \left( \Delta F_B^{(res)} + TI(\omega^{(N)}; S : B) + \Delta F^{(irr)} \right). \quad (172)$$

As expected, the second law is satisfied in the form  $T\Delta S \geq Q$ . The protocol minimizing heat dissipation  $-Q$  can be found by minimizing the negative terms in Eq. (172). Note however that the terms  $\Delta F_B^{(res)}$  and  $I(\omega^{(N)}; S : B)$  are fixed through  $H_S^{(N)}$ . Therefore, the only task left is to minimize  $\Delta F^{(irr)}$  over the Hamiltonian  $H_S^{(1)}$ , a problem that was already solved in the previous sections. Summarizing, the protocol minimizing heat dissipation can be described as:  $S$  and  $B$  are put in contact through the choice (105), which ensures minimal heat dissipation. Afterwards an isothermal transformation is implemented until the desired Hamiltonian  $H_S^{(N)}$  is reached.

As we did for work extraction, we now replace  $V$  by  $gV$  and find the first non-vanishing corrections of the penalizing terms in (172). First of all, to understand the scaling of the term involving the mutual information, let us write the mutual information as

$$I(\omega^{(N)}; S : B) = S \left( \omega^{(N)} \| \text{Tr}_B(\omega^{(N)}) \otimes \text{Tr}_S(\omega^{(N)}) \right) \geq 0. \quad (173)$$

Since we have  $\omega^{(N)} = \omega_S^{(N)} \otimes \omega_B^{(N)}$  if and only if  $g = 0$ , the function obtains its minimal value 0 only at  $g = 0$ . Also using (141), we know that the first derivative of  $I(g)$  at  $g = 0$  is equal to zero. Therefore its Taylor expansion starts with the quadratic term,  $g^2$ . As argued in the previous sections, the corrections of the relative distance  $S(\cdot \| \cdot)$ , (141), are also of  $O(g^2)$ . Hence, without doing any explicit calculations, we can conclude that,

$$Q = T\Delta S - K_q g^2 - O(g^3), \quad (174)$$

where  $K_q$  is given by

$$K_q = \frac{1}{2} \frac{d^2}{dg^2} \left[ \Delta F_{\min}^{(irr)} + T \left( S(\rho_B^{(N)} \| \omega_\beta(H_B)) + I(\omega^{(N)}; S : B) \right) \right] \Big|_{g=0}. \quad (175)$$

## 4 Heat engine

We consider engines made up of two baths at different temperatures which can interact strongly with  $S$ . The task is then to maximize the efficiency of a cycle of the engine. The optimal cycle turns out to have the same form as a Carnot engine (this will later be proved). Let us optimize a Carnot engine in the strong coupling regime, by considering the presence of two baths,  $B_c$  and  $B_h$ , at temperatures  $T_c, T_h > 0$ , respectively. We consider a Carnot cycle consisting of four steps:

### 1. Coupling to $B_h$ + isothermal process + decoupling from $B_h$ .

At first the global Hamiltonian is  $H_h^{(0)} = H_S^{(A)} + H_{B_h}$ ; the two systems are uncorrelated. We switch on the interaction  $V_h$ , and the global Hamiltonian becomes  $H^{(A)} = H_S^{(A)} + V + H_{B_h}$ . We perform the isothermal transformation and then keep the system in contact with  $B_h$



long enough for it to relax to the Gibbs state  $\omega_{\beta_h}(H^{(B)})$  of the interacting Hamiltonian  $H^{(B)} = H_S^{(B)} + V_h + H_{B_h}$ . Then we turn the interaction off again. The reduced state of  $S$  after decoupling is therefore,

$$\rho_1 = \text{Tr}_{B_h} \left( \omega_{\beta_h}(H^{(B)}) \right). \quad (176)$$

During this reversible isothermal stroke the heat absorbed from the bath and the extracted work follow directly from (172) and (168),

$$Q_h = T_h \Delta S_h - T_h \left( S(\rho_{B_h}^{(B)} || \omega_{\beta_h}(H_{B_h})) + I(\omega_{\beta_h}(H^{(B)}) ; S : B_h) \right) - \Delta F_h^{(\text{irr})}, \quad (177)$$

$$W_h = F(\rho_S^{(A)} \otimes \omega_{\beta_h}(H_{B_h}), H_h^{(0)}) - F(\omega_{\beta_h}(H^{(B)}), H_h^{(0)}) - \Delta F_h^{(\text{irr})}, \quad (178)$$

where

$$\Delta S_h = S(\rho_1) - S(\rho_S^{(A)}), \quad (179)$$

$$\Delta F_h^{(\text{irr})} = F(\rho_S^{(A)} \otimes \omega_{\beta_h}(H_{B_h}), H^{(A)}) - F(\omega_{\beta_h}(H^{(A)}), H^{(A)}), \quad (180)$$

and  $\rho_{B_h}^{(B)} = \text{Tr}_S \left( \omega_{\beta_h}(H^{(B)}) \right)$ .

