The maximum efficiency of nano heat engines depends on more than temperature

Mischa Woods,[1,2,3] Nelly Ng,[2,3] and Stephanie Wehner[2,3]
[1]University College of London, Department of Physics & Astronomy, London WC1E 6BT, United Kingdom
[2]QuTech, Delft University of Technology, Lorentzweg 1, 2611 CJ Delft, Netherlands
[3]Centre for Quantum Technologies, National University of Singapore, 117543 Singapore

Sadi Carnot's theorem regarding the maximum efficiency of heat engines is considered to be of fundamental importance in thermodynamics. This theorem famously states that the maximum efficiency depends only on the temperature of the heat baths used by the engine - but not the specific details on how these baths are actually realized. Here, we show that at the nano and quantum scale, this law needs to be revised in the sense that more information about the bath other than its temperature is required to decide whether maximum efficiency can be achieved. In particular, we derive new fundamental limitations of the efficiency of heat engines that show that the Carnot efficiency can only be achieved under special circumstances, and we derive a new maximum efficiency for others. This new understanding of thermodynamics has implications for nanoscale engineering aiming to construct very small thermal machines.

Nicolas Léonard Sadi Carnot is often described as the "father of thermodynamics". In his only publication in 1824 [1] , Carnot gave the first successful theory of the maximum efficiency of heat engines. It was later used by Rudolf Clausius and Lord Kelvin to formalize the second law of thermodynamics and define the concept of entropy [2, 3] . In 1824 he concluded that the maximum efficiency attainable did not depend upon the exact nature of the working fluids [1] :

> *The motive power of heat is independent of the agents employed to realize it; its quantity is fixed solely by the temperatures of the bodies between which is effected, finally, the transfer of caloric.*

For his "motive power of heat", we would today say "the efficiency of a reversible heat engine", and rather than "transfer of caloric" we would say "the reversible transfer of heat." Carnot knew intuitively that his engine would have the maximum efficiency, but was unable to state what that efficiency should be. *Working fluids* refers to the substance (normally gas or liquid) which is at the hot or cold bath temperatures.

Carnot also defined a hypothetical heat engine (now known as the *Carnot engine*) which would achieve the maximum efficiency. Later, this efficiency - now known as the Carnot efficiency (C.E.) - was shown to be

$$\eta_C = 1 - \frac{\beta_{\text{Hot}}}{\beta_{\text{Cold}}}, \tag{1}$$

where β_{Cold}, β_{Hot} are the inverse temperatures of the cold and hot baths respectively.

Unlike the large scale heat engines that inspired thermodynamics, we are now able to build nanoscale quantum machines consisting of a mere handful of particles, prompting many efforts to understand quantum thermodynamics (see e.g. [4–19]). Such devices are too small to admit statistical methods, and many results have shown that the workings of thermodynamics become more intricate in such regimes [4–7, 15] .

We show in this report that unlike at the macroscopic scale - at which Carnot's fundamental results hold - there are new fundamental limitations to the maximal efficiency at the nanoscale. Most significantly, this new efficiency *depends* on the working substance. We find that the C.E. can be achieved, but only when the working substance is of a particular form. Otherwise, a reduced efficiency is obtained, highlighting the significant difference in the performance of heat engines as our devices decrease in size.

Work in the nanoregime

The basic components of a heat engine (H.E.) are detailed in Fig. 1. The definition of work when dealing with nanoscopic quantum systems has seen much attention lately [4–9] . Performing work is always understood as changing the energy of a system, which we call *battery*. In the macroregime, one often pictures raising a weight on a string. In the nanoregime, this corresponds to changing the energy of a quantum system by lifting it to an excited state (see Fig. 2).

One aspect of extracting work W is to bring the battery's initial state ρ_{W}^0 to some final state ρ_{W}^1 such that $W = \text{tr}(\rho_{\text{W}}^1 \hat{H}_{\text{W}}) - \text{tr}(\rho_{\text{W}}^0 \hat{H}_{\text{W}}) > 0$. However, a change in energy alone, does not yet correspond to performing work. It is implicit in our macroscopic understanding of work that the energy transfer takes place in an ordered form. When lifting a weight, we *know* its final position and can exploit this precise knowledge to transfer all the work onto a third system without - in principle - losing any energy in the process. In the quantum regime, such knowledge corresponds to ρ_{W}^1 being a pure state. When ρ_{W}^1 is diagonal in the energy eigenbasis of $\hat{H}_{\text{W}}$, then ρ_{W}^1 is an energy eigenstate. We can thus understand work as an energy transfer about which we have perfect information, while heat, in contrast, is an energy transfer about which we hold essentially no information. Clearly, there is also an intermediary regime in which we transfer energy, while having some - but not perfect - information.

To illustrate this idea, consider a two-level system battery, where we extract work by transiting from an initial energy eigenstate $|E_{\text{W}}^j\rangle\langle E_{\text{W}}^j|$ to another energy eigenstate $|E_{\text{W}}^k\rangle\langle E_{\text{W}}^k|$, where $E_{\text{W}}^k - E_{\text{W}}^j > 0$. Changing the energy, while having some amount of information corresponds to changing the state of the battery to a mixture $\rho_{\text{W}}^1 = (1-\varepsilon)|E_{\text{W}}^k\rangle\langle E_{\text{W}}^k| + \varepsilon|E_{\text{W}}^j\rangle\langle E_{\text{W}}^j|$ for some parameter $\varepsilon \in [0, 1]$. The case of $\varepsilon = 0$ corresponds to doing *perfect work*. The smaller ε is, the closer

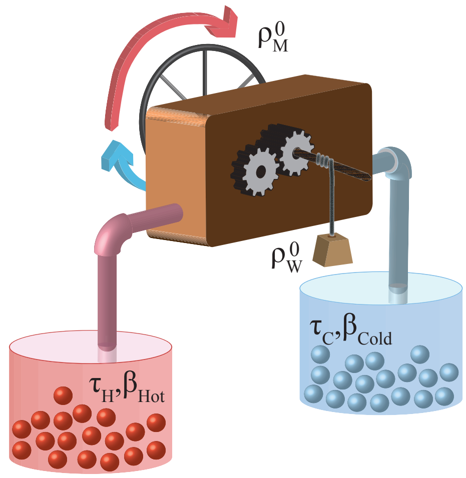

FIG. 1. **A heat engine** extracts work from the temperature difference of a hot bath (red) at temperature T_{Hot} and cold bath (blue) at temperature T_{Cold}. The term bath indicates that the initial states of these systems are thermal, $\tau_{\text{Hot}} = \exp(-\beta_{\text{Hot}}\hat{H}_{\text{Hot}})/Z_{\text{Hot}}$ and $\tau_{\text{Cold}} = \exp(-\beta_{\text{Cold}}\hat{H}_{\text{Cold}})/Z_{\text{Cold}}$ with inverse temperatures $\beta_{\text{Hot}} = 1/T_{\text{Hot}}$ and $\beta_{\text{Cold}} = 1/T_{\text{Cold}}$, and partition functions Z_{Hot} and Z_{Cold} respectively. The machine itself corresponds to a quantum system M with Hamiltonian $\hat{H}_{\text{M}}$, starting in an arbitrary state ρ^0_{M}. The battery indicates the system in which we will store work and is illustrated in detail (Figure 2). Let ρ^0_{W} be the starting state of the work system, and $\hat{H}_{\text{W}}$ its Hamiltonian. Operating the heat engine for one cycle for some particular time t, corresponds to applying a unitary transform $U(t)$ to the both baths, the actual machine, and also the work system. In order to account for all energy transfers we will demand that $[U(t), \hat{H}] = 0$, where $\hat{H} = \hat{H}_{\text{M}} + \hat{H}_{\text{Cold}} + \hat{H}_{\text{Hot}} + \hat{H}_{\text{W}}$. That is, $U(t)$ conserves total energy. We will furthermore demand that operating the heat engine does not affect the actual machine, i.e., the final state of the machine $\rho^1_{\text{M}} = \rho^0_{\text{M}}$, but the choice of ρ^0_{M} may affect the efficiencies we can obtain. In the language of [5] , we thus treat the machine as a catalyst. The *efficiency* $\eta = W_{\text{ext}}/\Delta H$ of the machine measures the amount of work extracted W_{ext} in relation to the energy invested as measured by the energy change of the hot bath ΔH. Given two temperatures T_{Hot} and T_{Cold}, finding the optimal heat engine for such temperatures would correspond to deciding on the best machine (ρ^0_{M}, $\hat{H}_{\text{M}}$), evolution $U(t)$, and bath structures $\hat{H}_{\text{Hot}}$ and $\hat{H}_{\text{Cold}}$. To understand the role of the structure of the cold bath, we here fix $\hat{H}_{\text{Cold}}$ while allowing an otherwise arbitrary choice of $\hat{H}_{\text{Hot}}$.

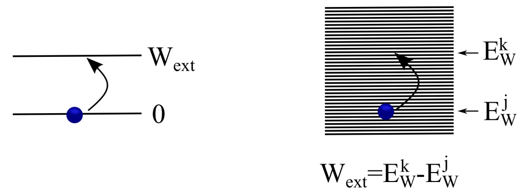

FIG. 2. **A battery** is a work-storage component of a heat engine. In the nanoregime, a minimal way of modeling the battery is as a two-level system [5] . Performing work corresponds to "lifting" the state from ground state to the excited state, where the energy gap is fine-tuned to the amount of work W_{ext} to be done. While an arbitrary energy spacing is difficult to realize in a two-level system, it can be done by picking two levels with the desired spacing from a quasi-continuum battery: this battery comprises of a large but finite number of discrete levels which form a quasi-continuum. Such a battery closely resembles the classical notion of a "weight attached to a string" as considered in [21] . The battery can be charged from a particular state (e.g. the ground state) to any of the higher levels.

we are to the situation of perfect work. One can characterize this intermediary regime by the von Neumann entropy $S(\rho^1_{\text{W}})$. For perfect work, $S(\rho^1_{\text{W}}) = 0$, while for heat transfer, under a fixed average energy, the two-level battery becomes thermal, since the thermal state maximizes entropy for a fixed energy [20] . Intuitively, we may also think about ε as the failure probability of extracting work. When $\varepsilon > 0$, what is relevant is not ε as an absolute, but relative to the energy W_{ext} that is extracted. We are thus interested in $\Delta S/W_{\text{ext}}$ where $\Delta S = S(\rho^1_{\text{W}}) - S(\rho^0_{\text{W}})$ is the change in entropy of the battery. For our investigations the limit $\Delta S/W_{\text{ext}} \to 0$ will be of particular interest, and corresponds to performing *near perfect work*. Our analysis applies to arbitrarily small heat engines, even if this machine was run for only one cycle. We emphasize that this is not a restriction of the analysis, but rather a strong and appealing feature because it is indeed the relevant case when we consider few qubit devices, and a small number of experimental trials.

No perfect work

Before establishing our main result, we first show that in the nanoscopic regime, no heat engine can perform perfect work ($\varepsilon = 0$). That is, the efficiency of any such heat engine is zero. More formally, it means that there exists no global energy preserving unitary (see Fig. 1) for which $W_{\text{ext}} > 0$.

Efficiency

Clearly, however, heat engines *can* be built, prompting the question how this might be possible. We show that for *any* $\varepsilon > 0$, there exists a process such that $W_{\text{ext}} > 0$. Therefore, a heat engine is possible if we ask only for near perfect work. Interestingly, even in the macroscopic regime, we can envision a heat engine that only extracts work with probability $1-\varepsilon$, but over many cycles of the engine we do not notice this feature when looking at the average work gained in each run.

To study the efficiency in the nanoscale regime, we make crucial use of the second laws of quantum thermodynamics [5] . It is apparent from these laws that we might only discover further limitations to the efficiency than we see at the macroscopic scale. Indeed they do arise, as we find that the efficiency no longer depends on just the temperatures of the heat baths. Instead, the explicit structure of the cold bath Hamiltonian $\hat{H}_{\text{Cold}}$ becomes important (a similar argument can be made for the hot bath). Consider a cold bath comprised of n two-level systems each with its own energy gap, where n can be arbitrarily large, but finite. Let us denote the spectral gap of the cold bath—the energy gap between its ground state and first excited state—by E_{min}. We can then define the quantity

$$\Omega = \frac{E_{\text{min}}(\beta_{\text{Cold}} - \beta_{\text{Hot}})}{1 + e^{-\beta_{\text{Cold}} E_{\text{min}}}}, \tag{2}$$

and study the efficiency in the *quasi-static* limit. This means that the final state of the cold bath is thermal, and its final temperature T_f is higher than T_{Cold} by only a positive infinitesimal amount.

Whenever $\Omega \leq 1$, we show that the maximum and attainable efficiency is indeed the familiar C.E., which can be expressed as

$$\eta = \left(1 + \frac{\beta_{\text{Hot}}}{\beta_{\text{Cold}} - \beta_{\text{Hot}}}\right)^{-1}. \tag{3}$$

However, when $\Omega > 1$, we find a new nanoscale limitation. In this situation, the efficiency is only

$$\eta = \left(1 + \frac{\beta_{\text{Hot}}}{\beta_{\text{Cold}} - \beta_{\text{Hot}}}\Omega\right)^{-1} \tag{4}$$

for a quasi-static heat engine. One might hope to obtain a higher efficiency compared to (4) by going away from the quasi-static setting, however we also show that such an efficiency is always strictly less than the C.E.

The restriction of near perfect work per cycle can now be further justified by examining how well the heat engine performs when the machine runs over many cycles: we find that if $\Omega \leq 1$, the heat engine can be run quasi-statically with an efficiency arbitrarily close to the C.E. while extracting *any* finite amount of work with an arbitrarily small entropy increase in the battery.

Comparison with standard entropy results

For any system in thermal contact with a bath at temperature T, consider the Helmholtz free energy $F(\rho) = \text{Tr}(\hat{H}\rho) - kTS(\rho)$, where $S(\rho) = -\text{tr}(\rho \ln \rho)$ is the von Neumann entropy of ρ. In the macroregime, the usual second law states that the Helmholtz free energy never increases,

$$F(\rho_0) \geq F(\rho_1), \tag{5}$$

when the system goes from a state ρ_0 to a state ρ_1. This, however, is but one of many conditions necessary for a state transformation [5] . The limitations we observe are a consequence of the fact that in the nanoregime, possible transitions are governed by a family of generalized second laws. The fact that more laws appear in this regime can intuitively be understood as being analogous to the fact that when performing a probabilistic experiment only a handful of times, not just the average, but other moments of a distribution become relevant. Indeed, all second laws converge to the standard second law in the limit of infinitely many particles [5], illustrating why we are traditionally accustomed to only this second law. The standard second law also emerges in some regimes of inexact catalysis [5, 22], however, this corresponds to a degradation of the machine in each cycle.

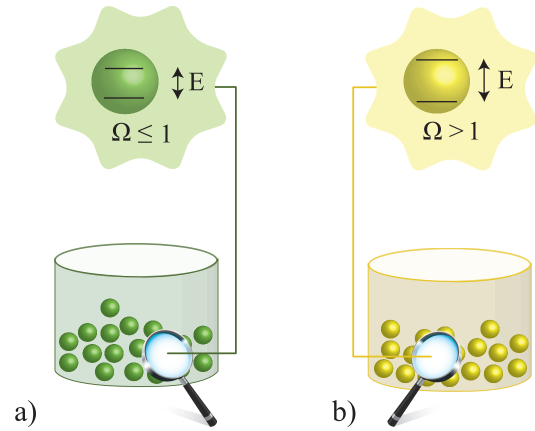

FIG. 3. **For fixed temperatures T_{Cold}, T_{Hot}, the efficiency of a nanoscale heat engine depends on the structure of the cold bath.** At the nano/quantum scale, Carnot's statement about the universality of heat engines does not hold. We find that the maximum efficiency of a heat engine, *does* depend on the "working fluid". In (a) the energy gaps are small enough to allow the heat engine to achieve C.E., i.e., $\Omega \leq 1$. In (b) the efficiency of the heat engine is reduced below the C.E. because the energy gap of the qubits are above the critical value $\Omega > 1$.

It is illustrative to analyze our problem when we apply just the standard second law in (5) to derive bounds on the efficiency, which is indeed a matter of textbook thermodynamics [23] . However, we here apply the law precisely to the heat engine model as given in Fig. 1, in which all energy flows are accounted for and (near) perfect work is performed. One might wonder whether the limitations we observe are just due to an inaccurate model or our demand for near perfect instead of average work, and might thus also arise in the macroregime. That is, are these newfound limitations really a consequence of the need to obey a wider family of second laws, or would the standard free energy predict the same things when the energy is quantized?

We show independently of whether we consider perfect ($\varepsilon = 0$) or near perfect ($\Delta S/W_{\text{ext}} \to 0$) work - that according to the standard free energy in (5), the maximum achievable efficiency is the C.E. Furthermore, we recover Carnot's famous statement that the C.E. can be achieved for *any* cold bath (i.e. for a cold bath with any finite dimensional pure point spectrum). We also see that C.E. is only achieved for quasi-static H.E.s. We prove this without invoking any additional assumptions than those laid out here, such as reversibility or that the system is in thermodynamic equilibrium at all times. Therefore, with our setup we recover exactly what Carnot predicted, namely that the maximum efficiency of the H.E. is independent of the working substance. This rules out that our inability to achieve what Carnot predicted according to the macroscopic laws of thermodynamics is not only a consequence of an overly stringent model.

Extensions to the setup

One could ask whether at the nano regime, if a less stringent model would allow one to recover Carnot's predictions. Specifically, what if we consider any final state of the battery ρ_{W}^{1},which is ε away in trace distance from the desired final battery state $|E_{\mathrm{W}}^{k}\rangle\langle E_{\mathrm{W}}^{k}|$? In this case, we show that as long as one still considers the extraction of near perfect work $\Delta S/W_{\mathrm{ext}} \rightarrow 0$, our findings remain unchanged: when $\Omega > 1$, C.E. cannot be obtained.

In a similar vein, one could imagine that the final components of the heat engine become correlated between themselves, and that this would allow one to always achieve the C.E.. According to macroscopic laws of thermodynamics, correlations between the final components always inhibit one from achieving the C.E.. We show that at the nanoscale such correlations can also be ruled out as a means to achieve the C.E. when $\Omega > 1$.

These results thus show the inevitability that the maximum efficiency of a nanoscale heat engine depends on more information about the thermal baths rather than just the temperature.

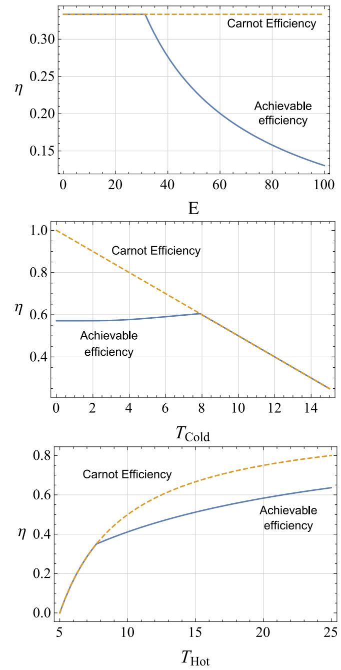

FIG. 4. **Comparison of the nano/quantum-scale efficiency versus macroscopic efficiency (C.E.), in the quasi-static regime. Top**: Efficiency vs. the energy gap $E_{\min}$ of $\hat{H}_{\mathrm{Cold}}$. According to Eqs. (3), (4), for any $\beta_{\mathrm{Hot}} < \beta_{\mathrm{Cold}}$ ($T_{\mathrm{Cold}} < T_{\mathrm{Hot}}$) one can achieve C.E. when the structure $\hat{H}_{\mathrm{Cold}}$ of the cold bath has sufficiently small energy gap. If the energy gap is too large, we find a reduced efficiency. This has been plotted for $T_{\mathrm{Hot}} = 15$ and $T_{\mathrm{Cold}} = 10$. **Middle**: Efficiency vs. T_{Cold}. For every $\hat{H}_{\mathrm{Cold}}$, there exists a temperature regime (T_{Cold} vs. T_{Hot}) such that C.E. cannot be achieved. This happens as T_{Cold} gets further from the temperature of the hot bath $T_{\mathrm{Hot}} = 20$, where we have used $E_{\min} = 15$. **Bottom**: Efficiency vs. T_{Hot}. Similarly, we see this feature of not being able to achieve C.E., as the temperature of the hot bath increases relative to $T_{\mathrm{Cold}} = 5$, where again $E_{\min} = 15$.