The expression for  $\Delta F_h^{(\text{irr})}$  is different from what's written in the Supplementary Material.

## 2. Adiabatic expansion.

With both baths disconnected,  $S$  is instantaneously quenched from  $H_S^{(B)}$  to  $H_S^{(C)}$ . The interaction is off, so no heat is exchanged. The work extracted is,

$$W_{B \rightarrow C} = -\text{Tr} \left( (\Delta H_S \otimes \mathbb{1})(\omega_{\beta_h} H^{(B)}) \right) = \text{Tr}_S \left( (H_S^{(B)} - H_S^{(C)}) \rho_1 \right). \quad (181)$$

## 3. Coupling to $B_c$ + isothermal process + decoupling from $B_c$ .

The same three-phase operation is now carried out with  $B_c$  and  $V_c$ . The isothermal transformation is applied from  $H_S^{(C)}$  to  $H_S^{(D)}$ . After decoupling the state of  $S$  is

$$\rho_2 = \text{Tr}_{B_c} \left( \omega_{\beta_c}(H^{(D)}) \right). \quad (182)$$

By direct application of (172) and (168), the heat discharged in the cold-isothermal stroke is

$$Q_c = -T_c \Delta S_c - T_c \left( S(\rho_{B_c}^{(D)} || \omega_{\beta_c}(H_{B_c})) + I(\omega_{\beta_c}(H^{(D)}) ; S : B_c) \right) - \Delta F_c^{(\text{irr})}, \quad (183)$$

while the extracted work during the same stroke is

$$W_c = F(\rho_S^{(C)} \otimes \omega_{\beta_c}(H_{B_c}), H_c^{(0)}) - F(\omega_{\beta_c}(H^{(D)}), H_c^{(0)}) - \Delta F_c^{(\text{irr})}, \quad (184)$$

where

$$\Delta S_c = S(\rho_2) - S(\rho_S^{(C)}), \quad (185)$$

$$\Delta F_c^{(\text{irr})} = F\left(\rho_S^{(C)} \otimes \omega_{\beta_c}(H_{B_c}), H^{(C)}\right) - F\left(\omega_{\beta_c}(H^{(C)}), H^{(C)}\right), \quad (186)$$

$$\text{and } \rho_{B_c}^{(D)} = \text{Tr}_S\left(\omega_{\beta_c}(H^{(D)})\right).$$

The expression for  $\Delta F_c^{(\text{irr})}$  is different from what's written in the Supplementary Material.

#### 4. Adiabatic compression.

Isolated again,  $S$  is quenched back from  $H_S^{(D)}$  to its original Hamiltonian  $H_S^{(A)}$ . Like step 2, this quench is heat-free and contributes only work,

$$W_{D \rightarrow A} = \text{Tr}_S\left((H_S^{(D)} - H_S^{(A)})\rho_2\right). \quad (187)$$

With regard to efficiency, such protocols, which only contain one isothermal process with each bath per cycle, are optimal. Intuitively, this follows since every time the working system is coupled to a heat bath, there is unavoidable dissipation, which decreases the efficiency. It is thus optimal to couple to each bath the minimal number of times so that work extraction remains possible. Later we will prove that this intuition is indeed correct.

Using the Clausius inequality we can write,

$$\Delta S_{\text{tot}} \geq \frac{Q_h}{T_h} + \frac{Q_c}{T_c}. \quad (188)$$

In the weak coupling regime, the cycle is reversible and the inequality turns into

$$\Delta S_{\text{tot}} = -\frac{Q_h}{T_h} - \frac{Q_c}{T_c} \geq 0. \quad (189)$$

Now we insert this into the definition of efficiency,

$$\eta = \frac{W}{Q_h} = 1 + \frac{Q_c}{Q_h} = 1 - \left(\frac{T_c}{T_h}\right) \left[\frac{\Delta S_{\text{tot}}}{Q_h/T_h} + 1\right]. \quad (190)$$

Therefore in order to maximize the efficiency we need to minimize the entropy term. Steps 2 and 4 of the protocol are purely unitary changes of the Hamiltonian that preserve entropy. According to Eq. (105), the system must enter the hot-isothermal stroke already in the Gibbs state of the hot bath. Hence,

$$\rho_S^{(A)} = \omega_{\beta_h}(H_S^{(A)}). \quad (191)$$

The same exact logic applies to the cold-thermal stroke,

$$\rho_S^{(C)} = \omega_{\beta_c}(H_S^{(C)}).$$

The quenches in step 2 and 4, change the Hamiltonian instantaneously but leave the state untouched, meaning

$$\rho_S^{(C)} = \rho_S^{(B)} = \omega_{\beta_h}(H_S^{(B)}), \quad (192)$$

$$\rho_S^{(A)} = \rho_S^{(D)} = \omega_{\beta_c}(H_S^{(D)}). \quad (193)$$

So, in the weak coupling regime, the efficiency of the engine is maximized (obtaining the Carnot bound  $T = 1 - \frac{\beta_h}{\beta_c}$ ) when

$$\omega_{\beta_h}(H_S^{(B)}) = \omega_{\beta_c}(H_S^{(C)}), \quad (194)$$

$$\omega_{\beta_c}(H_S^{(D)}) = \omega_{\beta_h}(H_S^{(A)}), \quad (195)$$

which ensures no dissipation at any point. Hence with Hamiltonians  $H_S^{(B)}$  and  $H_S^{(D)}$  fixed, we optimize over  $H_S^{(A)}$  and  $H_S^{(C)}$ .

From Eqs. (192) and (193), we conclude that  $\Delta S_c$  equals  $\Delta S_h$ . Thus, the efficiency is expressed as

$$\begin{aligned} \eta = 1 + \frac{Q_c}{Q_h} &= 1 + \frac{-T_c \left( \Delta S + S(\rho_{B_c}^{(D)}) \parallel \omega_{\beta_c}(H_{B_c}) \right) + I(\omega_{\beta_c}(H^{(D)}) ; S : B_c) + \Delta F_c^{(\text{irr})} / T_c}{T_h \left( \Delta S - S(\rho_{B_h}^{(B)}) \parallel \omega_{\beta_h}(H_{B_h}) \right) - I(\omega_{\beta_h}(H^{(B)}) ; S : B_h) - \Delta F_h^{(\text{irr})} / T_h} \\ &= 1 + \frac{-T_c \Delta S + T_c S(\rho_{B_c}^{(D)}) \parallel \omega_{\beta_c}(H_{B_c}) + T_c I(\omega_{\beta_c}(H^{(D)}) ; S : B_c) + \Delta F_c^{(\text{irr})}}{T_h \Delta S - T_h S(\rho_{B_h}^{(B)}) \parallel \omega_{\beta_h}(H_{B_h}) - T_h I(\omega_{\beta_h}(H^{(B)}) ; S : B_h) - \Delta F_h^{(\text{irr})}} \\ &= 1 - \frac{T_c}{T_h} \frac{1 + \frac{T_c S(\rho_{B_c}^{(D)}) \parallel \omega_{\beta_c}(H_{B_c}) + T_c I(\omega_{\beta_c}(H^{(D)}) ; S : B_c) + \Delta F_c^{(\text{irr})}}{T_c \Delta S}}{1 - \frac{T_h S(\rho_{B_h}^{(B)}) \parallel \omega_{\beta_h}(H_{B_h}) + T_h I(\omega_{\beta_h}(H^{(B)}) ; S : B_h) + \Delta F_h^{(\text{irr})}}{T_h \Delta S}} \\ &= 1 - \frac{T_c(1 + x_c)}{T_h(1 - x_h)}, \end{aligned} \quad (196)$$

with  $x_{c/h}$  defined by

$$x_c = \frac{T_c S(\rho_{B_c}^{(D)}) \parallel \omega_{\beta_c}(H_{B_c}) + T_c I(\omega_{\beta_c}(H^{(D)}) ; S : B_c) + \Delta F_c^{(\text{irr})}}{T_c \Delta S} \quad (197)$$

$$x_h = \frac{T_h S(\rho_{B_h}^{(B)}) \parallel \omega_{\beta_h}(H_{B_h}) + T_h I(\omega_{\beta_h}(H^{(B)}) ; S : B_h) + \Delta F_h^{(\text{irr})}}{T_h \Delta S}. \quad (198)$$

For an engine producing work, we have  $Q_c = -\Delta E_{B_h} > 0$ , and hence also  $\Delta S > 0$ , which implies that  $x_{c/h} > 0$ . It then follows that the efficiency is lower than the Carnot efficiency, i.e.,  $\eta < \eta^C := 1 - T_c/T_h$ , as expected. In order to maximize the efficiency  $\eta$ , we need to minimize both  $x_{c/h}$ . We fix