Conclusion

Our work establishes a fundamental result for the operation of nanoscale heat engines. We find all cold baths can be used in heat engines. However, for all temperatures T_{Cold} and T_{Hot} of the cold and hot baths, there exists an energy gap $E_{\min}(\beta_{\mathrm{Cold}}, \beta_{\mathrm{Hot}})$ of the two-level systems forming the cold bath above which the optimal efficiency is reduced below the C.E.. Viewed from another direction, for a fixed energy gap $E_{\min}(\beta_{\mathrm{Cold}}, \beta_{\mathrm{Hot}})$, whether the C.E. can be achieved depends on the relation between T_{Hot} and T_{Cold} as illustrated in Fig. 4. Loosely speaking, the C.E. can be achieved whenever the two temperatures are unequal but not too far apart. One might wonder why this restriction has not been observed before in the classical scenario. There, the energy spectrum is continuous or forms a quasi-continuum, and hence we can only access the C.E. regime.

Our result is a consequence of the fact that the second law takes on a more complicated form in the nanoregime. Next to the standard second law, many other laws become relevant and lead to additional restrictions. From a statistical perspective, small numbers require more refined descriptions than provided by averages, and as a result thermodynamics becomes more complicated when considering systems comprised of few particles. Similar effects can also be observed in information theory, where averaged quantities as given by the Shannon entropy need to be supplemented with refined quantities when we consider finitely many channel uses.

In the macroscopic regime, for completeness, we ruled out the possibility that the observed limitations on efficiency is a consequence of our demand for near perfect work, or the fact that we are using systems with discrete (sufficiently large spaced) spectra. This verification was achieved by showing that the C.E. can indeed always be attained (regardless to the size of an energy gap if present) when extracting near perfect work, when we are in such large systems that only the standard second law is relevant. One might wonder whether heat engines that do not operate quasi-statically, or employing quantum coherences would allow us to achieve the C.E. independent of the structure of the cold bath. As we show in the Supplementary Material, both do not help.

There are several works [15, 21, 24–26] that have analyzed the efficiencies of heat engines and obtained C.E. as the maximal efficiency. Common to all these approaches is that they consider an average notion of work, without directly accounting for a contribution from disordered energy (heat). Instead, one keeps the entropy of the battery low [21] , or bound the higher moments of the energy distribution [25] . These only limit contributions from heat, but do not fully prevent them. Our notion of (near) perfect work now makes this aspect of macroscopic work explicit in the nanoregime. Needless to say, imperfect work with some contribution of heat can also be useful. Yet, it does not quite constitute work if we cannot explicitly single out a contribution from heat. One could construct a machine which extracts some amount of energy, with some non-negligible amount of information. We can prove in this case that Carnot's efficiency can even be exceeded [27] .

This should not come as a surprise, because we are no longer asking for work - energy transfer about which we have (near) perfect information.

Our work raises many open questions. We see that the quasi-static efficiency has a discontinuous derivative with respect to T_{Hot}, T_{Cold} or E_{min} at $\Omega = 1$, as illustrated in Fig. 4, which is often associated with a phase transition. It is unclear whether this phenomenon can also be understood as a phase transition - absent at the macroscopic scale - and whether there is an abrupt change in the nature of the machine when crossing the $\Omega = 1$ boundary.

It would furthermore be satisfying to derive the explicit form of a hot bath, and machine attaining Carnot - or new Carnot - efficiency. One might wonder whether a non-trivial machine $(\rho_{\text{M}}, \hat{H}_{\text{M}})$ is needed at all in this case. To illustrate the dependence on the bath, it was sufficient to consider a bath comprising solely of qubits. The tools proposed in the Supplementary Material can also be used to study other forms of bath structures, yet it is a non-trivial question to derive efficiencies for such cold baths.

Most interestingly, there is the extremely challenging question of deriving a statement that is analogous to the C.E., but which makes explicit the trade-off between information and energy for all possible starting situations. In a heat engine, we obtain energy from two thermal baths about which we have minimal information. It is clear that the C.E. is thus a special case of a more general statement in which we start with two systems of a certain energy about which we may have some information, and we want to extract work by combining them. Indeed, the form that such a general statement should take is by itself a beautiful conceptual challenge, since what we understand as efficiency may not only be a matter of work obtained vs. energy wasted. Instead, we may want to take a loss of information about the initial states into account when formulating such a fully general efficiency.

[1] S. Carnot, *Reflections on the Motive Power of Fire* (1824).
[2] R. Clausius, Annalen der Physik **201**, 353 (1865).
[3] W. Thompson (Lord Kelvin), (1851).
[4] M. Horodecki and J. Oppenheim, Nature Communications **4** (2013).
[5] F. Brandão, M. Horodecki, N. Ng, J. Oppenheim, and S. Wehner, Proceedings of the National Academy of Sciences **112**, 3275 (2015).
[6] J. Åberg, Nature Communications **4** (2013).
[7] O. Dahlsten, R. Renner, E. Rieper, and V. Vedral, New Journal of Physics **13**, 053015 (2011).
[8] R. Gallego, J. Eisert, and H. Wilming, (2015), arXiv:1504.05056.
[9] J. Gemmer and J. Anders, arXiv:1504.05061 (2015).
[10] M. Lostaglio, D. Jennings, and T. Rudolph, Nature communications **6** (2015).
[11] M. Scully, M. Zubairy, G. Agarwal, and H. Walther, Science **299**, 862 (2003).
[12] M. Lostaglio, K. Korzekwa, D. Jennings, and T. Rudolph, Phys. Rev. X **5**, 021001 (2015).
[13] F. Binder, S. Vinjanampathy, K. Modi, and J. Goold, Phys. Rev. E **91**, 032119 (2015).
[14] S. Salek and K. Wiesner, (2015), arXiv:1504.05111.
[15] H. Tajima and M. Hayashi, arXiv:1405.6457 (2014).
[16] J. Gemmer, M. Michel, and G. Mahler, *Quantum thermodynamics* (Springer, 2009).
[17] J. Roßnagel, O. Abah, F. Schmidt-Kaler, K. Singer, and E. Lutz, Phys. Rev. Lett. **112**, 030602 (2014).
[18] B. Gardas and S. Deffner, Phys. Rev. E **92**, 042126 (2015).
[19] M. Perarnau-Llobet, K. V. Hovhannisyan, M. Huber, P. Skrzypczyk, N. Brunner, and A. Acín, Phys. Rev. X **5**, 041011 (2015).
[20] E. Jaynes, Physical Review **106**, 620 (1957).
[21] P. Skrzypczyk, A. Short, and S. Popescu, Nature Communications **5** (2014).
[22] M. Lostaglio, M. P. Müller, and M. Pastena, Phys. Rev. Lett. **115**, 150402 (2015).
[23] F. Reif, *Fundamentals of statistical and thermal physics*, McGraw-Hill series in fundamentals of physics (McGraw-Hill, 1965).
[24] P. Skrzypczyk, N. Brunner, N. Linden, and S. Popescu, Journal of Physics A: Mathematical and Theoretical **44**, 492002 (2011).
[25] N. Brunner, N. Linden, S. Popescu, and P. Skrzypczyk, Phys. Rev. E **85**, 051117 (2012).
[26] N. Brunner, M. Huber, N. Linden, S. Popescu, R. Silva, and P. Skrzypczyk, Phys. Rev. E **89**, 032115 (2014).
[27] N. Ng, M. Woods, and S. Wehner, (2015), in preparation.
[28] V. Shoup, *A computational introduction to number theory and algebra* (Cambridge university press, 2009) p. 50.
[29] T. M. Cover and J. A. Thomas, *Elements of information theory* (John Wiley & Sons, 2012).
[30] C. Shannon, Bell System Technical Journal **27**, 379 (1948).

Acknowledgments

We thank J. Oppenheim, M. Horodecki, R. Renner and A. Winter for useful comments and discussions. This research was funded by Singapore's MOE Tier 3A Grant MOE2012-T3-1-009, and STW, Netherlands.

Methods

Setup of heat engine The workings of a heat engine have been described in Fig. 1, which we expand in mathematical detail here. Consider the initial global system

$$\rho^0_{\text{ColdHotMW}} = \rho^0_{\text{Cold}} \otimes \rho^0_{\text{Hot}} \otimes \rho^0_{\text{M}} \otimes \rho^0_{\text{W}}. \tag{6}$$

The hot bath Hamiltonian $\hat{H}_{\text{Hot}}$ can be chosen arbitrarily, as long as ρ^0_{Hot} is the corresponding thermal state of temperature T_{Hot}. Similarly, the machine and its Hamiltonian $(\rho^0_{\text{M}}, \hat{H}_{\text{M}})$ can be chosen arbitrarily. Given any cold bath $(\rho^0_{\text{Cold}}, \hat{H}_{\text{Cold}})$ such that ρ^0_{Cold} is a thermal state at temperature $T_{\text{Cold}} < T_{\text{Hot}}$, one can extract work and store it in system W. This process then corresponds to inducing the transition

$$\rho^0_{\text{ColdHotMW}} \to \rho^1_{\text{ColdHotMW}}, \tag{7}$$

where $\text{tr}_{\text{Hot}}(\rho^1_{\text{ColdHotMW}}) = \rho^1_{\text{ColdW}} \otimes \rho^1_{\text{M}}$. We have $\rho^0_{\text{M}} = \rho^1_{\text{M}}$, that is, the machine is not degraded in the process. This also means that we preserve the tensor product structure between ColdW and M: if the machine is initially correlated with some other system, we do not want to destroy such correlations since this would also mean that we degrade the machine.

To quantify the amount of extractable work, we apply the generalized second laws derived in [5] . The initial cold bath ρ^0_{Cold} is thermal, and therefore diagonal in the energy eigenbasis, while the initial battery state ρ^0_{W} is also a pure energy eigenstate (see Fig. 2). Since here energy conserving $U(t)$ never increase coherences between energy eigenstates [5] , we can therefore conclude that ρ^1_{ColdW} is also diagonal in the energy eigenbasis. We can thus invoke the *necessary and sufficient conditions* for a transformation to be possible [5] . Specifically, $\rho^0_{\text{Cold}} \otimes \rho^0_{\text{W}} \to \rho^1_{\text{ColdW}}$ iff $\forall \alpha \geq 0$,

$$F_\alpha(\rho^0_{\text{Cold}} \otimes \rho^0_{\text{W}}, \tau^h_{\text{ColdW}}) \geq F_\alpha(\rho^1_{\text{ColdW}}, \tau^h_{\text{ColdW}}), \tag{8}$$

where τ^h_{ColdW} is the thermal state of the joint system (cold bath and battery) at temperature T_{Hot}. The generalized free energy F_α is defined as

$$F_\alpha(\rho, \tau) := \frac{1}{\beta_{\text{Hot}}} \left[D_\alpha(\rho\|\tau) - \ln Z_{\text{Hot}} \right], \tag{9}$$

where $D_\alpha(\rho\|\tau)$ are known as α-Rényi divergences. For states ρ, τ which are diagonal in the same eigenbasis, the Rényi divergences can be simplified to

$$D_\alpha(\rho\|\tau) = \frac{1}{\alpha - 1} \ln \sum_i p_i^\alpha q_i^{1-\alpha}, \tag{10}$$

where p_i, q_i are the eigenvalues of ρ and τ respectively. The case $\alpha = 1$ is defined by continuity in α. Taking the limit $\alpha \to 1$ for (9), one recovers the Helmholtz free energy, $F(\rho) = \langle \hat{H} \rangle_\rho - \beta^{-1}_{\text{Hot}} S(\rho)$. Using the second laws [5] is a powerful tool, since when searching for the optimum efficiency, we do not have to optimize explicitly over the possible machines $(\rho_{\text{M}}, \hat{H}_{\text{M}})$, the form of the hot bath $\hat{H}_{\text{Hot}}$, or the energy conserving unitary $U(t)$. Whenever (8) is satisfied, then we are guaranteed a suitable choice exists and hence we can focus solely on the possible final states ρ^1_{ColdW}.

Since we know that ρ^1_{ColdW} is a diagonal in the energy eigenbasis, the correlations between cold bath and battery can only be classical (w.r.t. energy eigenbasis). However, even such correlations cannot improve the efficiency: we show in the Supplementary Material that we may take the output state to have the form $\rho^1_{\text{ColdW}} = \rho^1_{\text{Cold}} \otimes \rho^1_{\text{W}}$ in order to achieve the maximum efficiency. According to Fig. 2, consider $\rho^0_{\text{W}} = |E^j_{\text{W}}\rangle\langle E^j_{\text{W}}|$ and $\rho^1_{\text{W}} = (1-\varepsilon)|E^k_{\text{W}}\rangle\langle E^k_{\text{W}}| + \varepsilon|E^j_{\text{W}}\rangle\langle E^j_{\text{W}}|$. From the second laws (8), we see that the maximum amount of extractable work is given by the largest value of $W_{\text{ext}} = E^k_{\text{W}} - E^j_{\text{W}}$ such that the state transition $\rho^0_{\text{Cold}} \otimes \rho^0_{\text{W}} \to \rho^1_{\text{Cold}} \otimes \rho^1_{\text{W}}$ is possible. The form of W_{ext} (derived in the Supplementary Material) is

$$W_{\text{ext}} = \inf_{\alpha \geq 0} W_\alpha, \tag{11}$$

$$W_\alpha = \frac{1}{\beta_{\text{Hot}}(\alpha - 1)} [\ln(A - \varepsilon^\alpha) - \alpha \ln(1-\varepsilon)], \tag{12}$$

$$A = \frac{\sum_i p_i^\alpha q_i^{1-\alpha}}{\sum_i p_i'^\alpha q_i^{1-\alpha}}, \tag{13}$$

where $p_i = \frac{e^{-\beta_{\text{Cold}} E_i}}{Z_{\beta_{\text{Cold}}}}$, $q_i = \frac{e^{-\beta_{\text{Hot}} E_i}}{Z_{\beta_{\text{Hot}}}}$ are probabilities of the thermal state of the cold bath at temperatures $T_{\text{Cold}}, T_{\text{Hot}}$ respectively, and p_i' are the probability amplitudes of state ρ^1_{Cold} when written in the energy eigenbasis of $\hat{H}_{\text{Cold}}$. The quantity W_{ext} is dependent on the initial and final cold bath $\rho^0_{\text{Cold}}, \rho^1_{\text{Cold}}$, the hot bath temperature T_{Hot}, and the allowed failure probability ε. The main difficulty of evaluating W_{ext} comes from the infimum over α, - indeed we know examples in which it can be obtained at any $\alpha \geq 0$ depending on $\beta_{\text{Hot}}, \beta_{\text{Cold}}$ and other parameters.

The efficiency η, however, is not determined the maximum extractable work, but rather by a tradeoff between W_{ext} and the energy drawn from the hot bath. More precisely,

$$\eta := \frac{W_{\text{ext}}}{\Delta\text{Hot}}, \tag{14}$$

where

$$\Delta\text{Hot} := \text{tr}(\hat{H}_{\text{Hot}}\rho^0_{\text{Hot}}) - \text{tr}(\hat{H}_{\text{Hot}}\rho^1_{\text{Hot}}) \tag{15}$$

is the mean energy *drawn* from the hot bath. Since $\hat{H}_{\text{ColdHotMW}} = \hat{H}_{\text{Cold}} + \hat{H}_{\text{Hot}} + \hat{H}_{\text{M}} + \hat{H}_{\text{W}}$ is void of interaction terms, and since total energy is preserved, we can also write the change of energy in the hot bath, in terms of the energy change in the remaining systems. That is,

$$\Delta\text{Hot} = \Delta\text{Cold} + \Delta W. \tag{16}$$

where

$$\Delta\text{Cold} := \text{tr}\left[H_{\text{Cold}}\rho^1_{\text{Cold}}\right] - \text{tr}\left[H_{\text{Cold}}\rho^0_{\text{Cold}}\right], \tag{17}$$

and

$$\Delta W := \mathrm{tr}(\hat{H}_{\mathrm{W}}\rho_{\mathrm{W}}^{1}) - \mathrm{tr}(\hat{H}_{\mathrm{W}}\rho_{\mathrm{W}}^{0}). \tag{18}$$

is the *change* in average energy of the cold bath and battery. We thus see that the efficiency can be described wholely in terms of the battery and the cold bath.

Macroscopic second law We first analyze the efficiency in the macroscopic regime, where only the usual free energy ($\alpha = 1$) dictates if a certain state transition is possible. The main question is then: given an initial cold bath Hamiltonian $\hat{H}_{\mathrm{Cold}}$, what is the maximum attainable efficiency considering all possible final states ρ_{Cold}^{1}? In both cases of perfect and near perfect work, we find that the efficiency is only maximized whenever ρ_{Cold}^{1} is (A) a thermal state, and (B) has a temperature β_f arbitrarily close to β_{Cold}. We refer to this situation as a *quasi-static* heat engine. Moreover, we find that the maximum is the C.E. and that it can be achieved by any given $\hat{H}_{\mathrm{Cold}}$. These results rigorously prove Carnot's findings when only the usual free energy is relevant.

Nanoscale second laws Here, when considering perfect work, we are immediately faced with an obstacle: the constraint at $\alpha = 0$ implies that $W_{\mathrm{ext}} > 0$ is not possible, whenever ρ_{Cold}^{0} is of full rank. This is due to the discontinuity of D_0 in the state probability amplitudes, and is similar to effects observed in information theory in lossy vs. lossless compression: no compression is possible if no error however small is allowed. However, when considering near perfect work, the D_0 constraint is satisfied automatically. We thus consider the limit $\Delta S/W_{\mathrm{ext}} \to 0$.

The results for the macroscopic second law implies an upper bound on both the maximum extractable work and efficiency for nanoscopic second laws, since the constraint of generalized free energy at $\alpha = 1$ is simply one of the many constraints described by all $\alpha \geq 0$. It thus follows from the macroscopic second law results, that if we can achieve the C.E., we can only do so when both (A) and (B) are satisfied. Consequently, we analyze the quasi-static regime. Furthermore, we specialize to the case where the cold bath consists of multiple identical two-level systems, each of which are described by a Hamiltonian with energy gap E.

Firstly, we identify characteristics that ε should have, such that near perfect work is extracted in the limit $\beta_f \to \beta_{\mathrm{Cold}}$ (i.e. when (A) and (B) are satisfied). We then show two technical results:

1. The choice of ε (as a function of β_f) simplifies the minimization problem in (11), by reducing the range the variable α appearing in the optimization of W_{ext}. Under the consideration of near perfect work, ε can be chosen such that the optimization of α is over $\alpha \geq \kappa$ for some $\kappa \in (0, 1]$, instead of $\alpha \geq 0$. The larger κ is for a chosen ε, the slower $\Delta S/W_{\mathrm{ext}}$ converges to zero.

2. We analyze the following cases separately:

    - For $\Omega \leq 1$, ε can always be chosen such that the infimum in (11) is obtained in the limit $\alpha \to 1$. Evaluating the efficiency in the limit $\alpha \to 1$ corresponds to the C.E..
    - For $\Omega > 1$, we show that for the best choice of ε, the infimum in (11) for W_{ext} is obtained at $\alpha \to \infty$. Furthermore, $\Omega > 1$ means that up to leading order terms, $W_1 > W_\infty$ for W_α defined in (12). But we know that the quantity W_1 gives us C.E.. Therefore, the efficiency is strictly less than Carnot.

The maximum efficiency of nano heat engines depends on more than temperature: Appendix

Contents

In this Supplementary Material, we detail our findings. Sections A-C are aimed at giving the reader an overview of the important concepts regarding heat engines, and to introduce the quantities of interest. Firstly, in Section A we describe the setup of our heat engine, the systems involved, and how work is extracted and stored. By using this general setup, we then proceed in Section B to introduce conditions for thermodynamical state transitions in a cycle of a heat engine. In Section C, we introduce the formal definition of efficiency, and specify how can this quantity be maximized over a set of free parameters (involving the bath Hamiltonian structure).

After providing these guidelines, we start in Section D to apply the macroscopic law of thermodynamics. We have performed the analysis with the generalization of allowing for an arbitrarily small probability of failure. The results in this section might be familiar and known to the reader, however from a technical perspective, their establishment is helpful for proving our main results (in Section E) about nanoscale systems. In Section E, we apply the recently discovered generalizations of the second law for small quantum systems. The results in Section D and Section E are summarized at the beginning of each section, for the reader to have a concise overview of the distinction between thermodynamics of macroscopic and nanoscopic systems. Finally, in Section F, we show that even when considering a more general setup, these results obtained in Section E remain unchanged.

A. The general setting for a heat engine

A heat engine is a procedure for extracting work from a temperature difference. It comprises of four basic elements: two thermal baths at distinct temperatures T_{Hot} and T_{Cold} respectively, a machine, and a battery. The machine interacts with these baths in such a way that utilizes the temperature difference between the two baths to perform work extraction. The extracted work can then be transferred and stored in the battery, while the machine returns to its original state.

In this section, we describe a fully general setup, where all involved systems and changes in energy are accounted for explicitly. Let us begin with the total Hamiltonian

$$\hat{H}_t = \hat{H}_{\text{Cold}} + \hat{H}_{\text{Hot}} + \hat{H}_{\text{M}} + \hat{H}_{\text{W}}, \tag{A1}$$

where the indices Hot, Cold, M, W represent a hot thermal bath (Hot), a cold thermal bath (Cold), a machine (M), and a battery (W) respectively. Let us also consider an initial state

$$\rho^0_{\text{ColdHotMW}} = \tau^0_{\text{Cold}} \otimes \tau^0_{\text{Hot}} \otimes \rho^0_{\text{M}} \otimes \rho^0_{\text{W}}. \tag{A2}$$

The state τ^0_{Hot} (τ^0_{Cold}) is the initial thermal state at temperature T_{Hot} (T_{Cold}), corresponding to the hot (cold) bath Hamiltonians $\hat{H}_{\text{Hot}}, \hat{H}_{\text{Cold}}$. More generally, given any Hamiltonian $\hat{H}$ and temperature T, the thermal state is defined as $\tau = \frac{1}{\text{tr}(e^{-\hat{H}/k_BT})} e^{-\hat{H}/k_BT}$. For notational convenience, we shall often use inverse temperatures, defined as $\beta_h := 1/k_BT_{\text{Hot}}$ and $\beta_c := 1/k_BT_{\text{Cold}}$ where k_B is the Boltzmann constant. Without loss of generality we set $T_{\text{Cold}} < T_{\text{Hot}}$. The initial machine $(\rho^0_{\text{M}}, \hat{H}_{\text{M}})$ can be chosen arbitrarily, as long as its final state is preserved (and therefore the machine acts like a catalyst).

We adopt a resource theory approach, which allows all energy-preserving unitaries $U(t)$ on the global system, i.e. such unitaries obey $[U(t), \hat{H}_{\text{ColdHotMW}}] = 0$. If $(\tau^0_{\text{Hot}}, \hat{H}_{\text{Hot}})$ and $(\rho^0_{\text{M}}, \hat{H}_{\text{M}})$ can be arbitrarily chosen, these correspond to the set of *catalytic thermal operations* [5] one can perform on the joint state ColdW. This implies that the cold bath is used as a resource state. By catalytic thermal operations that act on the cold bath, using the hot bath as a thermal reservoir, and the machine as a catalyst, one can possibly extract work and store it in the battery.

The aim is to achieve a final *reduced* state $\rho^1_{\text{ColdHotMW}}$, such that

$$\rho^1_{\text{ColdMW}} = \text{tr}_{\text{Hot}}(\rho^1_{\text{ColdHotMW}}) = \rho^1_{\text{Cold}} \otimes \rho^1_{\text{M}} \otimes \rho^1_{\text{W}}, \tag{A3}$$

where $\rho^1_{\text{M}} = \rho^0_{\text{M}}$, i.e. the machine is preserved, and ρ^1_{Cold}, ρ^1_{W} are the final states of the cold bath and battery. In Section F, we will consider the case in which there are correlations between the final state of the cold bath, hot bath, battery and or machine. We will find that the correlations do not change our results. For any bipartite state ρ_{AB}, we use the notation of reduced states $\rho_{\text{A}} := \text{tr}_{\text{B}}(\rho_{\text{AB}})$, $\rho_{\text{B}} := \text{tr}_{\text{A}}(\rho_{\text{BA}})$.

Finally, we describe the battery such that the state transformation from $\rho^0_{\text{ColdHotMW}}$ to $\rho^1_{\text{ColdHotMW}}$ stores work in the battery. This is done as follows: consider the battery which has a Hamiltonian (written in its diagonal form)

$$\hat{H}_{\text{W}} := \sum_{i=1}^{n_{\text{W}}} E_i^{\text{W}} |E_i\rangle\langle E_i|_{\text{W}}. \tag{A4}$$

For some parameter $\varepsilon \in [0, 1)$, we consider the initial and final states of the battery to be

$$\rho^0_{\text{W}} = |E_j\rangle\langle E_j|_{\text{W}} \tag{A5}$$

$$\rho^1_{\text{W}} = (1-\varepsilon)|E_k\rangle\langle E_k|_{\text{W}} + \varepsilon|E_j\rangle\langle E_j|_{\text{W}} \tag{A6}$$

respectively. The parameter W_{ext} is defined as the energy difference

$$W_{\text{ext}} := E_k^{\text{W}} - E_j^{\text{W}}. \tag{A7}$$

where we define $E_k^{\text{W}} > E_j^{\text{W}}$ such that $W_{\text{ext}} > 0$. In the case where W_{ext} is a value such that the transition $\rho^0_{\text{ColdW}} \to \rho^1_{\text{ColdW}}$ is possible via catalytic thermal operations, it corresponds to extracting work. We refer to the parameter ε as the probability of failure of work extraction. Note that ε in Eq. (A6) is also the trace distance

$$d(\rho, \sigma) = \frac{1}{2}\|\rho - \sigma\|_1 \tag{A8}$$

between $|E_j\rangle\langle E_j|_{\text{W}}$ and $|E_k\rangle\langle E_k|_{\text{W}}$. In Section F, we will generalize this definition to include *all* final states of the battery ρ^1_{W},

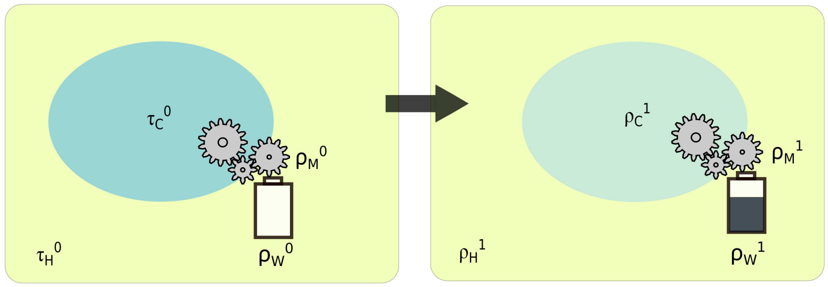

FIG. 5. The setting of a working heat engine.

which are a trace distance ε from the ideal final battery state $|E_k\rangle\langle E_k|_{\mathrm{W}}$. We show that our findings regarding the achievability of C.E. remains unchanged.

Throughout our analysis, we deal with two distinct scenarios of work extraction as defined below.

Definition 1. *(Perfect work) An amount of work extracted W_{ext} is referred to as* perfect work *when $\varepsilon = 0$.*

The next definition of work involves a condition regarding the *von Neumann entropy* of the final battery state. Let ΔS be the von Neumann entropy of the final battery state. When the initial state ρ^0_{W} is pure, we have

$$\Delta S := -\mathrm{tr}(\rho^1_{\mathrm{W}} \ln \rho^1_{\mathrm{W}}). \tag{A9}$$

When the final battery state is given by Eq. (A6), its probability distribution has its support on a two-dimensional subspace of the battery system, this definition also coincides with the binary entropy of ε,

$$\mathrm{h}_2(\varepsilon) = -\varepsilon \ln \varepsilon - (1-\varepsilon)\ln(1-\varepsilon) = \Delta S. \tag{A10}$$

Definition 2. *(Near perfect work) An amount of work extracted W_{ext} is referred to as* near perfect work *when*

1) $0 < \varepsilon \leq l$, for some fixed $l < 1$ and

2) $0 < \dfrac{\Delta S}{W_{\text{ext}}} < p$ for any $p > 0$, i.e. $\dfrac{\Delta S}{W_{\text{ext}}}$ is arbitrarily small.

In the main text, we have provided a detailed discussion regarding the physical meaning of perfect work and near perfect work, and the necessity for considering these quantities. As we will see later in the proof to Lemma 5, 1) and 2) in Def. 2 are both satisfied if and only if

$$\lim_{\varepsilon\to 0^+} \frac{\Delta S}{W_{\text{ext}}} = 0. \tag{A11}$$

Since the initial state $\rho^0_{\mathrm{ColdHotMW}}$ is diagonal in the energy eigenbasis, and since catalytic thermal operations do not create coherences between energy eigenstates, therefore ρ^1_{ColdMW} has to be diagonal in the energy eigenbasis. Furthermore, (as already stated above) in Section F, we extend the setup to include correlation in the final state between the battery, cold bath and machine and more general final battery states.

Note that in our model we allow the battery to have arbitrarily many (but finite) eigenvalues. One can compare this to the two-dimensional battery used in [5], referred to as the wit. Having a minimal dimension, the wit is a conceptually very useful tool to visualize work extraction. However, it has the disadvantage that the energy spacing, i.e. the amount of work to be extracted, has to be known a priori to the work being extracted in order to tune the energy gap of the wit. The more general battery, which we describe in Eq. (A4), requires a higher system dimension, but has the advantage that it can form a quasi-continuum and thus effectively any amount of work (i.e. any $W_{\text{ext}} > 0$) can be stored in it without prior knowledge of the work extraction process. We will see that our results are independent of $n_{\mathrm{W}} \geq 2$.

To summarize, so far we have made the following minimal assumptions:

(A.1) Product state: There are no initial nor final correlations between the cold bath, machine and battery. Initial correlations we assume do not exist, since each of the initial systems are brought independently into the process. This is an advantage of our setup, since if one assumed initial coherence, one would then have to use unknown resources to generate them in the first place. We later also show that correlations between the final cold bath and battery do not provide improvements in maximum extractable work or efficiency.

(A.2) Perfect cyclicity: The machine undergoes a cyclic process, i.e. $\rho^0_{\rm M} = \rho^1_{\rm M}$.

(A.3) Isolated quantum system: The heat engine as a whole, is isolated from and does not interact with the world. This assumption ensures that all possible resources in a work extraction process has been accounted for.

(A.4) Finite dimension: The Hilbert space associated with $\rho^0_{\rm ColdHotMW}$ is finite dimensional but can be arbitrarily large. Moreover, the Hamiltonians $\hat{H}_{\rm Cold}$, $\hat{H}_{\rm Hot}$, $\hat{H}_{\rm M}$ and $\hat{H}_{\rm W}$ all have bounded pure point spectra, meaning that these Hamiltonians have eigenvalues which are bounded.

After defining the set of allowed operations, and describing the desired state transformation process, one can then ask: what conditions should be fulfilled such that there exists a hot bath $(\tau^0_{\rm Hot}, \hat{H}_{\rm Hot})$, and a machine $(\rho^0_{\rm M}, \hat{H}_{\rm M})$ such that $\rho^0_{\rm ColdW} \rightarrow \rho^1_{\rm ColdW}$ is possible? Throughout this document we use "$\rightarrow$" to denote a state transition via catalytic thermal operations.

In Section D, by assuming the macroscopic law of thermodynamics governs the heat engine, we derive the efficiency of a heat engine, and verify the long known Carnot efficiency as the optimal efficiency. We do this for both cases where $\varepsilon = 0$ and when ε is arbitrarily small. In Section E, we analyze the same problem under recently derived second laws, which hold for small quantum systems. We show that these new second laws lead to fundamental differences to the efficiency of a heat engine.

Throughout our analysis, a particular notion that describes thermodynamical transitions will be important towards achieving maximum efficiency. We therefore define this technical term, which will be used throughout the manuscript.

Definition 3. *(Quasi-static) A heat engine is* quasi-static *if the final state of the cold bath is a thermal state and its inverse temperature β_f only differs infinitesimally from the initial cold bath temperature, i.e. $\beta_f = \beta_c - g$, where $0 < g \ll 1$.*

Since throughout this analysis we frequently deal with arbitrarily small paramaters ε, g, we also introduce beforehand the notation of order function $\Theta(x)$, $o(x)$, which denotes the growth of a function.

Definition 4. *(Big Θ, small o notation [28]) Consider two real-valued functions $P(x), Q(x)$. We say that*
1. $P(x) = \Theta(Q(x))$ in the limit $x \to a$ iff there exists $c_1, c_2 > 0$ and $\delta > 0$ such that for all $|x-a| \leq \delta$, $c_1 \leq \left|\frac{P(x)}{Q(x)}\right| \leq c_2$.
2. $P(x) = o(Q(x))$ in the limit $x \to a$ iff there exists $c_3 \geq 0$ such that $\displaystyle\lim_{x\to a} \left|\frac{P(x)}{Q(x)}\right| = c_3$.

Remark 1. *In Def.4, if the limit of x is unspecified, by default we take $a = 0$. In [28], these order terms were only defined for $x \to \infty$. However, choosing a general limit $x \to a$ can be done by simply defining the variable $x' = 1/(x-a)$, and $x \to a^+$ is the same as taking $x' \to \infty$.*

We also list a few properties of these functions here for $x \to 0$, which will help us throughout the proof:
a) For any $c \neq 0$, $\Theta(c \cdot P(x)) = \Theta(P(x))$.
b) For any functions $P_1(x)$ and $P_2(x)$, $\Theta(P_1(x)) + \Theta(P_2(x)) = \Theta\left(\max\{|P_1(x)|, |P_2(x)|\}\right)$.
c) For any functions $P_1(x)$ and $P_2(x)$, $\Theta(P_1(x)) \cdot \Theta(P_2(x)) = \Theta(P_1(x)P_2(x))$.
d) For any functions $P_1(x)$ and $P_2(x)$, $\Theta(P_1(x))/\Theta(P_2(x)) = \Theta(P_1(x)/P_2(x))$.

Definition 3 has two direct implications for a quasi-static heat engine:

(i) The temperature of the final state of the cold bath T_f, only increases w.r.t. its initial temperature by an infinitesimal amount, i.e. $T_f = T_{\rm Cold} + T^2_{\rm Cold}\, g + \Theta(g^2)$.

(ii) The amount of work extracted is infinitesimal: as we shall see later, the extractable perfect and near perfect work $W_{\rm ext} > 0$ (see Defs. (1), (2)) is of order $\Theta(g)$. This follows from using Eq. (D5) for the case where $\rho^1_{\rm Cold}$ is a thermal state with inverse temperature $\beta_f = \beta_c - g$, and calculating the Taylor expansion of $W_{\rm ext}$ about $g = 0$.

B. The conditions for thermodynamical state transitions

In this section, we briefly state the laws which govern the transitions from initial, $\rho^0_{\rm ColdHotMW}$ to final, $\rho^1_{\rm ColdHotMW}$ states for one cycle of our heat engine. By applying these laws, the amount of extractable work $W_{\rm ext}$ can be quantified and expressed as a function of the cold bath.

1. Second law for macroscopic systems

The cold bath, machine and battery form a *closed* but not isolated thermodynamic system. This means only heat exchange (and not mass exchange) occurs between these systems and the hot bath. Therefore, a transition from ρ^0_{ColdMW} to ρ^1_{ColdMW} will be possible if and only if the Helmholtz free energy, F does not increase

$$F(\rho^0_{\text{ColdMW}}) \geq F(\rho^1_{\text{ColdMW}}), \tag{B1}$$

where

$$F(\rho) := \langle \hat{H} \rangle_\rho - \frac{1}{\beta} S(\rho), \tag{B2}$$

and $S(\rho) := -\text{tr}(\rho \ln \rho)$ and $\langle \hat{H} \rangle_\rho := \text{tr}(\hat{H}\rho)$ being the entropy and the mean energy of state ρ respectively. Throughout the manuscript, whenever the state is a thermal state at temperature β, we use the shorthand notation $\langle \hat{H}_{\text{Cold}} \rangle_\beta$ and $S(\beta)$.

The Helmholtz free energy bears a close relation to the *relative entropy*,

$$D(\rho \| \sigma) = \text{tr}(\rho \ln \rho) - \text{tr}(\rho \ln \sigma). \tag{B3}$$

Whenever ρ and σ are diagonal in the same basis, the relative entropy can be written as

$$D(\rho \| \sigma) = \sum_i p_i \ln \frac{p_i}{q_i}, \tag{B4}$$

where p_i, q_i are the eigenvalues of ρ and σ respectively. Now, for any Hamiltonian $\hat{H}$, consider $\tau_\beta = \mathrm{e}^{-\beta \hat{H}}/Z_\beta$, which is the thermal state at some inverse temperature β, with partition function $Z_\beta = \text{tr}[\mathrm{e}^{-\beta \hat{H}}]$, and denote its eigenvalues as q_i. Then for any diagonal state ρ with eigenvalues p_i, and denoting $\{E_i\}_i$ as the eigenvalues of $\hat{H}$,

$$D(\rho \| \tau_\beta) = \sum_i p_i \ln \frac{p_i}{q_i} = -S(\rho) + \sum_i p_i (\beta E_i + \ln Z_\beta) = \beta F(\rho) + \ln Z_\beta. \tag{B5}$$

This implies that

$$F(\rho) = \frac{1}{\beta} [D(\rho \| \tau_\beta) - \ln Z_\beta]. \tag{B6}$$

In Section D we will solve Eq. (B1) in order to evaluate the maximum efficiency.

2. Second laws for nanoscopic systems

In the microscopic quantum regime, where only a few quantum particles are involved, it has been shown that macroscopic thermodynamics is not a complete description of thermodynamical transitions. More precisely, not only the Helmholtz free energy, but a whole other family of generalized free energies have to decrease during a state transition [5]. This places further constraints on whether a particular transition is allowed. In particular, these laws also give necessary and sufficient conditions, when a system with initial state ρ^0_{ColdW} can be transformed to final state ρ^1_{ColdW} (both diagonal in the energy eigenbasis), with the help of any catalyst/machine which is returned to its initial state after the process.