$H_S^{(B)}$  and  $H_S^{(D)}$ , so that states  $\rho_1$  and  $\rho_2$  are fixed, and optimize over  $H_S^{(A)}$  and  $H_S^{(C)}$ . Consequently, the only quantity that still needs to be optimized is the irreversible free-energy term  $\Delta F_{c/h}^{(\text{irr})}$ , which was already minimized leading to (105). This implies that the conditions (192) and (193) naturally generalize in the strong coupling to

$$\text{Tr}_B \left( \omega_{\beta_h}(H^{(B)}) \right) = \text{Tr}_B \left( \omega_{\beta_c}(H^{(C)}) \right), \quad (199)$$

$$\text{Tr}_B \left( \omega_{\beta_c}(H^{(D)}) \right) = \text{Tr}_B \left( \omega_{\beta_h}(H^{(A)}) \right). \quad (200)$$

These conditions define the choice of  $H_S^{(A)}$  and  $H_S^{(C)}$  in order to maximize  $\eta$  in the strong coupling regime.

In order to study the limit of small interactions for the Carnot engine, we first note that  $\Delta S$  does depend on  $g$ , as the initial and final states of  $S$ , taken just before and just after contact with the bath, are the reduced states of a global thermal state. We can then expand  $\Delta S$  over  $g$  obtaining

$$T_c \Delta S = T_c \Delta S^{(\text{weak})} + K_S g + O(g^2), \quad (201)$$

$$T_h \Delta S = T_h \Delta S^{(\text{weak})} + K_S g + O(g^2), \quad (202)$$

where  $\Delta S^{(\text{weak})} = S(\omega_{\beta_h}(H_S^{(B)})) - S(\omega_{\beta_h}(H_S^{(A)}))$  and  $K_S$  is the coefficient of the term that is linear in the coupling strength  $g$ ,

$$K_S = \left. \frac{d}{dg} (T_{c/h} \Delta S(g)) \right|_{g=0}. \quad (203)$$

Then, using (175), we can write

$$\begin{aligned} x_{c/h} &= \frac{K_q^{c/h} g^2 + O(g^3)}{T_{c/h} \Delta S^{(\text{weak})} + K_S g + O(g^2)} \text{!!!} = \frac{K_q^{c/h} g^2}{T_{c/h} \Delta S^{(\text{weak})}} \left[ \frac{1}{1 + \frac{K_S}{T_{c/h} \Delta S^{(\text{weak})}} g + O(g^2)} \right] \\ &= \frac{K_q^{c/h} g^2}{T_{c/h} \Delta S^{(\text{weak})}} \left[ 1 - \frac{K_S}{T_{c/h} \Delta S^{(\text{weak})}} g + O(g^2) \right] = \frac{K_q^{c/h}}{T_{c/h} \Delta S^{(\text{weak})}} g^2 + O(g^3), \end{aligned} \quad (204)$$

and

$$\frac{1+x_c}{1-x_h} = \left( 1 - \frac{K_q^c}{Q_c^{(\text{weak})}} g^2 \right) \left( 1 + \frac{K_q^h}{Q_h^{(\text{weak})}} g^2 \right) = 1 + \left( \frac{K_q^c}{Q_c^{(\text{weak})}} + \frac{K_q^h}{Q_h^{(\text{weak})}} \right) g^2 + O(g^3) \text{!!!}, \quad (205)$$

where  $Q_{c/h}^{(\text{weak})} = T_{c/h} \Delta S^{(\text{weak})}$ .

Equation (100) in the Supplementary Material incorrectly reads  $1 - \left( \frac{K_q^c}{Q_c^{(\text{weak})}} + \frac{K_q^h}{Q_h^{(\text{weak})}} \right) g^2 +$

$O(g^3)$ . The correct expansion is the one we obtained.

Using Eqs. (196) and (205) we obtain the desired corrections to Carnot efficiency due to strong coupling,

$$\begin{aligned}\eta &= \underbrace{1 - \frac{T_c}{T_h}}_{\eta^c} - g^2 \frac{T_c}{T_h} \left( \frac{K_q^c}{Q_c^{(\text{weak})}} + \frac{K_q^h}{Q_h^{(\text{weak})}} \right) + O(g^3) \\ &= \eta^c - g^2 \frac{T_c}{T_h} \left( \frac{K_q^c}{Q_c^{(\text{weak})}} + \frac{K_q^h}{Q_h^{(\text{weak})}} \right) + O(g^3).\end{aligned}\tag{206}$$

To conclude, we extended the classical Carnot engine by including strong interactions between the system and the baths. Using a perturbative expansion in the coupling strength  $g$ , we found explicit expression for the loss parameters  $x_{c/h}$ , which capture the main corrections to the ideal efficiency. These corrections appear at second order in  $g$ , and reflect how strong coupling introduces irreversibility.

Now we move on to the next subsection, showing that Carnot-like protocols remain optimal even in the strong-coupling regime.

## 4.1 Carnot-like protocols are optimal

In this section we prove that the Carnot-like protocols considered above are optimal from the point of view of efficiency. To do this, let us first discuss some preliminaries. Consider an arbitrary cyclic protocol. In any such protocol, there are parts of the protocol where the system remains coupled to one of the heat baths, while multiple quenches from some initial Hamiltonian  $H$  to some other Hamiltonian  $H'$  are done on the system. The first observation to make is that the protocol can only become more efficient if we replace this part of the protocol with an isothermal reversible process from  $H$  to  $H'$ , as we showed in the previous sections. (Recall that an isothermal process corresponds to a sequence of infinitesimally small quenches followed by equilibration steps.) This shows that optimal protocols will consist only two kinds of operations:

1. isothermal reversible processes in contact with one of the baths,
2. quenches while not being coupled to the baths (adiabatic compression).

Any such protocol is thus composed of  $n_h$  isothermal process with the hot bath and  $n_c$  isothermal process with the cold bath, with adiabatic quenches in-between. We can then describe any such protocol by  $n_h$  pairs of Hamiltonians  $(H_{h,j}^{(i)}, H_{h,j}^{(f)})$ , denoting the initial and final Hamiltonian of the  $j$ -th isothermal reversible process with the hot bath, and analogously  $n_c$  pairs of Hamiltonians  $(H_{c,j}^{(i)}, H_{c,j}^{(f)})$  for the cold bath. Here, we take the convention that

$$H_{c/h,j}^{(i/f)} = H_{c/h,j,S}^{(i/f)} + V + H_B.\tag{207}$$

Suppose now that a given protocol would have two isothermal processes with one of the baths after each other, only separated by an adiabatic quench. Then this part of the protocol would in general be irreversible and have fixed initial and final states. For concreteness, suppose this would happen with the cold bath and suppose the two initial and final Hamiltonians would be given by  $(H_{c,j}^{(i)}, H_{c,j}^{(f)})$  with  $j = 1, 2$ . Then we could replace this part of the protocol by an isothermal reversible process from  $H_{c,1}^{(i)}$  to  $H_{c,2}^{(f)}$  and the efficiency could only increase. This shows that in optimal protocols, the isothermal reversible process at the two different baths alternate, so that an isothermal process at the hot bath is necessarily followed by an adiabatic quench and an isothermal process at the cold bath (and vice-versa).

Since the protocol has to be cyclic, we already know that  $n_h = n_c$ . What remains to be shown is that optimal protocols have  $n_c = n_h = 1$ . To see this, suppose, for concreteness, that  $n_h = n_c = 2$ . We will now show that we can always describe such a protocol by two sub-cycles with  $n_h = n_c = 1$  that are run sequentially. To do this, let us start the description of the total protocol at the end of the second isothermal at the cold bath, thus starting with state  $\omega_{\beta_{c,2}}(H_{c,2}^{(f)})$ . Then the original protocol proceeds as follows:

1. Adiabatic quench to  $H_{h,1}^{(i)}$ ,
2. Isothermal process to  $H_{h,1}^{(f)}$ ,
3. Adiabatic quench to  $H_{c,1}^{(i)}$ ,
4. Isothermal process to  $H_{c,1}^{(f)}$ ,
5. Adiabatic quench to  $H_{h,2}^{(i)}$ ,
6. Isothermal process to  $H_{h,2}^{(f)}$ ,
7. Adiabatic quench to  $H_{c,2}^{(i)}$ ,
8. Isothermal process back to  $H_{c,2}^{(f)}$ .

To keep things simple, we treat the process of decoupling, changing the Hamiltonian  $H_S$ , and recoupling as a single adiabatic quench. Let us now replace this protocol by the following protocol, which has exactly the same efficiency:

1. Adiabatic quench to  $H_{h,1}^{(i)}$ ,
2. Isothermal process to  $H_{h,2}^{(f)}$ ,
3. Isothermal process from  $H_{h,2}^{(f)}$  to  $H_{h,1}^{(f)}$ ,
4. Adiabatic quench to  $H_{c,1}^{(i)}$ ,

5. Isothermal process to  $H_{c,1}^{(f)}$ ,
6. Adiabatic quench to  $H_{h,2}^{(i)}$ ,
7. Isothermal process to  $H_{h,1}^{(f)}$ ,
8. Isothermal process from  $H_{h,1}^{(f)}$  to  $H_{h,2}^{(f)}$ ,
9. Adiabatic quench to  $H_{c,2}^{(i)}$ ,
10. Isothermal process back to  $H_{c,2}^{(f)}$ .