We can apply these second laws to our scenario by associating the catalyst with ρ^0_{M}, and considering the state transition $\rho^0_{\text{W}} \otimes \tau^0_{\text{Cold}} \rightarrow \rho^1_{\text{W}} \otimes \rho^1_{\text{Cold}}$ as described in Section A. Note that the initial state $\rho^0_{\text{W}} \otimes \tau^0_{\text{Cold}}$ is block-diagonal in the energy eigenbasis (for the battery by our choice, and for the cold bath because it is a thermal state). By catalytic thermal operations, the final state is also block-diagonal in the energy eigenbasis. Furthermore, according to the second laws in [5], the transition from $\rho^0_{\text{W}} \otimes \tau^0_{\text{Cold}} \rightarrow \rho^1_{\text{W}} \otimes \rho^1_{\text{Cold}}$ is then possible iff

$$F_\alpha(\tau^0_{\text{Cold}} \otimes \rho^0_{\text{W}}, \tau^h_{\text{ColdW}}) \geq F_\alpha(\rho^1_{\text{Cold}} \otimes \rho^1_{\text{W}}, \tau^h_{\text{ColdW}}) \quad \forall \alpha \geq 0, \tag{B7}$$

where τ^h_{ColdW} is the thermal state of the system at temperature T_{Hot} of the surrounding bath. The quantity $F_\alpha(\rho, \sigma)$ for $\alpha \geq 0$

corresponds to a family of free energies defined in [5], which can be written in the form

$$F_\alpha(\rho,\tau) = \frac{1}{\beta_h}\left[D_\alpha(\rho\|\tau) - \ln Z_h\right], \tag{B8}$$

where $D_\alpha(\rho\|\tau)$ are known as α-Rényi divergences. Sometimes we will use the short hand $F_\infty := \lim_{\alpha\to\infty} F_\alpha$. On occasion, we will refer to a particular transition as being possible/impossible according to the F_α free energy constraint. By this, we mean that for that particular value of α and transition, Eq. (B7) is satisfied/not satisfied. The α-Rényi divergences can be defined for arbitrary quantum states, giving us necessary (but insufficient) second laws for state transitions [5, 10]. However, since we are analyzing states which are diagonal in the same eigenbasis, the Rényi divergences can be simplified to

$$D_\alpha(\rho\|\tau) = \frac{1}{\alpha-1}\ln\sum_i p_i^\alpha q_i^{1-\alpha}, \tag{B9}$$

where p_i, q_i are the eigenvalues of ρ and the state τ. The cases $\alpha=0$ and $\alpha\to 1$ are defined by continuity, namely

$$D_0(\rho\|\tau) = \lim_{\alpha\to 0^+} D_\alpha(\rho\|\tau) = -\ln\sum_{i:p_i\neq 0} q_i, \tag{B10}$$

$$D_1(\rho\|\tau) = \lim_{\alpha\to 1} D_\alpha(\rho\|\tau) = \sum_i p_i \ln\frac{p_i}{q_i}, \tag{B11}$$

and we also define D_∞ as

$$D_\infty(\rho\|\tau) = \lim_{\alpha\to\infty^+} D_\alpha(\rho\|\tau) = \ln\max_i \frac{p_i}{q_i}. \tag{B12}$$

The quantity $D_1(\rho\|\tau)$ coincides with $D(\rho\|\tau)$, as we have defined in Eq. (B3), and evaluated in Eq. (B4) for diagonal states. We will often use this convention. Furthermore, since we are considering initial states which are block-diagonal in the energy eigenbasis, these generalized second laws are both necessary and *sufficient* conditions for state transformations. Therefore, in Section E 2 a we will solve Eq. (B7) explicitly to find an expression for W_{ext} with the ultimate goal of evaluating the maximum efficiency in this regime.

The reader should note that for both Section B 1 and B 2, the conditions for state transformation place upper bounds on the quantity W_{ext}. In particular, this allows us to express the maximum values W_{ext} can take (such that the joint state transformation of cold bath and battery is possible) in terms of quantities related to the cold bath, and the error probability ε. It is also worth comparing the conditions for state transformation in Section B 1 and B 2, which are stated in Eqs. (B1) and (B7). In particular, Eq. (B1) is but a particular instance of Eq. (B7), and therefore the nanoscopic second laws always place a stronger upper bound on W_{ext} compared to the macroscopic second law.

C. Efficiency, maximum efficiency and how to evaluate it

The central quantity of interest in this letter is the efficiency of heat engines. Since we have already introduced the notion of a heat engine in Section A, and the rules which govern the possibility of thermodynamical transitions of one cycle of a heat engine in Section B, it is timely to define the efficiency. After defining this quantity, we demonstrate how go about calculating its maximum value under different conditions, such as for perfect work, near perfect work, in both the macroscopic and nanoscopic regimes. This will prepare the scene for Sections D and E, where we evaluate the maximum efficiency more explicitly.

1. Definition of efficiency and maximum efficiency

As stated in the main text, the efficiency of a particular heat engine (recall that a heat engine is defined by its initial and final states $\rho^0_{\text{ColdHotMW}}$, $\rho^1_{\text{ColdHotMW}}$ as described in Section A) is defined as

$$\eta := \frac{W_{\text{ext}}}{\Delta H}, \tag{C1}$$

where W_{ext} is the amount of work extracted which is defined in Eq. (A7), and ΔH is the amount of mean energy drawn from the hot bath, namely $\Delta H := \text{tr}(\hat{H}_{\text{Hot}}\rho^0_{\text{Hot}}) - \text{tr}(\hat{H}_{\text{Hot}}\rho^1_{\text{Hot}})$, where ρ^1_{Hot} is the reduced state of the hot bath.

Now, consider the set of conditions on state transformations given by Eq. (B7) for nanoscale systems. As discussed in Section

B, these conditions place a restriction on the range of values W_{ext} can take. Therefore, for any fixed ρ^1_{Cold}, we define $\eta^{\text{nano}}(\rho^1_{\text{Cold}})$ as the maximum achievable efficiency as a function of the final state of the cold bath. More precisely,

$$\eta^{\text{nano}}(\rho^1_{\text{Cold}}) \tag{C2}$$
$$= \sup_{W_{\text{ext}}} \eta(\rho^1_{\text{Cold}}) \quad \text{subject to} \quad F_\alpha(\rho^0_{\text{W}} \otimes \tau^0_{\text{Cold}}, \tau^h_{\text{ColdW}}) \geq F_\alpha(\rho^1_{\text{W}} \otimes \rho^1_{\text{Cold}}, \tau^h_{\text{ColdW}}) \quad \forall \alpha \geq 0. \tag{C3}$$

In Eq. (C2), we have written the quantity in Eq. (C1) as $\eta = \eta(\rho^1_{\text{Cold}})$ to remind ourselves of its explicit final cold bath state dependency. Therefore, the maximum efficiency will correspond to maximizing over the final state of the cold bath:

$$\eta_{\max} = \sup_{\rho^1_{\text{Cold}} \in \mathcal{S}} \eta^{\text{nano}}(\rho^1_{\text{Cold}}), \tag{C4}$$

where $\mathcal{S}$ is the space of all quantum states in $\mathcal{H}_{\text{Cold}}$. By analyzing this quantity in Section E, we show that perfect work cannot be extracted. Therefore, when we calculate the maximization in Eq. (C4) we will consider near perfect work (see Def. 2).

In the macro regime, we have to satisfy a less stringent requirement, namely the macroscopic second law of thermodynamics. And hence we have that for fixed ρ^1_{Cold}, $\eta^{\text{mac}}(\rho^1_{\text{Cold}})$ is the maximum efficiency as a function of ρ^1_{Cold}

$$\eta^{\text{mac}}(\rho^1_{\text{Cold}}) = \sup_{W_{\text{ext}}} \eta(\rho^1_{\text{Cold}}) \quad \text{subject to} \quad F(\rho^0_{\text{ColdMW}}) \geq F(\rho^1_{\text{ColdMW}}) \tag{C5}$$
$$\text{and} \quad \text{tr}(\hat{H}_t \rho^0_{\text{ColdHotMW}}) = \text{tr}(\hat{H}_t \rho^1_{\text{ColdHotMW}}), \tag{C6}$$

where $\hat{H}_{\text{ColdHotMW}}$ is defined in Eq. (A1). Similarly to the nanoscale setting, the maximum efficiency is

$$\eta_{\max} = \sup_{\rho^1_{\text{Cold}} \in \mathcal{S}} \eta^{\text{mac}}(\rho^1_{\text{Cold}}). \tag{C7}$$

We can also define the maximum quasi-static efficiencies for the macro and nano scale. The maximum efficiency of a quasi-static heat engine (see Def. 3), is

$$\eta^{\text{stat}}_{\max} = \lim_{g \to 0^+} \eta^{\text{nano}}(\tau(g)), \tag{C8}$$
$$\eta^{\text{stat}}_{\max} = \lim_{g \to 0^+} \eta^{\text{mac}}(\tau(g)), \tag{C9}$$

for the *nanoscopic* and *macroscopic* cases respectively. $\tau(g) \in \mathcal{H}_{\text{Cold}}$ is the thermal state with Hamiltonian $\hat{H}_{\text{Cold}}$ at temperature $\beta_f = \beta_c - g$ and $\eta^{\text{nano}}, \eta^{\text{mac}}$ are defined in Eqs. (C2) and (C5) respectively. Since we can extract perfect and near perfect work in the macroscopic setting, we will derive the efficiency for both cases in Section D.

2. Finding a simplified expression for the efficiency

We can find a more useful expression for ΔH appearing in Eq. (C1). This can be obtained by observing that since only energy preserving operations are allowed, we have

$$\text{tr}(\hat{H}_t \rho^0_{\text{ColdHotMW}}) = \text{tr}(\hat{H}_t \rho^1_{\text{ColdHotMW}}), \tag{C10}$$

where $\hat{H}_t = \hat{H}_{\text{Hot}} + \hat{H}_{\text{Cold}} + \hat{H}_{\text{M}} + \hat{H}_{\text{W}}$. Since the Hamiltonian does not contain interaction terms between these systems, the mean energy depends only on the *reduced states* of each system. Mathematically, it means that Eq. (C10) can be written as

$$\text{tr}(\hat{H}_{\text{Hot}} \rho^0_{\text{Hot}}) + \text{tr}(\hat{H}_{\text{Cold}} \rho^0_{\text{Cold}}) + \text{tr}(\hat{H}_{\text{M}} \rho^0_{\text{M}}) + \text{tr}(\hat{H}_{\text{W}} \rho^0_{\text{W}}) = \tag{C11}$$
$$\text{tr}(\hat{H}_{\text{Hot}} \rho^1_{\text{Hot}}) + \text{tr}(\hat{H}_{\text{Cold}} \rho^1_{\text{Cold}}) + \text{tr}(\hat{H}_{\text{M}} \rho^1_{\text{M}}) + \text{tr}(\hat{H}_{\text{W}} \rho^1_{\text{W}}). \tag{C12}$$

Also, note that since $\rho^0_{\text{M}} = \rho^1_{\text{M}}$, therefore $\text{tr}(\hat{H}_{\text{M}} \rho^0_{\text{M}}) = \text{tr}(\hat{H}_{\text{M}} \rho^1_{\text{M}})$. This implies that we have

$$\Delta H = \Delta C + \Delta W, \tag{C13}$$

where

$$\Delta C := \text{tr}\left[\hat{H}_{\text{Cold}}\rho^1_{\text{Cold}}\right] - \text{tr}\left[\hat{H}_{\text{Cold}}\tau^c_{\text{Cold}}\right], \tag{C14}$$

and

$$\Delta W := \text{tr}(\hat{H}_{\text{W}}\rho^1_{\text{W}}) - \text{tr}(\hat{H}_{\text{W}}\rho^0_{\text{W}}). \tag{C15}$$

are the change in average energy of the cold bath and battery. We can thus write Eq. (C1) as

$$\eta = \frac{W_{\text{ext}}}{\Delta W + \Delta C}. \tag{C16}$$

Furthermore, from Eqs. (A5), (A6), (A7) and (C15), we have $\Delta W = (1-\varepsilon)W_{\text{ext}}$, and hence we can write the inverse efficiency as

$$\eta^{-1}(\rho^1_{\text{Cold}}) = 1 - \varepsilon + \frac{\Delta C(\rho^1_{\text{Cold}})}{W_{\text{ext}}(\rho^1_{\text{Cold}})}, \tag{C17}$$

where we have made explicit the ρ^1_{Cold} dependency. We already know from the setting that ρ^0_{Cold} is thermal. If ρ^1_{Cold} is also a thermal state at some temperature β according to the cold bath Hamiltonian $\hat{H}_{\text{Cold}}$, we will sometimes use the shorthand notation $\eta(\beta)$ for $\eta(\rho^1_{\text{Cold}})$ and $\Delta W(\beta)$, $\Delta C(\beta)$ for $\Delta W(\rho^1_{\text{Cold}})$, $\Delta C(\rho^1_{\text{Cold}})$ respectively.

In Section D, we will derive an expression for W_{ext} and solve Eqs. (C5), (C7). In Section E, we will derive a new expression for W_{ext} in the nanoscopic regime, and solve Eqs. (C2), (C4).

D. Efficiency of a heat engine according to macroscopic thermodynamics

In this section, we study the efficiency of the setup detailed in Section A under the constraints of macroscopic thermodynamics, as described in Section B 1. This implies that the Helmholtz free energy solely dictates whether $\rho^0_{\text{ColdW}} \rightarrow \rho^1_{\text{ColdW}}$ is possible. We find that in both cases of extracting perfect and near perfect work,

(1) The maximum achievable efficiency is the Carnot efficiency.

(2) The Carnot efficiency can be achieved for any cold bath Hamiltonian.

(3) The Carnot efficiency is only achieved when the final state of the cold bath is thermal (according to a different temperature T_f).

(4) The Carnot efficiency is only achieved for quasi-static heat engines, meaning in the limit where $T_f \gtrsim T_{\text{Cold}}$. A technical definition of quasi-static heat engines can be found in Def. 3. Roughly speaking, this means that there is only infinitesimal change in the final temperature of the cold bath, compared to its original state.

This section can be summarized as follows: in Section D 1, we first apply the macroscopic law of thermodynamics, namely the fact that Helmholtz free energy is non-increasing, to our heat engine setup. By making use of energy conservation, we can derive the amount of maximum extractable work as shown in Eq. (D4). Next, in Section D 2 we show that when considering the extraction of perfect work, we show the points (1)-(4) as stated above. In Section D 3, we show that points (1)-(4) hold also when considering near perfect work.

The main results can be found in Theorem 1 and Lemma 6. One may think points (1)-(4) are obvious since it has long been known that the optimal achievable efficiency of a heat engine operating between two thermal baths is the Carnot efficiency, and that this efficiency can only be achieved quasi-statically. The motivations for proving these results here are two-fold. Firstly, this is a rigorous and mathematical proof of optimality, while usually one encounters arguments such as reversibility, or that the heat engine must remain in thermal equilibrium at all times during the working of the heat engine. Secondly, we will find later on at the nano/quantum scale that the Carnot efficiency can be achieved but observation (2) does not hold anymore. For these reasons, it is worthwhile proving that one can actually achieve points (1)-(4) in this setting for any cold bath Hamiltonian according to macroscopic thermodynamics. From a practical point of view, many of the technical results proved here will be needed in the proofs of Section E, where we derive results involving a more refined set of generalized free energies, which describes thermodynamic transitions for nanoscale quantum systems.

1. Maximum extractable work according to macroscopic law of thermodynamics

Our first task is to find an expression for W_{ext} in the macro regime. We do so by solving Eq. (B1) for W_{ext} such that

$$\langle \hat{H}_{\text{ColdMW}} \rangle_{\rho^1_{\text{ColdMW}}} - \frac{1}{\beta_h} S(\rho^1_{\text{ColdMW}}) \leq \langle \hat{H}_{\text{ColdMW}} \rangle_{\rho^0_{\text{ColdMW}}} - \frac{1}{\beta_h} S(\rho^0_{\text{ColdMW}}). \tag{D1}$$

The entropy is an additive quantity under tensor product, meaning that $S(\rho_1 \otimes \rho_2) = S(\rho_1) + S(\rho_2)$ for any states ρ_1, ρ_2. Furthermore, since the joint Hamiltonian does not contain interaction terms, therefore the mean energy also depends only on the reduced states. In summary, both S and $\langle \hat{H} \rangle$ are additive under a tensor product structure of ρ^0_{ColdMW} and ρ^1_{ColdMW} as described in Eqs. (A2) and (A3). This means one can rewrite Eq. (D1) by expanding its terms,

$$\begin{aligned} \langle \hat{H}_{\text{Cold}} \rangle_{\rho^1_{\text{Cold}}} + \langle \hat{H}_{\text{M}} \rangle_{\rho^1_{\text{M}}} + \langle \hat{H}_{\text{W}} \rangle_{\rho^1_{\text{W}}} - \frac{1}{\beta_h} \left[S(\rho^1_{\text{Cold}}) + S(\rho^1_{\text{M}}) + S(\rho^1_{\text{W}}) \right] \leq \\ \langle \hat{H}_{\text{Cold}} \rangle_{\rho^0_{\text{Cold}}} + \langle \hat{H}_{\text{M}} \rangle_{\rho^0_{\text{M}}} + \langle \hat{H}_{\text{W}} \rangle_{\rho^0_{\text{W}}} - \frac{1}{\beta_h} \left[S(\rho^0_{\text{Cold}}) + S(\rho^0_{\text{M}}) + S(\rho^0_{\text{W}}) \right], \end{aligned} \tag{D2}$$

Furthermore, note that $\rho^0_{\text{M}} = \rho^1_{\text{M}}$, and therefore $S(\rho^0_{\text{M}}), \langle \hat{H}_{\text{M}} \rangle_{\rho^0_{\text{M}}}$ are common terms on both sides of Eq. (D2) which can be cancelled out. Furthermore, by our construction of the battery in Eqs. (A4)-(A7), we have that $S(\rho^0_{\text{W}}) = 0$, $S(\rho^1_{\text{W}}) = \Delta S = \text{h}_2(\varepsilon)$ and $\langle \hat{H}_{\text{W}} \rangle_{\rho^0_{\text{W}}} = E^{\text{W}}_j$ and $\langle \hat{H}_{\text{W}} \rangle_{\rho^1_{\text{W}}} = E^{\text{W}}_k$. Thus, Eq. (D2) can be simplified to

$$W_{\text{ext}} + \langle \hat{H}_{\text{Cold}} \rangle_{\rho^1_{\text{Cold}}} - \frac{1}{\beta_h} S(\rho^1_{\text{Cold}}) \leq \langle \hat{H}_{\text{Cold}} \rangle_{\rho^0_{\text{Cold}}} - \frac{1}{\beta_h} S(\rho^0_{\text{Cold}}) + \frac{1}{\beta_h} \text{h}_2(\varepsilon), \tag{D3}$$

where W_{ext} has been defined in Eq. (A7). In other words, $W_{\text{ext}} \leq F(\rho^0_{\text{Cold}}) - F(\rho^1_{\text{Cold}}) + \frac{1}{\beta_h} \text{h}_2(\varepsilon)$.

We can also express W_{ext} with the *relative entropy* instead, by using Eq. (B6). We can apply this identity to Eq. (D3) whenever the initial and final states are diagonal in the energy eigenbasis. Note that the initial ρ^0_{Cold} is a thermal state (of some temperature), and therefore diagonal in the energy eigenbasis. Since we start with a state $\tau^0_{\text{Cold}} \otimes \rho^0_{\text{W}}$ which is diagonal w.r.t. the Hamiltonian, and since catalytic thermal operations can never increase coherences between energy eigenstates (or in the macro setting, since we only demand mean energy conservation), we know that the final state $\rho^1_{\text{Cold}} \otimes \rho^1_{\text{W}}$ is also diagonal in the energy eigenbasis. Therefore, Eq. (D3) can be rewritten w.r.t. the relative entropies as follows

$$W_{\text{ext}} \leq F(\rho^0_{\text{Cold}}) - F(\rho^1_{\text{Cold}}) + \frac{1}{\beta_h} \text{h}_2(\varepsilon) = \frac{1}{\beta_h} \left[D(\rho^0_{\text{Cold}} \| \tau^h_{\text{Cold}}) - D(\rho^1_{\text{Cold}} \| \tau^h_{\text{Cold}}) + \text{h}_2(\varepsilon) \right]. \tag{D4}$$

2. Maximum efficiency for perfect work is Carnot efficiency

In this section, we want to find the maximum efficiency according to Eqs. (C1), (C5) and (C7), for the case of $\varepsilon = 0$ which implies $\text{h}_2(\varepsilon) = 0$. We do this by the following steps:

1. *Evalaute* W_{ext}. According to Eq. (D4), we know that

$$W_{\text{ext}} = F(\rho^0_{\text{Cold}}) - F(\rho^1_{\text{Cold}}) = \frac{1}{\beta_h} \left[D(\rho^0_{\text{Cold}} \| \tau^h_{\text{Cold}}) - D(\rho^1_{\text{Cold}} \| \tau^h_{\text{Cold}}) \right], \tag{D5}$$

where recall that we have defined τ^h_{Cold} previously as the thermal state of system C with temperature T_{Hot}. Note that here equality can be achieved because in macroscopic thermodynamics, satisfying the free energy constraint is a necessary and sufficient condition for the possibility of a state transformation. Note that since by construction the initial and final states of the battery are pure energy eigenstates, namely $\varepsilon = 0$ and therefore

$$W_{\text{ext}} = \Delta W. \tag{D6}$$

2. *Write inverse maximum efficiency as optimization problem.* By substituting the simplified expression for efficiency derived in Eq. (C17) into Eq. (C7), we have

$$\eta^{-1}_{\max} = \inf_{\rho^1_{\text{Cold}}} (\eta^{\text{mac}})^{-1} = 1 + \inf_{\rho^1_{\text{Cold}}} \frac{\Delta C}{W_{\text{ext}}}. \tag{D7}$$

3. *Maximize W_{ext} given a fixed value of ΔC.* This is done in Lemma 1, where we show that given a fixed ΔC, the final cold bath state that maximizes W_{ext} is uniquely a thermal state, corresponding to a certain inverse temperature β'.