For clarity, note that the subscripts in  $H_{h,1}^{(i)}$ ,  $H_{h,1}^{(f)}$ ,  $H_{h,2}^{(i)}$  and  $H_{h,2}^{(f)}$  index successive strokes with the same hot bath at temperature  $T_h$ , and do not denote different hot baths. The only change that occurred is a splitting off of isothermal processes into two pieces. Due to reversibility of isothermal processes, this does not change the total work and heat flows and hence this protocol has the same efficiency. However, by rearranging the steps, we can see that we have turned the protocol into two simple sub-cycles.

1. Adiabatic quench to  $H_{c,2}^{(i)}$ ,
2. Isothermal process to  $H_{c,2}^{(f)}$ ,
3. Adiabatic quench to  $H_{h,1}^{(i)}$ ,
4. Isothermal process to  $H_{h,2}^{(f)}$ ,
5. Isothermal process from  $H_{h,2}^{(f)}$  to  $H_{h,1}^{(f)}$ ,
6. Adiabatic quench to  $H_{c,1}^{(i)}$ ,
7. Isothermal process to  $H_{c,1}^{(f)}$ ,
8. Adiabatic quench to  $H_{h,2}^{(i)}$ ,
9. Isothermal process to  $H_{h,1}^{(f)}$ ,
10. Isothermal process from  $H_{h,1}^{(f)}$  to  $H_{h,2}^{(f)}$ .

The first cycle is composed of one isothermal in contact with the hot bath from Hamiltonian  $H_{h,1}^{(i)}$  to  $H_{h,2}^{(f)}$  and an isothermal with the cold bath connecting the Hamiltonian  $H_{c,2}^{(i)}$  to  $H_{c,2}^{(f)}$ . The second cycle consists of the isothermal process in contact with the hot bath connecting the Hamiltonians  $H_{h,2}^{(i)}$  and  $H_{h,1}^{(f)}$  and an isothermal with the cold bath connecting the Hamiltonians  $H_{c,1}^{(i)}$  and  $H_{c,1}^{(f)}$  as before. The two cycles are connected by one isothermal from  $H_{h,1}^{(f)}$  to  $H_{h,2}^{(i)}$  and the same isothermal

run backwards. Due to reversibility, these two isothermals cancel out when calculating the heat and work. We thus conclude that the total work of the protocol and the total heat absorbed from the hot bath are given by

$$W = W_1 + W_2, \quad (208)$$

$$Q = Q_1 + Q_2, \quad (209)$$

where  $W_1$  and  $W_2$  denote the work in each cycles and  $Q_1$  and  $Q_2$  denote the corresponding heat exchanges. The efficiency of each sub-cycle is

$$\eta_1 = \frac{W_1}{Q_1}, \quad \eta_2 = \frac{W_2}{Q_2}, \quad (210)$$

and the total efficiency then yields

$$\eta = \frac{W_1 + W_2}{Q_1 + Q_2} = \frac{\eta_1 Q_1 + \eta_2 Q_2}{Q_1 + Q_2} = \alpha_1 \eta_1 + \alpha_2 \eta_2, \quad (211)$$

where  $\alpha_j := Q_j / (Q_1 + Q_2)$ . Because  $Q_1, Q_2 > 0$ , the coefficients  $\alpha_j$  satisfy  $\alpha_j \geq 0$  and  $\alpha_1 + \alpha_2 = 1$ , making (211) a convex combination. Since such combinations cannot exceed their largest component, we conclude that

$$\eta \leq \max\{\eta_1, \eta_2\} = \max\left\{\frac{W_1}{Q_1}, \frac{W_2}{Q_2}\right\}. \quad (212)$$

A similar construction can be made for any protocol with  $n_h = n_c > 1$ . Hence, the efficiency of any such protocol can be bounded as

$$\eta \leq \max\{\eta_j\} \quad (213)$$

where  $\eta_j$  denotes the efficiency of the  $j$ -th sub-cycle. We conclude that optimal protocols are simple Carnot-cycles as analyzed in the previous sections.

## 5 Power and lower bound on the equilibration time

In this section we show that the equilibration time satisfies  $\tau \geq C/g$  for some constant  $C > 0$ . By equilibration we mean the process in which the initial expectation of any operator  $A(0)$  evolves in time towards a certain value  $\bar{A}$  in which it (approximately) remains.

### 5.1 Lower bound on the equilibration time for a single equilibration

In order to give a lower bound for the equilibration time let us consider the fastest rate of change of  $\langle A(t) \rangle$ , i.e, the quantity

$$v = \sup_t \left| \frac{d}{dt} \langle A(t) \rangle \right|, \quad (214)$$



which, using the Ehrenfest theorem, can be bounded by

$$v = \sup_t |\text{Tr}([A(t), H]\rho)| \leq \sup_t \|[A(t), H]\| = \|e^{iHt}[A, H]e^{-iHt}\| = \|[A, H]\|. \quad (215)$$

Let  $\Delta A := |\langle A(0) - \bar{A} \rangle|$  be the total change that has to occur. Because  $v$  is the maximum instantaneous rate of change, we have

$$\Delta A = \left| \int_0^\tau \frac{d}{dt} \langle A(t) \rangle dt \right| \leq \int_0^\tau \left| \frac{d}{dt} \langle A(t) \rangle \right| dt \leq v\tau. \quad (216)$$

Rearranging Eq. (216) yields a lower bound on the equilibration time,

$$\tau \geq \frac{|\exp(A(0)) - \bar{A}|}{v}. \quad (217)$$

In the particular case of  $A = H_S$ , the rate at which the energy of the system changes during the equilibration is bounded by

$$v \leq \|[H_S, H]\| = g\|[H_S, V]\| = gc, \quad (218)$$

with  $c = \|[H_S, V]\|$ . This leads to an equilibration time lower bounded by

$$\tau \geq \frac{|\Delta E_S|}{gc}, \quad (219)$$

where  $\Delta E_S := |\exp(H_S(0)) - \bar{H}_S|$ .

## 5.2 Lower bound on the equilibration time for the entire cycle of the engine

In order to find the lower bound for the equilibration time of the entire cycle, it will be useful to use the bound of the equilibration time of a single equilibration by means of the energy change of the bath

$$\tau \geq \frac{\delta E_B}{gr}, \quad (220)$$

where now  $r := \|[H_B, V]\|$ . Although in general  $r$  could scale as  $\|H_B\|$ , in practice the Hamiltonian of the bath has a locality structure and the commutator is only non-trivial on the degrees of freedom close to the boundary and  $r$  is independent of the bath's size. Let us decompose the cycle in a heat engine described previously into two main parts.

- **Coupling to the cold bath + isothermal reversible process with the cold bath + decoupling from the cold bath.**

The system is initially already decoupled from the hot bath and thus the global Hamiltonian is non-interacting. After performing the isothermal process with the cold bath, the system is also decoupled from the bath. The first law of thermodynamics for a closed engine cycle yields  $Q_c = -\Delta E_{B_c}$ . Hence, the accumulated energy variation of the bath during all the

protocol steps is given by

$$\sum_i |\delta E_{B_c}^{(i)}| \geq \left| \sum_i \delta E_{B_c}^{(i)} \right| = |\Delta E_{B_c}| = |Q_c|. \quad (221)$$

A lower bound on the time required to perform this part of the protocol is given by

$$\Delta t_c \geq \sum_i \frac{|\delta E_{B_c}^{(i)}|}{gr_c} \geq \frac{|Q_c|}{gr_c}, \quad (222)$$

where  $r_c = \|[H_{B_c}, V_c]\|$  is the maximum rate at which the bath loses or gains energy,  $H_{B_c}$  is the Hamiltonian of the cold bath, and  $V_c$  is the interaction that couples the system to the cold bath.

- **Coupling to the hot bath + isothermal reversible process with the bath+ decoupling from the hot bath.**

Similarly, the time needed for the second part of the cycle can be lower bounded by

$$\Delta t_h \geq \frac{|Q_h|}{gr_h}, \quad (223)$$

where  $r_h = \|[H_{B_h}, V_h]\|$ ,  $H_{B_h}$  is the Hamiltonian of the hot bath, and  $V_h$  is the interaction that couples the system to the hot bath.

The total run-time of the cycle is then bounded by

$$\Delta t = \Delta t_c + \Delta t_h \geq \frac{|Q_c|}{gr_c} + \frac{|Q_h|}{gr_h} = |Q_h| \left( \frac{1}{gr_c} + \frac{1}{gr_h} \right) - \frac{|W|}{gr_c}, \quad (224)$$

where we have considered that for an engine  $|Q_c| = |Q_h| - |W|$ .