4. *Show that 3) implies that efficiency is maximized by a thermal state of the cold bath.* This is proven in Lemma 2. Therefore, this implies one only needs to optimize Eq. (D7) over one variable, i.e. β_f, the final temperature of the cold bath.

5. *Show that the efficiency is strictly increasing with β_f.* This is done first by proving several identities, which are summarized in Corollary 1. Using these identities, we prove in Lemma 4 that the first derivative of efficiency w.r.t. β_f is always positive over the range where $W_{\text{ext}} > 0$. This leads us to conclude, in Theorem 1, that maximum efficiency is achieved in the limit $\beta_f \to \beta_c$, and evaluating the efficiency at this limit gives us the Carnot efficiency.

Firstly, let us develop a technical Lemma 1, which concerns the unique solution towards maximizing W_{ext} for a fixed ΔC. By applying Lemma 1, we show in Lemma 2 that the maximal efficiency is achieved when ρ^1_{Cold} is a thermal state. The reader can easily find similar proofs in [29].

Lemma 1. *Given any Hamiltonian $\hat{H}_{\text{Cold}}$, a corresponding thermal state τ^h_{Cold} of some temperature β_h, and a fixed initial state ρ^0_{Cold}, consider the maximization over final states ρ^1_{Cold},*

$$\max_{\rho^1_{\text{Cold}}} W_{\text{ext}} \tag{D8}$$

over all states ρ^1_{Cold} which are diagonal in the energy eigenbasis, subject to the constraint that ΔC is a constant. Then the solution for ρ^1_{Cold} is unique, and ρ is a thermal state according to the Hamiltonian $\hat{H}_{\text{Cold}}$ at a certain temperature β'.

Proof. Firstly, from Eq. (C14) we see that the constraint ΔC being a constant, is the same as $\mathrm{tr}\left[\hat{H}_{\text{Cold}}\rho^1_{\text{Cold}}\right]$ being a constant. This is because they differ only by a constant term. On the other hand, from Eq. (C15) and (D6), we can see that Eq. (D8) is equal to

$$\max_{\rho^1_{\text{Cold}}} W_{\text{ext}} = \frac{1}{\beta_h}\left[D(\rho^0_{\text{Cold}}\|\tau^h_{\text{Cold}}) - \min_{\rho^1_{\text{Cold}}} D(\rho^1_{\text{Cold}}\|\tau^h_{\text{Cold}})\right]. \tag{D9}$$

Since ρ^1_{Cold} and τ are both diagonal in the energy eigenbasis (ρ^1_{Cold} by the statement in the lemma, and τ by it being a thermal state), one can evaluate the relative entropy by using Eq. (B3). Denote the eigenvalues of our variable ρ^1_{Cold} to be $\{p_i\}_i$, and the eigenvalues of the thermal state τ to be $\{q_i\}_i$. We can then write the optimization problem as

$$\min_{\{p_i\}} \sum_i p_i(\ln p_i - \ln q_i); \quad \text{subject to} \sum_i p_i E_i = c \quad \text{constant, and} \sum_i p_i = 1.$$
$$\text{where} \quad q_i = \frac{e^{-\beta E_i}}{Z_\beta}; \quad Z_\beta = \sum_i e^{-\beta E_i}.$$

We can now employ techniques of Lagrange multipliers to solve this optimization. The constrained Lagrange equation is

$$L(\{p_i\},\lambda) = \sum_i p_i(\ln p_i - \ln q_i) + \lambda\left(\sum_i E_i p_i - c\right) + \mu\left(\sum_i p_i - 1\right), \tag{D10}$$

$$\frac{dL}{dp_i} = (\ln p_i - \ln q_i + 1 + \lambda E_i + \mu) = 0, \tag{D11}$$

$$\frac{dL}{d\lambda} = \sum_i E_i p_i - c = 0. \tag{D12}$$

$$\frac{dL}{d\mu} = \sum_i p_i - 1 = 0. \tag{D13}$$

We find that the normalized solution is

$$p_i = \frac{e^{-\beta' E_i}}{Z_{\beta'}}, \quad Z_{\beta'} = e^{(1+\mu)Z_\beta}, \tag{D14}$$

and p_i are probabilities corresponding to the Boltzmann distribution, according to inverse temperature $\beta' = \beta + \lambda$. Depending on the mean energy constraint c and normalization condition, one can solve for the Lagrange multipliers λ and μ. With this we conclude that the state ρ which maximizes $D(\rho^1_{\text{Cold}}\|\tau)$ is a thermal state, where its temperature is such that the constraint on mean energy is satisfied. □

Lemma 2. *Consider the work extraction process described by the state transformation $\rho^0_{\text{ColdMW}} \to \rho^1_{\text{ColdMW}}$, where ρ^0_{Cold}, ρ^0_{W} and ρ^1_{W} have been described in Section A. Denote $\mathcal{H}_{\text{Cold}}$ as the Hilbert space of the cold bath. Then the maximal efficiency in Eq. (D7) is obtained for a final state of the cold bath ρ^1_{Cold}, which is thermal:*

$$\eta^{-1}_{\max} = 1 + \inf_{\rho^1_{\text{Cold}}\in\mathcal{S}_\tau} \frac{\Delta C}{W_{\text{ext}}}, \tag{D15}$$

where $\mathcal{S}_\tau$ the set of all thermal states (for $\hat{H}_{\text{Cold}}$ with any temperature $T > 0$) in $\mathcal{H}_{\text{Cold}}$. Furthermore, all non-thermal states do not achieve the maximum efficiency, i.e.

$$\eta^{-1}_{\max} < 1 + \left.\frac{\Delta C}{W_{\text{ext}}}\right|_{\rho^1_{\text{Cold}}} \qquad \textit{for any } \rho^1_{\text{Cold}} \in \mathcal{S}\setminus\mathcal{S}_\tau. \tag{D16}$$

where $\mathcal{S}$ is the space of all quantum states in $\mathcal{H}_{\text{Cold}}$

Proof. First of all, note that without loss of generality we can always consider only diagonal states, as explained in the paragraph before Eq. (D4) that catalytic thermal operations do not increase coherences between energy eigenstates. We begin by substituting Eqs. (C14) and (D5) into Eq. (D7), and finding

$$\eta^{-1}_{\max} = 1 + \inf_{\rho^1_{\text{Cold}}} \frac{\Delta C}{W_{\text{ext}}} \tag{D17}$$

$$= 1 + \inf_{\rho^1_{\text{Cold}}} \frac{\beta_h \Delta C}{D_1(\tau^c_{\text{Cold}}\|\tau^h_{\text{Cold}}) - D_1(\rho^1_{\text{Cold}}\|\tau^h_{\text{Cold}})} \tag{D18}$$

$$= 1 + \beta_h \left[\sup_{\rho^1_{\text{Cold}}} \frac{D_1(\tau^c_{\text{Cold}}\|\tau^h_{\text{Cold}}) - D_1(\rho^1_{\text{Cold}}\|\tau^h_{\text{Cold}})}{\text{tr}(\hat{H}_{\text{Cold}}\rho^1_{\text{Cold}}) - \text{tr}(\hat{H}_{\text{Cold}}\tau^c_{\text{Cold}})}\right]^{-1}. \tag{D19}$$

In the last line of Eq. (D19), we see that only two terms depend on the maximization variable ρ^1_{Cold}. This means we can perform the maximization in two steps:

$$\sup_{\rho^1_{\text{Cold}}} \frac{D_1(\tau^c_{\text{Cold}}\|\tau^h_{\text{Cold}}) - D_1(\rho^1_{\text{Cold}}\|\tau^h_{\text{Cold}})}{\text{tr}(\hat{H}_{\text{Cold}}\rho^1_{\text{Cold}}) - \text{tr}(\hat{H}_{\text{Cold}}\tau^c_{\text{Cold}})} = \sup_{A>0} \frac{D_1(\tau^c_{\text{Cold}}\|\tau^h_{\text{Cold}}) - B(A)}{A} \tag{D20}$$

where $B(A)$ is the optimal value of a separate minimization problem:

$$B(A) = \inf_{\substack{\rho^1_{\text{Cold}}\in\mathcal{S} \\ tr(H_{\text{Cold}}\rho^1_{\text{Cold}}) - \text{tr}(\hat{H}_{\text{Cold}}\tau^c_{\text{Cold}}) = A}} D_1(\rho^1_{\text{Cold}}\|\tau^h_{\text{Cold}}) \tag{D21}$$

From Lemma 1, we know that the solution of the sub-minimization problem in Eq. (D21) has a unique form, namely $\rho^1_{\text{Cold}} = \tau^f_{\text{Cold}}$ is a thermal state of some temperature β_f. Therefore, Eq. (D20) can be simplified to

$$\sup_{\rho^1_{\text{Cold}}} \frac{D_1(\tau^c_{\text{Cold}}\|\tau^h_{\text{Cold}}) - D_1(\rho^1_{\text{Cold}}\|\tau^h_{\text{Cold}})}{\text{tr}(\hat{H}_{\text{Cold}}\rho^1_{\text{Cold}}) - \text{tr}(\hat{H}_{\text{Cold}}\tau^c_{\text{Cold}})} = \sup_{\beta_f} \frac{D_1(\tau^c_{\text{Cold}}\|\tau^h_{\text{Cold}}) - D_1(\tau^f_{\text{Cold}}\|\tau^h_{\text{Cold}})}{\text{tr}(\hat{H}_{\text{Cold}}\tau^f_{\text{Cold}}) - \text{tr}(\hat{H}_{\text{Cold}}\tau^c_{\text{Cold}})}. \tag{D22}$$

Whats more, for every constant A, the function

$$f(x) = \left(1 + \beta_h \left[\frac{D_1(\tau^c_{\text{Cold}}\|\tau^h_{\text{Cold}}) - x}{A}\right]^{-1}\right)^{-1} \tag{D23}$$

is bijective in $x \in \mathbb{R}$ and thus due to the uniqueness of the sub-minimization problem in Eq. (D21), we conclude that for all non-thermal states ρ^1_{Cold}, the corresponding efficiency will be strictly less than that of Eq. (D19). Thus from Eq. (D22) and (D19) we conclude the lemma. □

After establishing Lemma 2, we can continue to solve the optimization problem in Eq. (D7) by only looking at final states which are thermal (according to some final temperature β_f which we optimize over). In the next Lemma 3 and Corollary 1, we derive some useful and interesting identities. These identities concern quantities such as the derivatives of mean energy and entropy of the thermal state (with respect to inverse temperature), and relates them to the variance of energy. We later use them in Lemma 4 to prove that the Carnot efficiency can only be achieved for quasi-static heat engines. The reader can find similar proofs in any standard thermodynamic textbook (For example in Sections 6.5, 6. of [23]), but we derive them here for completeness.

Lemma 3. *For any cold bath Hamiltonian $\hat{H}_{\text{Cold}}$, consider the thermal state $\tau_\beta = \frac{1}{Z_\beta} e^{-\beta \hat{H}_{\text{Cold}}}$ with inverse temperature β. Define $\langle \hat{H}_{\text{Cold}} \rangle_\beta = \operatorname{tr}(\hat{H}_{\text{Cold}} \tau_\beta)$, and $S(\beta) = -\tau_\beta \ln \tau_\beta$ to be the mean energy and entropy of τ_β. Then the following identities hold:*

$$\frac{d\langle \hat{H}_{\text{Cold}} \rangle_\beta}{d\beta} = -\operatorname{var}(\hat{H}_{\text{Cold}})_\beta \tag{D24}$$

$$\frac{dS(\beta)}{d\beta} = -\beta \cdot \operatorname{var}(\hat{H}_{\text{Cold}})_\beta, \tag{D25}$$

where $\operatorname{var}(\hat{H}_{\text{Cold}})_\beta = \langle \hat{H}^2_{\text{Cold}} \rangle_\beta - \langle \hat{H}_{\text{Cold}} \rangle^2_\beta$ is the variance of energy for τ_β.

Proof. Intuitively we know that the expectation value of energy increases as temperature increases (or as the inverse temperature decreases). More precisely, consider the probabilities of τ_β for each energy level of the Hamiltonian E_i,

$$p_i = \frac{e^{-\beta E_i}}{Z_\beta}, \quad \text{where } Z_\beta = \sum_i e^{-\beta E_i}$$

$$\frac{dp_i}{d\beta} = \frac{1}{Z_\beta^2}\left[-E_i e^{-\beta E_i} \cdot Z_\beta - \frac{dZ_\beta}{d\beta} \cdot e^{-\beta E_i}\right] = -p_i E_i - \frac{1}{Z_\beta}\frac{dZ_\beta}{d\beta} p_i = -p_i E_i + p_i \langle \hat{H}_{\text{Cold}} \rangle_\beta. \tag{D26}$$

The last equality holds because of the following identity:

$$\frac{-1}{Z}\frac{dZ}{d\beta} = \frac{-1}{Z}\sum_i (-E_i) e^{-\beta E_i} = \sum_i p_i E_i = \langle \hat{H}_{\text{Cold}} \rangle_\beta. \tag{D27}$$

Therefore, we have

$$\frac{d\langle \hat{H}_{\text{Cold}} \rangle_\beta}{d\beta} = \sum_i \frac{d\langle \hat{H}_{\text{Cold}} \rangle_\beta}{dp_i}\frac{dp_i}{d\beta} = \sum_i E_i \cdot \left[-p_i E_i + p_i \langle \hat{H}_{\text{Cold}} \rangle_\beta\right] \tag{D28}$$

$$= -\langle \hat{H}^2_{\text{Cold}} \rangle_\beta + \langle \hat{H}_{\text{Cold}} \rangle^2_\beta = -\operatorname{var}(\hat{H}_{\text{Cold}})_\beta. \tag{D29}$$

On the other hand, similarly, one can prove the second identity by writing down the expression of entropy for the thermal state,

$$S(\beta) = -\sum_i \frac{e^{-\beta E_i}}{Z_\beta} \ln \frac{e^{-\beta E_i}}{Z_\beta} = \sum_i \beta E_i \frac{e^{-\beta E_i}}{Z_\beta} + \ln Z_\beta \sum_i \frac{e^{-\beta E_i}}{Z_\beta} = \beta \langle \hat{H}_{\text{Cold}} \rangle_\beta + \ln Z_\beta. \tag{D30}$$

Therefore, the derivative of $S(\beta)$ w.r.t. β is

$$\frac{dS(\tau_\beta)}{d\beta} = \langle \hat{H}_{\text{Cold}} \rangle_\beta + \beta \frac{d\langle \hat{H}_{\text{Cold}} \rangle_\beta}{d\beta} + \frac{1}{Z_\beta}\frac{dZ_\beta}{d\beta} = \beta \cdot \frac{d\langle \hat{H}_{\text{Cold}} \rangle_\beta}{d\beta} = -\beta \cdot \operatorname{var}(\hat{H}_{\text{Cold}})_\beta. \tag{D31}$$

□

By using Lemma 3 in a special case, we obtain the following corollary:

Corollary 1. *Given any Hamiltonian $\hat{H}_{\text{Cold}}$, consider the quantities*

$$\Delta C(\beta_f) = \operatorname{tr}(\hat{H}_{\text{Cold}} \tau_{\beta_f}) - \operatorname{tr}(\hat{H}_{\text{Cold}} \tau_{\beta_c}) = \langle \hat{H}_{\text{Cold}} \rangle_{\beta_f} - \langle \hat{H}_{\text{Cold}} \rangle_{\beta_c} \tag{D32}$$

and

$$W_{\text{ext}}(\beta_f) = F(\tau_{\beta_c}) - F(\tau_{\beta_f}) = \frac{1}{\beta_h}\left[D(\tau_{\beta_c} \| \tau_{\beta_h}) - D(\tau_{\beta_f} \| \tau_{\beta_h})\right], \tag{D33}$$

where τ_β corresponds to the thermal state defined by $\hat{H}_{\text{Cold}}$ at inverse temperature β. Then

$$\frac{d\Delta C(\beta_f)}{d\beta_f} = -\text{var}(\hat{H}_{\text{Cold}})_{\beta_f} \tag{D34}$$

$$\frac{dW_{\text{ext}}(\beta_f)}{d\beta_f} = \frac{\beta_h - \beta_f}{\beta_h}\text{var}(\hat{H}_{\text{Cold}})_{\beta_f}. \tag{D35}$$

Proof. For $\Delta C(\beta_f)$, it is straightforward from Lemma 3 that

$$\frac{d\Delta C(\beta_f)}{d\beta_f} = \frac{d\langle \hat{H}_{\text{Cold}}\rangle_{\beta_f}}{d\beta_f} = -\text{var}(\hat{H}_{\text{Cold}})_{\beta_f}. \tag{D36}$$

On the other hand, $\Delta W(\beta_f)$ can be simplified by substituting Eq. (B6) into Eq. (D33),

$$W_{\text{ext}}(\beta_f) = F(\tau_{\beta_c}) - F(\tau_{\beta_f}) = \langle \hat{H}_{\text{Cold}}\rangle_{\beta_c} - \langle \hat{H}_{\text{Cold}}\rangle_{\beta_f} - \frac{1}{\beta_h}\left[S(\tau_{\beta_c}) - S(\tau_{\beta_f})\right]. \tag{D37}$$

With this, we can evaluate the derivative

$$\begin{aligned}\frac{dW_{\text{ext}}(\beta_f)}{d\beta_f} &= -\frac{d\langle \hat{H}_{\text{Cold}}\rangle_{\beta_f}}{d\beta_f} + \frac{1}{\beta_h}\frac{dS(\tau_{\beta_f})}{d\beta_f} \\ &= \text{var}(\hat{H}_{\text{Cold}})_{\beta_f} - \frac{\beta_f}{\beta_h}\text{var}(\hat{H}_{\text{Cold}})_{\beta_f} \\ &= \frac{\beta_h - \beta_f}{\beta_h}\text{var}(\hat{H}_{\text{Cold}})_{\beta_f}.\end{aligned}$$

The second equality is obtained by Lemma 3 for $\frac{d\langle \hat{H}_{\text{Cold}}\rangle_{\beta_f}}{d\beta_f}$, and the third by grouping common factors together. $\square$

In the next step, by using Corollary 1, we show that the optimal efficiency is achieved only in the quasi-static limit, i.e. in the limit $\beta_f \to \beta_c$.

Lemma 4. *Evaluate the efficiency expressed in Eq. (C17) for the situation where the final state of the cold bath is a thermal state at inverse temperature β_f:*

$$\eta(\beta_f) = \frac{W_{\text{ext}}(\beta_f)}{\Delta C(\beta_f) + W_{\text{ext}}(\beta_f)}. \tag{D38}$$

Then for all $\beta_f < \beta_c$, $\frac{d\eta(\beta_f)}{d\beta_f} > 0$.