### 5.3 Upper bound on the power of the heat engine

Although non-zero interactions between  $S$  and  $B$  tend to increase dissipation, they can enhance power by decreasing the time scale of thermalization  $\tau$ . We now derive an upper bound for such a power enhancement in a Carnot-like engine. From Eq. (224) we obtain a limit in the power of a heat engine in terms of its efficiency and coupling strength, that is,

$$\Delta t \geq \frac{|Q_h|}{g} \left( \frac{1}{r_h} + \frac{1-\eta}{r_c} \right) \implies P = \frac{\eta |Q_h|}{\Delta t} \leq \frac{\eta |Q_h|}{\frac{|Q_h|}{g} \left( \frac{1}{r_h} + \frac{1-\eta}{r_c} \right)} = g\eta \frac{r_c}{1-\eta + \frac{r_c}{r_h}}, \quad (225)$$

where we have used the definition of efficiency  $\eta = W/Q_h$  (with  $W, Q_h > 0$  for an engine). Because  $r_c < r_h$ , the fraction  $r_c/r_h$  is  $< 1$ , giving a looser bound for power,

$$P < gr_h\eta. \quad (226)$$

Define the positive constant

$$\mathcal{C} := \frac{T_c}{T_h} \left( \frac{K_q^c}{Q_c^{(\text{weak})}} + \frac{K_q^h}{Q_h^{(\text{weak})}} \right) > 0,$$

so that the efficiency becomes

$$\eta(g) = \eta^C - \mathcal{C} g^2 + O(g^3). \quad (227)$$

Introduce

$$D(g) := 1 - \eta(g) + \frac{r_c}{r_h} = D_0 + \mathcal{C} g^2 + O(g^3), \quad D_0 := 1 - \eta^C + \frac{r_c}{r_h}. \quad (228)$$

Since  $0 < T_c < T_h$  and  $r_c, r_h > 0$ , we have  $D_0 > 0$ . Expanding  $D(g)$  gives,

$$\frac{1}{D(g)} = \frac{1}{D_0} \frac{1}{1 + (\mathcal{C}/D_0)g^2 + O(g^3)} = \frac{1}{D_0} \left( 1 - \frac{\mathcal{C}}{D_0} g^2 \right) + O(g^3). \quad (229)$$

Using the power bound (Eq. (225)),

$$P(g) \leq \frac{r_c}{D(g)} \eta(g) g,$$

we obtain, up to cubic order in  $g$ ,

$$P(g) \leq \frac{r_c}{D_0} \left( \eta^C g - \frac{\mathcal{C}(D_0 + \eta^C)}{D_0} g^3 \right) + O(g^4). \quad (230)$$

Thus, for small  $g$  the power bound grows linearly with slope  $\frac{r_c}{D_0} \eta^C$ , while the first correction is negative and scales as  $-g^3$ . In other words, power increases in the weak-coupling regime, but the cubic term causes it to bend down as  $g$  becomes larger.

## 6 Conclusion

In this report, we first examined the protocol for work extraction in the strong coupling regime. Through a detailed derivation, we identified the correction terms to the extracted work. Notably, we showed that the act of switching the interaction on and off itself represents a dissipative process at strong coupling. To achieve optimal work extraction, it is crucial to prepare the local Hamiltonians  $H_S^{(1)}$  and  $H_S^{(N)}$  in such a way that they effectively cancel out all  $O(g)$  penalties. This leads to the following expression for the work:

$$W(g) = W_{\text{weak}} - \Delta F_{\min}^{(\text{irr})}(g) - \Delta F_{\min}^{(\text{res})}(g),$$

where both penalties  $\Delta F_{\min}^{(\text{irr})}(g)$  and  $\Delta F_{\min}^{(\text{res})}(g)$  scale as  $O(g^2)$ . In the strong coupling case, the heat dissipated into the bath is corrected by two terms that approach zero in the reversible limit, along with the irreversibility penalty  $\Delta F^{(\text{irr})}$ . Consequently, minimizing dissipated heat coincides with

the optimization of work extraction, as both depend on the same set of conditions.

Next, we established that Carnot-like protocols remain optimal even in the strong coupling regime, meaning the fundamental structure of the process does not change. The analysis of the time and power, however, reveals new constraints: the thermalization speed cannot be arbitrarily increased, which places a limit on the power of any Carnot-like engine. Specifically, we derived a bound on the power in Eq. (226), which reflects how the power scales with coupling strength. Comparing this to the classical case, where the efficiency and power are maximized without the constraints of strong coupling, we observe that while the classical Carnot engine operates at its ideal efficiency, the strong coupling introduces additional terms that lower the power. These terms grow with  $g^2$ , meaning that for small coupling strengths, the power increases linearly but eventually decreases as  $g$  grows larger. Thus, although power initially improves with stronger coupling, it reaches a turning point beyond which it diminishes due to the negative correction terms. This insight is crucial because it shows that while the coupling strength can enhance power up to a certain point, there is a trade-off, and the optimal performance still requires carefully tuning the interaction to avoid excessive dissipation.

In conclusion, the optimal performance of a heat engine in the strong coupling regime, as in the classical case, depends on minimizing irreversible losses and dissipated heat, which ultimately maximizes work.