Proof. To prove this, we show that $\frac{d\eta^{-1}}{d\beta_f} < 0$, where $\eta^{-1} = 1 + \frac{\Delta C}{W_{\text{ext}}}$. Evaluating the derivative of η^{-1} w.r.t. β_f, we obtain

$$\frac{d\eta^{-1}}{d\beta_f} = \frac{1}{W_{\text{ext}}^2}\cdot\left[\frac{d\Delta C(\beta_f)}{d\beta_f}W_{\text{ext}} - \frac{dW_{\text{ext}}(\beta_f)}{d\beta_f}\Delta C\right] \tag{D39}$$

$$= \frac{\text{var}(\hat{H}_{\text{Cold}})_{\beta_f}}{W_{\text{ext}}^2}\cdot\left[-W_{\text{ext}} - \frac{\beta_h - \beta_f}{\beta_h}\Delta C\right] \tag{D40}$$

$$= \frac{\text{var}(\hat{H}_{\text{Cold}})_{\beta_f}}{W_{\text{ext}}^2}\cdot\left[\Delta C + \frac{1}{\beta_h}\left[S(\tau_{\beta_c}) - S(\tau_{\beta_f})\right] - \frac{\beta_h - \beta_f}{\beta_h}\Delta C\right] \tag{D41}$$

$$= \frac{\text{var}(\hat{H}_{\text{Cold}})_{\beta_f}}{W_{\text{ext}}^2}\frac{\beta_f}{\beta_h}\cdot\left[\Delta C - \frac{1}{\beta_f}[S(\tau_{\beta_f}) - S(\tau_{\beta_c})]\right]. \tag{D42}$$

(D43)

The first equality is obtained by invoking the chain rule of differentiation. The second equality is obtained by substituting $\frac{dW_{\text{ext}}}{d\beta_f}'$, $\frac{d\Delta C}{d\beta_f}$, as evaluated earlier in Corollary 1. The third equality is obtained by expressing W_{ext} according to Eq. (D37), plus recognizing that $\langle \hat{H}_{\text{Cold}}\rangle_{\tau_{\beta_f}} - \langle \hat{H}_{\text{Cold}}\rangle_{\tau_{\beta_c}} = \Delta C$. The last inequality is obtained, simply by taking out a common term β_f/β_h.

We then make the following observations:
1) The factor

$$\frac{\beta_f}{\beta_h W_{\text{ext}}^2} > 0, \tag{D44}$$

2) The variance of energy for any positive temperature

$$\text{var}(\hat{H}_{\text{Cold}})_{\beta_f} > 0, \tag{D45}$$

3) and the last term $\Delta C - \frac{1}{\beta_f}[S(\tau_{\beta_f}) - S(\tau_{\beta_c})]$ can be written as $F(\tau_{\beta_f}) - F(\tau_{\beta_c})$, where F is the free energy of a system w.r.t. a bath with inverse temperature β_f. But then, since τ_{β_f} is the thermal state with the same inverse temperature, this means that τ_{β_f} is the *unique* state that minimizes free energy. Therefore, $F(\tau_{\beta_c}) - F(\tau_{\beta_f}) > 0$ for any τ_{β_c}.

□

From Lemma 2 and Lemma 4, we conclude that the maximization of efficiency for any Hamiltonian $\hat{H}$ happens for a final state which is thermal, and the greater its inverse temperature β_f, the higher efficiency is. With these lemmas we can now prove the main result of this section (Theorem 1).

In the next theorem, we evaluate the efficiency at the limit $\beta_f \to \beta_c^-$, and show that it corresponds to the Carnot efficiency.

Theorem 1 (Carnot Efficiency)**.** *Consider all heat engines which extract perfect work (see Definition 1). Then according to the macroscopic second law of thermodynamics, the maximum achievable efficiency (see Eq. (C7)) is the Carnot efficiency*

$$\eta_{\max} = 1 - \frac{\beta_h}{\beta_c}. \tag{D46}$$

It can be obtained for all *cold bath Hamiltonians* $\hat{H}_{\text{Cold}}$*, but* only *for quasi-static heat engines (as defined in Def. 3 and Eq. (C9) for quasi-static maximum efficiency), where an infinitesimal amount of work is extracted.*

Proof. From Eq. (C7), we have an expression for the optimal efficiency in terms of a maximization over final cold bath states $\rho^1_{\text{Cold}} \in \mathcal{S}$. By Lemma 2, we know that the optimal solution is obtained only for thermal states. Subsequently, by Lemma 4, it is shown that when the final cold bath is of temperature β_f, the corresponding efficiency is strictly increasing w.r.t. β_f. Also note that since by definition $W_{\text{ext}} > 0$, this implies that $\beta_f < \beta_c$. Intuitively, this is because heat cannot flow from a cold to hot system without any work input. One can also see this mathematically, by showing that for any $\beta \geq \beta_h$,

$$\frac{dF(\tau_\beta)}{d\beta} = \frac{d}{d\beta}\left[\langle \hat{H}_{\text{Cold}}\rangle_\beta - \frac{1}{\beta_h}S(\beta)\right] = \left(\frac{\beta}{\beta_h} - 1\right)\text{var}(\hat{H}_{\text{Cold}})_\beta \geq 0. \tag{D47}$$

This implies that if $\beta_f \geq \beta_c \geq \beta_h$, then $F(\beta_f) \geq F(\beta_c)$, and according to Eq. (D33) $W_{\text{ext}} \leq 0$. Therefore, the optimal efficiency must be achieved only when the ρ^1_{Cold} is a thermal state whose inverse temperature β_f approaches β_c from below. Let $\beta_f = \beta_c - g$, where $g > 0$. Then Thus we have

$$\eta_{\max}^{-1} = \lim_{g\to 0^+} (\eta^{\text{mac}})^{-1}(\beta_c - g), \quad (\eta^{\text{mac}})^{-1}(\beta_c - g) = 1 + \frac{\Delta C}{W_{\text{ext}}}\bigg|_{\rho^1_{\text{Cold}} = \tau_{(\beta_c - g)}}. \tag{D48}$$

Since as $g \to 0^+$, both the numerator and denominator vanish, we can evaluate this limit by first applying L'Hôspital rule, the chain rule for derivatives (for any function F, $\frac{dF}{dg} = -\frac{dF}{d\beta_f}$), and then Corollary 1 to obtain

$$\lim_{g\to 0^+} \frac{\Delta C}{W_{\text{ext}}} = \lim_{g\to 0^+} \frac{\frac{d\Delta C}{dg}}{\frac{dW_{\text{ext}}}{dg}} = \lim_{\beta_f \to \beta_c^-} \frac{\frac{d\Delta C}{d\beta_f}}{\frac{dW_{\text{ext}}}{d\beta_f}} = \frac{\beta_h}{\beta_c - \beta_h}.$$

This implies that

$$\eta_{\max}^{-1} = \lim_{g\to 0^+} (\eta^{\text{mac}})^{-1}(\beta_c - g) = 1 + \frac{\beta_h}{\beta_c - \beta_h} = \frac{\beta_c}{\beta_c - \beta_h} \tag{D49}$$

and hence $\eta_{\max} = 1 - \frac{\beta_h}{\beta_c}$.

□

3. Maximum efficiency for near perfect work is still Carnot efficiency

In this section, we show that even while allowing a non-zero failure probability $\varepsilon > 0$ in the near perfect work scenario, the maximum achievable efficiency is still the Carnot efficiency. It is worth noting that this result is also important later, as an upper bound to maximum efficiency in the nanoscopic regime. We first prove it in Lemma 5 for the case where the final state of the battery is fixed as in Eq. (A6). Then later, we show in Lemma 6 that Carnot is still the maximum, even if we allow a more general final battery state. Before we present the proof, it is useful for the reader to recall the definition of near perfect work (Def. 2) and quasi-static heat engines (Def. 3).

Lemma 5. *Consider all heat engines which extract near perfect work (see Def. 2). Then according to the macroscopic second law of thermodyanmics, the maximum efficiency of a heat engine, $\eta_{\rm max}$ is the Carnot efficiency*

$$\eta_{\rm max} = \sup_{\rho^1_{\rm C}\in\mathcal{S}} \eta^{\rm mac}(\rho^1_{\rm C}) = 1 - \frac{\beta_h}{\beta_c}, \tag{D50}$$

and the supremum can only *be achieved for quasi-static heat engines (see Def. (3) and Eq. (C9)).*

Proof. The ideas in this proof are very similar to that of Section D 2, and the main complication comes from proving that even if we allow $\varepsilon > 0$, as long as $\Delta S/W_{\rm ext}$ is arbitrarily small, the maximum efficiency cannot surpass the Carnot efficiency.

Let us begin by establishing the relevant quantities for near perfect work extraction. The amount of work extractable from the heat engine, when we have a probability of failure, according to the standard free energy can be obtained by solving Eq. (D5). We thus have that the maximum $W_{\rm ext}$ is

$$W_{\rm ext} = \beta_h^{-1}(1-\varepsilon)^{-1}\left[D(\tau_{\beta_c}\|\tau_{\beta_h}) - D(\rho^1_{\rm C}\|\tau_{\beta_h}) + \Delta S\right], \tag{D51}$$

where ΔS is defined in Eq. (A9).

Before we continue with the analysis, we will note a trivial consequence of Eq. (D51). Condition 1) in Def 2 implies that $(1-\varepsilon)^{-1}$ is upper bounded. The terms in square brackets in Eq. (D51) are also clearly upper bounded for finite β_c, β_h. Hence $W_{\rm ext}$ is bounded from above. ΔS is solely a function of ε and only approaches zero in the limits $\varepsilon\to 0^+$, $\varepsilon \to 1^-$; and $\varepsilon\to 1^-$ is forbidden by 1) in Def 2. Thus if 1) and 2) in Def 2 are satisfied,

$$\lim_{\varepsilon\to 0^+} \frac{\Delta S}{W_{\rm ext}} = 0. \tag{D52}$$

In turn, if Eq. (D52) is satisfied, then we have near perfect work by Def. 2. Thus Eq. (D52) is satisfied iff we have near perfect work. We will use this result later in the proof.

Extracting a positive amount of near perfect work implies that we can rule out all states $\rho^1_{\rm C}$ such that $D(\tau_{\beta_c}\|\tau_{\beta_h}) \leq D(\rho^1_{\rm C}\|\tau_{\beta_h})$ from the analysis. This can be proven by contradiction: if $D(\tau_{\beta_c}\|\tau_{\beta_h}) \leq D(\rho^1_{\rm C}\|\tau_{\beta_h})$, then from Eq. (D51) $\beta_h W_{\rm ext} \leq \Delta S/(1-\varepsilon)$ and together with 2) in Def 2 this would imply

$$0 < \beta_h(1-\varepsilon) \leq \frac{\Delta S}{W_{\rm ext}} < p. \tag{D53}$$

However, since from 1) Def. 2 we have $\varepsilon \leq l$, Eq. (D53) cannot be satisfied for all $p > 0$, leading to a contradiction.

From Eq. (D7) we have

$$\eta^{-1}_{\rm max} = 1-\varepsilon + \inf_{\rho^1_c\in\mathcal{S}} \frac{\Delta C}{W_{\rm ext}} = (1-\varepsilon)\cdot\left[1 + \frac{\beta_h \Delta C}{D(\tau_{\beta_c}\|\tau_{\beta_h}) - D(\rho^1_{\rm C}\|\tau_{\beta_h}) + \Delta S}\right], \tag{D54}$$

where $\Delta C = \Delta C(\rho^1_{\rm C})$ and is defined in Eq. (C14).

Firstly, let us show that with a similar analysis as shown in Lemma 2, the maximum efficiency occurs when $\rho^1_{\rm C}$ is a thermal state. From Eq. (D54), we have

$$\eta^{-1}_{\rm max} = (1-\varepsilon)\left[1 + \beta_h \inf_{\rho^1_{\rm C}\in\mathcal{S}} \frac{\Delta C}{D(\tau_{\beta_c}\|\tau_{\beta_h}) - D(\rho^1_{\rm C}\|\tau_{\beta_h}) + \Delta S}\right] \tag{D55}$$

$$= (1-\varepsilon)\left[1 + \beta_h \inf_{A>0} \frac{A}{D(\tau_{\beta_c}\|\tau_{\beta_h}) - B(A) + \Delta S}\right] \tag{D56}$$

where

$$B(A) = \inf_{\substack{\rho_{\rm C}^1 \in \mathcal{S} \\ tr(\hat{H}_{\rm Cold}\rho_{\rm C}^1) - {\rm tr}(\hat{H}_{\rm Cold}\tau_{\beta_c}) = A}} D(\rho_{\rm C}^1 \| \tau_{\beta_h}). \tag{D57}$$

We can split this minimization problem to Eqs. (D56) and (D57) because $D(\tau_{\beta_c}\|\tau_{\beta_h})$ and ΔS do not depend on the variable $\rho_{\rm C}^1$. Furthermore, when $\rho_{\rm Cold}^1$ is a thermal state of inverse temperature β_f, we have seen in the beginning of the proof in Theorem 1 that for $W_{\rm ext} > 0$, $\beta_f < \beta_c$. This implies that the variable $A = \Delta C = {\rm tr}(\hat{H}_{\rm Cold}\tau_{\beta_f}) - {\rm tr}(\hat{H}_{\rm Cold}\tau_{\beta_c}) > 0$.

By Lemma 1, for any fixed $A > 0$ we conclude that the infimum in Eq. (D57) is achieved *uniquely* when $\rho_{\rm C}^1$ is a thermal state. Therefore, our optimization problem is simplified to optimization over final temperatures β_f (or $g = \beta_c - \beta_f$),

$$\eta_{\max}^{-1} = (1-\varepsilon)\cdot\left[1 + \beta_h \inf_{\substack{\beta_f \\ \Delta C > 0}} \frac{\Delta C}{D(\tau_{\beta_c}\|\tau_{\beta_h}) - D(\tau_{\beta_f}\|\tau_{\beta_h}) + \Delta S}\right] \tag{D58}$$

Consider cases of β_f, where $D(\tau_{\beta_c}\|\tau_{\beta_h}) - D(\tau_{\beta_f}\|\tau_{\beta_h})$ is non-vanishing (finite), which are *non quasi-static*. Note that this always corresponds to extracting near perfect work, since when $\varepsilon \to 0^+$, we have $\varepsilon, \Delta S \to 0$ and these contributions dissapear from Eq. (D58). However, by Lemma 2 we also know that the infimum over β_f occurs uniquely at the quasi-static limit, when $g \to 0^+$. This means that for all non quasi-static cases, Carnot efficiency cannot be achieved.

What remains, is then to consider the quasi-static heat engine, namely the limit $g \to 0^+$. Extracting near perfect work in this case corresponds to requiring that $\lim_{g\to 0^+} \frac{\Delta S}{W_{\rm ext}} = 0$, where $\varepsilon = \varepsilon(g)$ and $\lim_{g\to 0^+}\varepsilon(g) = 0$. Equivalently

$$\lim_{g\to 0^+} \frac{W_{\rm ext}}{\Delta S} = \infty. \tag{D59}$$

Substituting Eq. (D51) into Eq. (D59),

$$\lim_{g\to 0^+} (1-\varepsilon(g))^{-1}\left[1 + \frac{D(\tau_{\beta_c}\|\tau_{\beta_h}) - D(\tau_{\beta_f}\|\tau_{\beta_h})}{\Delta S}\right] = \infty \tag{D60}$$

which implies that $\lim\limits_{g\to 0^+} \dfrac{D(\tau_{\beta_c}\|\tau_{\beta_h}) - D(\tau_{\beta_f}\|\tau_{\beta_h})}{\Delta S} = \infty$, or equivalently,

$$\lim_{\varepsilon\to 0^+}\lim_{g\to 0^+} \frac{\Delta S}{D(\tau_{\beta_c}\|\tau_{\beta_h}) - D(\tau_{\beta_f}\|\tau_{\beta_h})} = 0. \tag{D61}$$

Finally, we evaluate the inverse efficiency at the quasi-static limit,

$$\eta^{-1} = \lim_{g\to 0^+}(1-\varepsilon(g))\cdot\left[1 + \beta_h \frac{\Delta C}{D(\tau_{\beta_c}\|\tau_{\beta_h}) - D(\tau_{\beta_f}\|\tau_{\beta_h}) + \Delta S}\right] \tag{D62}$$

$$= 1 + \beta_h \lim_{g\to 0^+} \frac{\Delta C}{D(\tau_{\beta_c}\|\tau_{\beta_h}) - D(\tau_{\beta_f}\|\tau_{\beta_h}) + \Delta S} \tag{D63}$$

$$= 1 + \beta_h \lim_{g\to 0^+} \frac{\Delta C}{\left[D(\tau_{\beta_c}\|\tau_{\beta_h}) - D(\tau_{\beta_f}\|\tau_{\beta_h})\right]}\cdot\left(1 + \frac{\Delta S}{D(\tau_{\beta_c}\|\tau_{\beta_h}) - D(\tau_{\beta_f}\|\tau_{\beta_h})}\right)^{-1} \tag{D64}$$

$$= 1 + \beta_h \lim_{g\to 0^+} \frac{d\Delta C(\tau_{\beta_f})/dg}{dD(\tau_{\beta_f}\|\tau_{\beta_h})/dg} \tag{D65}$$

$$= 1 - \frac{\beta_h}{\beta_h - \beta_c}, \tag{D66}$$

where from Eq. (D64) to (D65), we make use of Eq. (D61) : the second term within the limit is simply 1, and the first term depends only on g, which we can obtain Eq. (D65) by invoking the L'Hôspital rule. The last equality in Eq. (D66) follows directly from the identities we derived for $\frac{dW_{\text{ext}}}{d\beta_f}$ and $\frac{d\Delta C}{d\beta_f}$ in Corollary 1,

$$\frac{d\Delta C}{dg} = -\frac{d\Delta C}{d\beta_f} = -\text{var}(\hat{H}_{\text{Cold}})_{\beta_f} \tag{D67}$$

$$\frac{dD(\tau_{\beta_f}\|\tau_{\beta_h})}{dg} = -\frac{dD(\tau_{\beta_f}\|\tau_{\beta_h})}{d\beta_f} = \beta_h \frac{dW_{\text{ext}}}{d\beta_f} = (\beta_h - \beta_f)\text{var}(\hat{H}_{\text{Cold}})_{\beta_f}, \tag{D68}$$

while in the limit $g \to 0,\ \beta_f = \beta_c$.

Finally, we now see that the quasi-static efficiency is

$$\eta = \left(\frac{\beta_h - \beta_c - \beta_h}{\beta_h - \beta_c}\right)^{-1} = \frac{\beta_c - \beta_h}{\beta_c} = 1 - \frac{\beta_h}{\beta_c} \tag{D69}$$

which is exactly the Carnot efficiency. □

Later, in Section F 2 we will need Lemma 2 to hold in a more general scenario, i.e. instead of the final battery state being $\rho_{\text{W}}^1 = (1-\varepsilon)\,|E_k\rangle\langle E_k|_{\text{W}} + \varepsilon\,|E_j\rangle\langle E_j|_{\text{W}}$, we want to allow the final battery state to be any energy block-diagonal state with trace distance ε. Next we state and prove this generalized lemma.

Lemma 6. *Consider all heat engines which extract near perfect work (see Definition 2), but allowing for any final battery state with a trace distance ε to the ideal final pure state $|E_k\rangle\langle E_k|_{\text{W}}$. Then according to the macroscopic second law of thermodynamics, the maximum efficiency of a heat engine, η_{max} is the Carnot efficiency*

$$\eta_{\text{max}} = \sup_{\rho_{\text{C}}^1 \in \mathcal{S}} \eta^{\text{mac}}(\rho_{\text{C}}^1) = 1 - \frac{\beta_h}{\beta_c}, \tag{D70}$$

and the supremum is only *achieved for quasi-static heat engines (see Def.* (3) *and Eq.* (C9)*).*

Proof. Firstly, let us note that since the initial state ρ_{ColdW}^0 we start out with is energy block-diagonal, the final state has to also be block-diagonal. Therefore, given the product structure between the cold bath and battery, it is sufficient to consider the case when the final battery state is energy block-diagonal. Next, let us note that any final state ρ_{W}^2 which is energy block-diagonal, and has trace distance ε with $|E_k\rangle\langle E_k|_{\text{W}}$ can be written as,

$$\rho_{\text{W}}^2 = (1-\varepsilon)\,|E_k\rangle\langle E_k|_{\text{W}} + \varepsilon\rho_{\text{W}}^{\text{junk}}, \text{ where } \rho_{\text{W}}^{\text{junk}} = \sum_i p_i\,|E_i\rangle\langle E_i|_{\text{W}}\,, \quad \sum_i p_i = 1 \text{ and } p_k = 0. \tag{D71}$$

Next, one can calculate W_{ext} given by the standard free energy condition, i.e.

$$F(\tau_{\beta_c}) + F(\rho_{\text{W}}^0) \geq F(\rho_{\text{Cold}}^1) + F(\rho_{\text{W}}^1). \tag{D72}$$

Using the identity $F(\rho) = \text{tr}(\hat{H}\rho) - \beta^{-1}S(\rho)$, we have that

$$F(\tau_{\beta_c}) + E_j \geq F(\rho_{\text{Cold}}^1) + (1-\varepsilon)E_k + \varepsilon\text{tr}(\hat{H}_{\text{W}}\rho_{\text{W}}^{\text{junk}}) - \beta_h^{-1}S(\rho_{\text{W}}^2). \tag{D73}$$

Substituting $W_{\text{ext}} = E_k - E_j$, and rearranging terms, we have

$$(1-\varepsilon)W_{\text{ext}} \leq F(\tau_{\beta_c}) - F(\rho_{\text{Cold}}^1) + \beta_h^{-1}\Delta S - \varepsilon[\text{tr}(\hat{H}_{\text{W}}\rho_{\text{W}}^{\text{junk}}) - E_j]. \tag{D74}$$

Finally, by using the identity (in Eq. (B6)) that $F(\rho) = \beta_h^{-1}[D(\rho\|\tau_{\beta_h}) - \ln Z_{\beta_h}]$, the maximum amount of extractable work is given by

$$W_{\text{ext}} = (1-\varepsilon)^{-1}\beta_h^{-1} \cdot [D(\tau_{\beta_c}\|\tau_{\beta_h}) - D(\rho_{\text{Cold}}^1\|\tau_{\beta_h}) + \Delta S - \varepsilon\tilde{E}], \tag{D75}$$

where $\tilde{E} = \text{tr}(\hat{H}_{\text{W}}\rho_{\text{W}}^{\text{junk}}) - E_j$.

Following the steps in Lemma 5, in particular the derivations in Eq. (D55) and (D56), we have

$$\eta_{\text{max}}^{-1} = (1-\varepsilon) \cdot \left[1 + \beta_h \inf_{\substack{\beta_f \\ \Delta C > 0}} \frac{\Delta C}{D(\tau_{\beta_c}\|\tau_{\beta_h}) - D(\tau_{\beta_f}\|\tau_{\beta_h}) + \Delta S - \varepsilon\tilde{E}}\right]. \tag{D76}$$

To show Eq. (D76) gives the Carnot efficiency, we show that 1) for non quasi-static cases where $\beta_f < \beta_c$, Carnot efficiency is not attained, and 2) in the quasi-static limit, Carnot efficiency is attained.

Let us first consider the case of extracting a non-vanishing amount of near perfect work, i.e. for all cases where $\beta_f < \beta_c$. Then near perfect work, by Def. 2, corresponds to the limit $\varepsilon \to 0$,

$$\eta^{-1} = \lim_{\varepsilon\to 0}\ (1-\varepsilon)\cdot\left[1+\beta_h\frac{\Delta C}{D(\tau_{\beta_c}\|\tau_{\beta_h})-D(\tau_{\beta_f}\|\tau_{\beta_h})+\Delta S-\varepsilon\tilde{E}}\right] \tag{D77}$$

$$= 1+\beta_h\frac{\Delta C}{D(\tau_{\beta_c}\|\tau_{\beta_h})-D(\tau_{\beta_f}\|\tau_{\beta_h})}. \tag{D78}$$

In this limit, all terms involving ε vanish, and the inverse efficiency has the same expression as the efficiency for perfect work. We already know from Lemma 4 that the infimum over β_f cannot be obtained in this regime, since the inverse efficiency is strictly decreasing with β_f.

Therefore, again we are left with analyzing the quasi-static limit for this problem. Following the derivation in Eq. (D64) for the quasi-static limit, we obtain

$$\eta_{\max}^{-1} = 1+\beta_h \lim_{g\to 0^+}\frac{\Delta C}{\left[D(\tau_{\beta_c}\|\tau_{\beta_h})-D(\tau_{\beta_f}\|\tau_{\beta_h})\right]}\cdot\left(1+\frac{\Delta S-\varepsilon\tilde{E}}{D(\tau_{\beta_c}\|\tau_{\beta_h})-D(\tau_{\beta_f}\|\tau_{\beta_h})}\right)^{-1}, \tag{D79}$$

where $\varepsilon=\varepsilon(g)$ and note that requiring near perfect work implies that

$$\lim_{g\to 0^+}\frac{\Delta S}{D(\tau_{\beta_c}\|\tau_{\beta_h})-D(\tau_{\beta_f}\|\tau_{\beta_h})} = 0. \tag{D80}$$

Next, we observe the relationship between ε and ΔS, in the regime where ε is small. Given any $\varepsilon > 0$ denoting the trace distance $d(\rho_{\rm W}^2, |E_k\rangle\langle E_k|_{\rm W}) = \varepsilon$, the smallest amount of entropy that can be produced corresponds to $\Delta S = h_2(\varepsilon)$. This is because if we try to distribute the weight ε over more energy eigenvalues, then by majorization the entropy only increases. But we also know that $\varepsilon \leq h_2(\varepsilon)$ for small values of ε, in particular over the regime $\varepsilon\in[0,\frac{1}{2}]$. Therefore, we have that in this regime, $\varepsilon\leq h_2(\varepsilon)\leq\Delta S$ holds. Therefore, we also know that

$$\lim_{g\to 0^+}\frac{\varepsilon\tilde{E}}{D(\tau_{\beta_c}\|\tau_{\beta_h})-D(\tau_{\beta_f}\|\tau_{\beta_h})} = 0, \tag{D81}$$

where $\varepsilon=\varepsilon(g)$. Plugging Eqns. (D80) and (D81) into Eq. (D79), we have that the quasi-static efficiency is $\eta = 1-\frac{\beta_h}{\beta_c}$. □

E. Efficiency of a nanoscopic quantum heat engine

In this section, we will be applying the conditions for state transitions for nanoscale systems, as detailed in Section B 2. The reader will see that due to these extra constraints from the generalized free energies, the fundamental limitations on efficiency will differ greatly from those observed in Section D.

Firstly, in Section E 1, we show that the extraction of a positive amount of perfect work is *impossible* using the setup. In Section E 2, we show that this can be resolved by considering near perfect work instead. Then we find that

(1) The maximum achievable efficiency is still the Carnot efficiency. This is proven in Section E 2 b.

(2) However, the Carnot efficiency *cannot* be achieved for all cold bath Hamiltonians. **This is our main result, which is stated in Theorem 2, found in Section E 2 f.** The results in Section E 2 d and E 2 e are more technical proofs, that pave the way for deriving this main result.

(3) The Carnot efficiency is only achieved when the final state of the cold bath is thermal (according to a different temperature T_f). This is proven in Section E 2 b.

(4) The Carnot efficiency is only achieved for quasi-static heat engines (see Def. 3), meaning in the limit where $T_f \gtrsim T_{\rm Cold}$. This is proven in Section E 2 b.

1. Impossibility of extracting perfect work

We will first show that with the general setup as described in Section A, no perfect work can ever be extracted. By this we mean that whenever ε as defined in Eq.(A6) equals zero, then for any value of $W_{\rm ext} > 0$, and for any final state $\rho^1_{\rm Cold}$, the transition $|E_j\rangle\langle E_j|_{\rm W} \otimes \tau^0_{\rm Cold} \rightarrow |E_k\rangle\langle E_k|_{\rm W} \otimes \rho^1_{\rm Cold}$ is not possible. Intuitively speaking, this occurs because the cold bath is initially in a state of full rank. Since thermal operations cannot decrease the rank of the system, therefore the final state of the cold bath $\rho^1_{\rm Cold}$ must also be of full rank. By directly solving Eq. (B7), we find that the amount of extractable work satisfies

$$W_{\rm ext} \leq kT_{\rm Hot} \inf_{\alpha\geq 0} \left[D_\alpha(\tau^0_{\rm Cold}\|\tau^h_{\rm Cold}) - D_\alpha(\rho^1_{\rm Cold}\|\tau^h_{\rm Cold})\right], \tag{E1}$$

where $\tau^h_{\rm Cold}$ is the thermal state of the cold bath (according to the cold bath Hamiltonian $\hat{H}_{\rm Cold}$), at temperature $T_{\rm Hot}$ (since the surrounding hot bath is of temperature $T_{\rm Hot}$). However from Eq. (B8), $D_0(\tau^0_{\rm Cold}\|\tau^h_{\rm Cold}) = D_0(\rho^1_{\rm Cold}\|\tau^h_{\rm Cold})$. Therefore according to Eq. (E1), the amount of work extractable satisfies $W_{\rm ext} \leq 0$.

We phrase this with more rigor in the following Lemmas 7 and 8, which proves that for perfect work, $W_{\rm ext} > 0$ is impossible. The proof holds for general initial states $\rho^0_{\rm Cold}$ of full rank, in particular, they need not even be diagonal in the energy eigenbasis.

Lemma 7. *For any $W_{\rm ext} > 0$, consider the Hamiltonian $\hat{H}_{\rm W}$ given by Eq. (A4). Then for any inverse temperature $\beta_h > 0$, the thermal state $\tau_{\rm W} = \frac{1}{{\rm tr}(e^{-\beta_h \hat{H}_{\rm W}})} e^{-\beta_h \hat{H}_{\rm W}}$ satisfies*

$${\rm tr}\left[\left(|E_j\rangle\langle E_j|_{\rm W} - |E_k\rangle\langle E_k|_{\rm W}\right) \tau_{\rm W}\right] > 0. \tag{E2}$$

Proof. Follows directly from the definitions. Since $W_{\rm ext} > 0$, we know that $E^{\rm W}_j < E^{\rm W}_k$. Evaluating the quantity above gives $\frac{1}{{\rm tr}(e^{-\beta_h \hat{H}_{\rm W}})} \cdot \left(e^{-\beta_h E^{\rm W}_j} - e^{-\beta_h E^{\rm W}_k}\right) > 0$. □

Lemma 8. *Consider any general quantum state $\rho^0_{\rm Cold}$ of full rank. Then for any $\rho^1_{\rm Cold}$, the transition from $\rho^0_{\rm Cold} \otimes \rho^0_{\rm W} \rightarrow \rho^1_{\rm Cold} \otimes \rho^1_{\rm W}$ is not possible via catalytic thermal operations if*

$${\rm tr}\left[\left(\Pi_{\rho^0_{\rm W}} - \Pi_{\rho^1_{\rm W}} \right) \tau_{\rm W} \right] > 0, \tag{E3}$$

where Π_ρ is the projector onto the support of state ρ, and $\tau_{\rm W}$ is the thermal state of the working body at the initial hot bath temperature.

Proof. One can show this by invoking the quantum second law for $\alpha = 0$ [5], which says that if $\rho_{\rm in} \rightarrow \rho_{\rm out}$ is possible via catalytic thermal operations, then

$$D_0(\rho_{\rm in}\|\tau) \geq D_0(\rho_{\rm out}\|\tau), \tag{E4}$$

where τ is the thermal state of the system at bath temperature, and

$$D_0(\rho\|\sigma) = \lim_{\alpha\rightarrow 0^+} \frac{1}{\alpha - 1} \ln {\rm tr}[\rho^\alpha \sigma^{1-\alpha}] = -\ln {\rm tr}[\Pi_\rho \sigma], \tag{E5}$$

is defined for arbitrary quantum states ρ, σ. Applying this law with $\rho_{\rm in} = \rho^0_{\rm W} \otimes \rho^0_{\rm Cold}$ and $\rho_{\rm out} = \rho^1_{\rm W} \otimes \rho^1_{\rm Cold}$, we arrive at

$$D_0(\rho^0_{\rm W}\|\tau^h_{\rm W}) - D_0(\rho^1_{\rm W}\|\tau^h_{\rm W}) \geq D_0(\rho^1_{\rm Cold}\|\tau^h_{\rm Cold}) - D_0(\rho^0_{\rm Cold}\|\tau^h_{\rm Cold}), \tag{E6}$$

where $\tau^h_{\rm Cold}$ and $\tau^h_{\rm W}$ are thermal states of the cold bath and battery at the temperature of surrounding hot bath ($T_{\rm Hot}$) respectively. Since $\rho^0_{\rm Cold}$ have full rank, and since $\tau^h_{\rm Cold}$ is normalized, therefore according to Eq. (E5), $D_0(\rho^0_{\rm Cold}\|\tau^h_{\rm Cold}) = 0$. Furthermore, since the $\alpha-$Rényi divergence D_0 is non-negative, therefore the r.h.s. of Eq. (E6) is lower bounded by 0. Thus, we have

$${\rm tr}[(\Pi_{\rho^0_{\rm W}} - \Pi_{\rho^1_{\rm W}})\tau_{\rm W}] \leq 0. \tag{E7}$$

Since this is a necessary condition for state transformations, we arrive at the conclusion that: when Eq. (E7) is violated, state transformations are not possible. But from Lemma 7, any type of perfect work extraction violates Eq. (E7). Therefore, in this setting, perfect work extraction is always impossible.

□

To summarize, Lemma 8 implies that if the initial state of the cold bath is thermal, and therefore of full rank, then any work extraction scheme via thermal operations bringing $\rho^0_{\rm W} = |j\rangle\langle j|_{\rm W}$ to $\rho^1_{\rm W} = |k\rangle\langle k|_{\rm W}$ where $W_{\rm ext} = E^{\rm W}_k - E^{\rm W}_j > 0$ is not possible.

In general, we see that if $\Pi_{\rho_{\rm W}^0} \neq \Pi_{\rho_{\rm W}^1}$, then when transition $\rho_{\rm W}^0$ to $\rho_{\rm W}^1$ is possible, transition $\rho_{\rm W}^1$ to $\rho_{\rm W}^0$ is not. Consequentially, we will have to consider near perfect work at the nano regime.

2. Efficiency for extracting near perfect work

As we have just seen in the previous Section E 1, we cannot extract perfect work. Due to the impossibility result, we consider the relaxation of extracting near perfect work in the nanoscale setting.

- We begin by evaluating the expression for efficiency according to the nanoscopic laws of thermodynamics, given a final state of the cold bath, and comparing it to the expression according to macroscopic laws of thermodynamics. This is done in Sections E 2 a and E 2 b, and the relation between two efficiencies are summarized in Eq. (E14). Since the nanoscopic efficiency is always smaller than the macroscopic efficiency, which attains Carnot efficiency only in the quasi-static limit, it will be possible only to attain Carnot efficiency in the quasi-static limit, when considering nanoscopic laws of thermodynamics.

- We analyze the quasi-static regime, focusing on the special case where the cold bath consists of n qubits. Since the quasi-static limit corresponds to the case of small $g>0$, and ε also has to be arbitrarily small for near perfect work extraction, we perform Taylor expansion of the analytical expressions for $W_{\rm ext}$ and ΔC w.r.t. g and ε. This is done in Section E 2 c.

- In Section E 2 d, we identify how to choose $\varepsilon(g)$ such that it corresponds to drawing near perfect work in the quasi-static limit. We first begin by observing that any continuous function $\varepsilon(g)$ that vanishes in the limit $g\rightarrow 0$ can be characterize with a real-valued parameter $\bar{\kappa}$ that determine how quickly ε goes to zero. This is shown in Lemma 11. In Lemma 12, we show that near perfect work is drawn only if $\bar{\kappa}\in[0,1]$.

- Lemma 12 gives us the analytical expression and minimization range in order to evaluate $W_{\rm ext}$, according to Eq. (E66). In Section E 2 e, we show how one can evaluate this optimization problem, by comparing the stationary points and endpoints of the function $\frac{\alpha B_\alpha}{\alpha-1}$ that gives the leading term in Eq. (E66). Lemma 13 proves a technical property of the first derivative of this function. Using it, we prove in Lemma 14 that one can always choose $\varepsilon(g)$ with some $\bar{\kappa}<1$ such that the infimum of $\frac{\alpha B_\alpha}{\alpha-1}$ is obtained at either $\alpha=\bar{\kappa}$ or $\alpha\rightarrow\infty$.

- Finally, in Section E 2 f, we use the results in Section E 2 e regarding the evaluation of $W_{\rm ext}$ to find the efficiency in the quasi-static limit.

a. An explicit expression for $W_{\rm ext}$

Our first task is to work out an explicit expression for $W_{\rm ext}$ depending on the initial and final states of the cold bath, ε and hot bath (inverse) temperature β_h. Such as expression is found by applying the generalized second laws as detailed in Section B 2.

Lemma 9. *Consider the transition*

$$\tau_{\rm Cold}^0\otimes\rho_{\rm W}^0\rightarrow\rho_{\rm Cold}^1\otimes\rho_{\rm W}^1\quad\text{with}\quad\varepsilon>0. \tag{E8}$$

where $\rho_{\rm W}^0$ and $\rho_{\rm W}^1$ are defined in Eqs. (A5), (A6) respectively. Let $W_{\rm ext}$ denote the maximum *possible value such that Eq. (E8) is possible via catalytic thermal operations, with a thermal bath of inverse temperature β_h. Let $\beta_c>\beta_h$. Then the final state $\rho_{\rm Cold}^1=\sum_i p_i'|E_i\rangle\langle E_i|_{\rm Cold}$ is block-diagonal in the energy eigenbasis, and*

$$W_{\rm ext}=\inf_{\alpha\geq 0}\ W_\alpha, \tag{E9}$$

$$W_\alpha=\frac{1}{\beta_h(\alpha-1)}[\ln(A-\varepsilon^\alpha)-\alpha\ln(1-\varepsilon)], \tag{E10}$$

$$A=\frac{\sum_i p_i^\alpha q_i^{1-\alpha}}{\sum_i p_i'^\alpha q_i^{1-\alpha}}, \tag{E11}$$

where $p_i=\frac{e^{-\beta_c E_i}}{Z_{\beta_C}}$, $q_i=\frac{e^{-\beta_h E_i}}{Z_{\beta_h}}$, and p_i' are the probability amplitudes of state $\rho_{\rm Cold}^1$ when written in the energy eigenbasis of $\hat{H}_{\rm Cold}$. The quantities W_1 and W_∞ are defined by taking the limit $\alpha\rightarrow 1,+\infty$ respectively.

Proof. Eq. (B7) is necessary and sufficient for Eq. (E8) to be satisfied. We can apply the additivity property of the Rényi divergence, to Eq. (B7) to find

$$D_\alpha(\rho_{\rm W}^0\|\tau_{\rm W}) + D_\alpha(\tau_{\beta_c}\|\tau_{\beta_h}) \geq D_\alpha(\rho_{\rm W}^1\|\tau_{\rm W}) + D_\alpha(\rho_{\rm Cold}^1\|\tau_{\beta_h}), \tag{E12}$$

where τ_{β_h} is the thermal state with Hamiltonian $\hat{H}_{\rm W}$ at inverse temperature β_h. We define W_α to be the value of $E_k^{\rm W} - E_j^{\rm W}$ that satisfies Eq. (E12) with equality. A straightforward manipulation of these equations gives the expression for W_α. Then $W_{\rm ext} = \inf_{\alpha\geq 0} W_\alpha$ is the maximum value that satisfies the inequalities Eq. (E12) for all $\alpha \geq 0$. □

As we will see later there exist $\rho_{\rm Cold}^1$ such that, $W_{\rm ext}$ given by Eq. (E9), has a solution (i.e. $W_{\rm ext} > 0$) for *any* $\varepsilon > 0$.

We can use this to write down an explicit solution to the maximization problem Eq. (C2). Using Eqs. (C2), (C17) and Lemma 9, we conclude

$$\eta^{\rm nano}(\rho_{\rm Cold}^1) = \left(1 - \varepsilon + \frac{\Delta C(\rho_{\rm Cold}^1)}{\inf_{\alpha\geq 0} W_\alpha(\rho_{\rm Cold}^1)}\right)^{-1} \tag{E13}$$

where W_α is given by Eqs. (E10), (E11) and recall ΔC can be found in Eq. (C14). From Eqs. (E13), (E10), (E11), we see that the optimization problem $\sup_{\rho_{\rm Cold}^1} \eta^{\rm nano}(\rho_{\rm Cold}^1)$ is still a formidable task. In the next section, see will show that we can use the results from Section D, to drastically simplify the problem.

b. An upper bound for the efficiency

Before moving on to solving the nanoscale efficiency explicitly, we will first use the results of Section D 3 to find upper bounds for the efficiency in the nanoscale regime, in the context of extracting near perfect work (Def. 2).

Recall how we have discussed in comparing Sections B 1 and B 2, that the solution for the family of free entropies F_α, in the case of F_1 is simply the Helmholts free energy. Therefore, from Lemma 9, it follows that W_1 is simply the maximum amount of extractable work according to Eq. (B1). From Eqs. (C5), (C17),

$$\eta^{\rm mac}(\rho_{\rm Cold}^1) = \left(1 - \varepsilon + \frac{\Delta C(\rho_{\rm Cold}^1)}{W_1(\rho_{\rm Cold}^1)}\right)^{-1}. \tag{E14}$$

One can now compare Eq. (E14) with Eq. (E13), and note that for any $\rho_{\rm Cold}^1 \in \mathcal{S}$, we have $W_1(\rho_{\rm Cold}^1) \geq \inf_{\alpha\geq 0} W_\alpha(\rho_{\rm Cold}^1)$. Therefore, we conclude that for any $\rho_{\rm Cold}^1 \in \mathcal{S}$,

$$\eta^{\rm nano}(\rho_{\rm Cold}^1) \leq \eta^{\rm mac}(\rho_{\rm Cold}^1). \tag{E15}$$

Eq. (E15) in conjunction with Lemma 5 has an important consequence. Namely,

$$\sup_{\rho_{\rm Cold}^1 \in \mathcal{S}} \eta^{\rm nano}(\rho_{\rm Cold}^1) \tag{E16}$$

$$\begin{cases} \leq 1 - \beta_h/\beta_c & \text{if the state } \rho_{\rm Cold}^1 \text{ that solves the supremum } \textit{is} \text{ that of a quasi-static heat engine,} \\ < 1 - \beta_h/\beta_c & \text{if the state } \rho_{\rm Cold}^1 \text{ that solves the supremum is } \textit{not} \text{ that of a quasi-static heat engine.} \end{cases} \tag{E17}$$

This tells us that if we cannot achieve the Carnot efficiency for a quasi-static heat engine, we can *never* achieve it, and can only achieve a lower efficiency. We therefore will only consider the quasi-static regime in the rest of Section E.

c. Evaluating near perfect work in the quasi-static heat engine

In light of the results from the previous section, we will now calculate the near perfect work $W_{\rm ext}$ for quasi-static heat engines, i.e. the case where $\varepsilon, g \ll 1$. Specifically, we make the following assumption about the cold bath Hamiltonian:
(A.5) The Hamiltonian is taken to be of n qubits:

$$\hat{H}_{\rm Cold} = \sum_{k=1}^{n} \mathbb{1}^{\otimes(k-1)} \otimes \hat{H}_{{\rm c},k} \otimes \mathbb{1}^{\otimes(n-k)}, \quad \text{where} \quad \hat{H}_{{\rm c},k} = E_k |E_k\rangle\langle E_k|, \tag{E18}$$

and $E_k > 0$ is the energy gap of the k-th qubit.
The tensor product structure in Assumption (A.5) allows us to simplify ρ^0_{Cold}, to

$$\rho^0_{\text{Cold}} = \bigotimes_{i=1}^{n} \tau_{i,\beta_c}, \tag{E19}$$

where τ_{i,β_c} is the thermal state of ith qubit Hamiltonian $\hat{H}_{i,\text{c}}$ at inverse temperature β_c. For the simplicity of following proofs, we present them in the special case of identical qubits, i.e. that $E_i = E$ for all $1 \leq i \leq n$. This means Eq. (E19) can be reduced to

$$\rho^0_{\text{Cold}} = \tau_{\beta_c}^{\otimes n}. \tag{E20}$$

Furthermore, since we consider quasi-static heat engines, the output state is

$$\rho^1_{\text{Cold}} = \tau_{\beta_f}^{\otimes n}, \tag{E21}$$

with $\beta_f = \beta_c - g$,where $0 < g \ll 1$. Eq. (E18) together with Eq. (E21) allows us to further simplify Eq. (E11) to

$$A = \left(\frac{\sum_i p_i^\alpha q_i^{1-\alpha}}{\sum_i p_i'^\alpha q_i^{1-\alpha}}\right)^n, \tag{E22}$$

where $p_i = \frac{e^{-\beta_c E_i}}{Z_{\beta_c}}$, $p_i' = \frac{e^{-\beta_f E_i}}{Z_{\beta_f}}$, $q_i = \frac{e^{-\beta_h E_i}}{Z_{\beta_h}}$ are the probabilities of thermal states (different temperatures) for the qubit Hamiltonian $\hat{H}_\text{c}$. The proof follows along the same lines as the proof to Lemma 9, but now noting that in Eq. (E12) we can replace $D_\alpha(\tau_{\text{Cold}}\|\tau_{\beta_h})$ and $D_\alpha(\rho^1_{\text{Cold}}\|\tau_{\beta_h})$ with $nD_\alpha(\tau_{\beta_c}\|\tau_{\beta_h})$ and $nD_\alpha(\tau_{\beta_f}\|\tau_{\beta_h})$ respectively. This follows from the additivity property of the Rényi divergences. After proving the special case of identical qubits, we show in Theorem 2 that it can be extended to non-identical qubits as generally described by Assumption (A.5).

Since we are dealing with near perfect work and quasi-static heat engines, both $g > 0$ and $\varepsilon > 0$ are infinitesimally small. Thus with the goal in find of finding a solution for W_{ext} from Eqs. (E9), (E10), and (E22); we will proceed to find an expansion of W_α for small ε and g.

i) The expansion of A in a quasi-static heat engine

To simplify our calculations of W_{ext}, especially that of efficiency, it is important to express A in Eq. (E22) in terms of its first order expansion w.r.t. the parameter g. Recall that this parameter $g = \beta_c - \beta_f$ is the difference of inverse temperature between the initial and final state of the cold bath.

Firstly, note that for any integer n, the expression in Eq. (E22) evaluates to $A|_{g=0} = 1$. This is because at $g = 0$, $\beta_f = \beta_c$ and therefore the probabilities p_i, p_i' are identical. To obtain an approximation in the regime $0 < g \ll 1$, we derive

$$\frac{dA}{dg} = -n\left(\sum_i p_i^\alpha q_i^{1-\alpha}\right)^n \left(\sum_i p_i'^\alpha q_i^{1-\alpha}\right)^{-n-1} \left[\sum_i \alpha p_i'^{\alpha-1} q_i^{1-\alpha} \frac{dp_i'}{dg}\right] \tag{E23}$$

$$= -\alpha n A \left(\sum_i p_i'^\alpha q_i^{1-\alpha}\right)^{-1} \left[\sum_i p_i'^\alpha q_i^{1-\alpha}(E_i - \langle \hat{H}_\text{c}\rangle_{\beta_f})\right]. \tag{E24}$$

The first inequality holds by noticing that only the probabilities p_i' depend on g, which means only the denominator in Eq. (E22) is differentiated, using the chain rule

$$\frac{dA(\{p_i'\})}{dg} = \sum_i \frac{dA(\{p_i'\})}{dp_i'} \frac{dp_i'}{dg}. \tag{E25}$$

The equality in Eq. (E24) makes use of the fact that $\frac{dp_i'}{dg} = -\frac{dp_i'}{d\beta_f} = p_i'(E_i - \langle \hat{H}_\text{c}\rangle_{\beta_f})$ as derived in Eq. (D26). Evaluated at

$g = 0$, implies that $p_i' = p_i$, and therefore this gives

$$\left.\frac{dA}{dg}\right|_{g=0} = \alpha n B_\alpha, \quad \text{where} \tag{E26}$$

$$B_\alpha = \frac{1}{\sum_i p_i^\alpha q_i^{1-\alpha}} \sum_i p_i^\alpha q_i^{1-\alpha} \left(\langle \hat{H}_\mathrm{c} \rangle_{\beta_c} - E_i \right). \tag{E27}$$

Recall that p_i, q_i are probabilties of the thermal states of $\hat{H}_\mathrm{c}$, at inverse temperatures β_c, β_h respectively. With this, we can write the expansion of A with respect to g as

$$A = 1 + \alpha n g B_\alpha + \Theta(g^2). \tag{E28}$$

Later on, we will also need to evaluate the derivative of B_α w.r.t. α. This quantity, when evaluated at $\alpha = 1$, has a close relation to the change in average energy of the cold bath (per copy), $\frac{\Delta C}{n}$.

Lemma 10. *Let*

$$\Delta C'(\beta_c) := \left.\frac{d}{dg} \Delta C(\beta_f)\right|_{g=0}, \tag{E29}$$

where recall $\beta_f = \beta_c - g$. Then

$$B_1' = \left.\frac{dB_\alpha}{d\alpha}\right|_{\alpha=1} = \frac{\beta_c - \beta_h}{n} \Delta C'(\beta_f) = (\beta_c - \beta_h) \cdot \mathrm{var}(\hat{H}_\mathrm{c})_{\beta_c}. \tag{E30}$$

Proof. From the definition of ΔC (Eq. (C14)) and using Eqs. (E18), (E19), (E21), we have

$$\frac{\Delta C}{n} = \mathrm{tr}[(\tau_{\beta_f} - \tau_{\beta_c}) \hat{H}_\mathrm{c}]. \tag{E31}$$

Recalling that $\beta_f = \beta_c - g$ and using Eq. (D36), from Eq. (E31) it follows

$$\frac{1}{n} \Delta C'(\beta_c) = \left.\frac{1}{n} \frac{d\Delta C}{dg}\right|_{g=0} = -\left.\frac{1}{n} \frac{d\Delta C}{d\beta_f}\right|_{\beta_f = \beta_c} = \mathrm{var}(\hat{H}_\mathrm{c})_{\beta_c}. \tag{E32}$$

Now, let us evaluate the partial derivative of B_α w.r.t. α. Denoting $r_i = \frac{p_i}{q_i}$, and invoking the chain rule of derivatives for Eq. (E27)

$$\frac{dB_\alpha}{d\alpha} = \left(\sum_i p_i^\alpha q_i^{1-\alpha} \right)^{-2} \left\{ \left[\sum_i q_i r_i^\alpha \ln r_i \left(\langle \hat{H}_\mathrm{c} \rangle_{\beta_c} - E_i \right) \right] \left[\sum_i p_i^\alpha q_i^{1-\alpha} \right] \right. \tag{E33}$$

$$\left. - \left[\sum_i q_i r_i^\alpha \ln r_i \right] \left[\sum_i p_i^\alpha q_i^{1-\alpha} \left(\langle \hat{H}_\mathrm{c} \rangle_{\beta_c} - E_i \right) \right] \right\}. \tag{E34}$$

Substituting $\alpha = 1$ into Eq. (E33), we obtain that $\sum_i p_i^\alpha q_i^{1-\alpha} = 1$. Also, $\sum_i p_i^\alpha q_i^{1-\alpha} \big(\langle \hat{H}_\mathrm{c} \rangle_{\beta_c} - E_i \big) = 0$ while the factor multiplied in front is finite. Therefore, we are left with the terms

$$B_1' = \sum_i p_i \ln r_i \left(\langle \hat{H}_\mathrm{c} \rangle_{\beta_c} - E_i \right) \tag{E35}$$

$$= \sum_i p_i \left[\ln \frac{Z_h}{Z_c} + (\beta_h - \beta_c) E_i \right] (\langle \hat{H}_\mathrm{c} \rangle_{\beta_c} - E_i) \tag{E36}$$

$$= (\beta_c - \beta_h) \mathrm{var}(\hat{H}_\mathrm{c})_{\beta_c} \tag{E37}$$

$$= \frac{\beta_c - \beta_h}{n} \Delta C'(\beta_c). \tag{E38}$$

The second equality comes from substituting $r_i = \frac{p_i}{q_i} = e^{(\beta_h - \beta_c) E_i} \cdot Z_h / Z_c$. In the third equality, $\ln \frac{Z_h}{Z_c}$ is brought out of the summation, while the summation yields 0. Subsequently, we invoke $\sum_i p_i E_i(\langle \hat{H}_\mathrm{c}\rangle_{\beta_c} - E_i) = \langle \hat{H}_\mathrm{c}\rangle_{\beta_c}^2 - \langle \hat{H}_\mathrm{c}^2\rangle_{\beta_c} = -\mathrm{var}(\hat{H}_\mathrm{c})_{\beta_c}$. □

ii) The expansion of W_α in the quasi-static heat engine

We now proceed to derive an expansion of W_α valid for small g, and ε. Note that W_1 is defined through continuity to be the limit of the Rényi divergences at $\alpha \to 1$, and the small ε and g expansion does not hold for $\alpha = 0$, we shall have to examine W_1 and W_0 separately.
(A) For $\varepsilon > 0$, $\alpha \in (0,1) \cup (1,\infty)$

$$\begin{aligned} W_\alpha &= \frac{1}{\beta_h(\alpha-1)}[\ln(A-\varepsilon^\alpha) - \alpha\ln(1-\varepsilon)] && \text{(E39)}\\ &= \frac{1}{\beta_h(\alpha-1)}\left[\ln\left(1+\alpha n g B_\alpha + \Theta(g^2) - \varepsilon^\alpha\right) - \alpha\ln(1-\varepsilon)\right] && \text{(E40)}\\ &= \frac{1}{\beta_h(\alpha-1)}\left[\alpha n g B_\alpha + \Theta(g^2) - \varepsilon^\alpha + \Theta(\varepsilon^{2\alpha}) + \Theta(g\varepsilon^\alpha) - \alpha\left(-\varepsilon + \Theta\left(\varepsilon^2\right)\right)\right], && \text{(E41)}\\ &= \frac{1}{\beta_h(\alpha-1)}\left[\alpha n g B_\alpha - \varepsilon^\alpha + \alpha\varepsilon\right] + \Theta(g^2) + \Theta(\varepsilon^{2\alpha}) + \Theta(g\varepsilon^\alpha) + \Theta(\varepsilon^2). && \text{(E42)} \end{aligned}$$

In the second equality, we have used the expansion of A derived in Eq. (E28). In the third equality, we use the Mercator series

$$\ln(1+x) = \sum_{k=1}^{\infty}\frac{(-1)^{k+1}}{k}x^k, \quad |x|<1, \tag{E43}$$

to expand both of the natural logarithms in line Eq. (E39). The order terms of $\Theta(g^3)$, $\Theta(g^4)$, $\Theta(g^2\varepsilon^\alpha)$ vanish because they are of higher order compared with $\Theta(g^2)$ and $\Theta(g\varepsilon^\alpha)$. The last equality occurs because $c\Theta(g(x)) = \Theta(g(x))$ for any $c \in \mathbb{R}\backslash 0$.
(B) For $\varepsilon > 0$, $\alpha = 1$
We are now interested in finding a small $\varepsilon > 0$, $g > 0$ expansion for W_1, which is defined through continuity of the Rényi divergences. Going back to Eq. (E12), note that W_1 is the maximum value such that Eq. (E12) holds with equality, when all D_α terms in Eq. (E12) are evaluated at $\alpha \to 1$. Recall that $\lim_{\alpha\to 1} D_\alpha(\rho\|\tau) = D(\rho\|\tau)$, the relative entropy we have derived in Section D (see Eq. (B10)). Therefore, one can write an equation for W_1 in a more compact form: W_1 is the value such that

$$n\cdot\left[\langle\hat{H}_\mathrm{c}\rangle_{\beta_c} - \frac{1}{\beta_h}S(\beta_c)\right] = n\cdot\left[\langle\hat{H}_\mathrm{c}\rangle_{\beta_f} - \frac{1}{\beta_h}S(\beta_f)\right] + (1-\varepsilon)W_1 - \frac{1}{\beta_h}\mathrm{h}_2(\varepsilon), \tag{E44}$$

where $\langle\hat{H}_\mathrm{c}\rangle_{\beta_c}$ is the mean energy evaluated at temperature T_{Cold}, $S(\beta_c)$ is the von Neumann entropy of the state τ_{β_c}, and $\mathrm{h}_2(\varepsilon)$ is the binary entropy function. Rearranging Eq. (E44), we get

$$W_1 = \frac{1}{1-\varepsilon}\left[n\langle\hat{H}_\mathrm{c}\rangle_{\beta_c} - n\langle\hat{H}_\mathrm{c}\rangle_{\beta_f} - n\frac{1}{\beta_h}\left(S(\beta_c) - S(\beta_f)\right) + \frac{1}{\beta_h}\mathrm{h}_2(\varepsilon)\right]. \tag{E45}$$

We can expand (E45) using a power law expansion in g and ε for the terms in Eq.(E45), obtaining

$$W_1 = \left[1+\varepsilon+\Theta(\varepsilon^2)\right]\cdot\left[n\frac{d(-\langle\hat{H}_\mathrm{c}\rangle_{\beta_f} + \beta_h^{-1}S(\beta_f))}{dg}\bigg|_{g=0} g + \Theta(g^2) + \frac{1}{\beta_h}\mathrm{h}_2(\varepsilon)\right]. \tag{E46}$$

To proceed, we recall that $\beta_f = \beta_c - g$ and evaluate the term

$$\begin{aligned} \frac{d(-\langle\hat{H}_\mathrm{c}\rangle_{\beta_f} + \beta_h^{-1}S(\beta_f))}{dg}\bigg|_{g=0} &= \frac{d(\langle\hat{H}_\mathrm{c}\rangle_{\beta_f} - \beta_h^{-1}S(\beta_f))}{d\beta_f}\bigg|_{\beta_f=\beta_c} = -\mathrm{var}(\hat{H}_\mathrm{c})_{\beta_c} + \frac{\beta_c}{\beta_h}\mathrm{var}(\hat{H}_\mathrm{c})_{\beta_c} && \text{(E47)}\\ &= \frac{\beta_c-\beta_h}{\beta_h}\mathrm{var}(\hat{H}_\mathrm{c})_{\beta_c}. && \text{(E48)} \end{aligned}$$

This implies that when fully expanded, Eq. (E46) reads as

$$W_1 = ng\frac{\beta_c-\beta_h}{\beta_h}\mathrm{var}(\hat{H}_\mathrm{c})_{\beta_c} + \beta_h^{-1}\mathrm{h}_2(\varepsilon) + \Theta(\varepsilon g) + \Theta(\varepsilon)\mathrm{h}_2(\varepsilon) + \Theta(g\varepsilon^2) + \Theta(\varepsilon^2)\mathrm{h}_2(\varepsilon) \tag{E49}$$

$$+ \Theta(g^2) + \Theta(\varepsilon g^2) + \Theta(\varepsilon^2 g^2) \tag{E50}$$

$$= ng\frac{\beta_c-\beta_h}{\beta_h}\mathrm{var}(\hat{H}_\mathrm{c})_{\beta_c} + \beta_h^{-1}(-\varepsilon\ln\varepsilon+\varepsilon) + \Theta(\varepsilon g) + \Theta(\varepsilon^2\ln\varepsilon) + \Theta(\varepsilon^2) + \Theta(g^2), \tag{E51}$$

where we have used $\mathrm{h}_2(\varepsilon) = -\varepsilon\ln\varepsilon + \Theta(\varepsilon)$, which follows from finding the power-law expansion of the second term in Eq. (A10).

Although Eq. (E42) is not defined for $\alpha = 1$, we can evaluate it in the limit $\alpha \to 1$ to see if it coincides with the correct expression of W_1 (in Eq. (E51)) at least for the leading order term (found in square brackets of Eq. (E42)). For the leading order term of Eq. (E42), we find

$$\lim_{\alpha\to1}\frac{1}{\beta_h(\alpha-1)}\left[\alpha ngB_\alpha - \varepsilon^\alpha + \alpha\varepsilon\right] = \beta_h^{-1}\left[ng\lim_{\alpha\to1}\frac{\alpha B_\alpha}{\alpha-1} - \lim_{\alpha\to1}\frac{\varepsilon^\alpha-\alpha\varepsilon}{\alpha-1}\right] \tag{E52}$$

$$= \beta_h^{-1}\left[ng\lim_{\alpha\to1}\frac{\alpha B_\alpha}{\alpha-1} + (-\varepsilon\ln\varepsilon+\varepsilon)\right], \tag{E53}$$

$$= ng\frac{\beta_c-\beta_h}{\beta_h}\mathrm{var}(\hat{H}_\mathrm{c})_{\beta_c} + \beta_h^{-1}(-\varepsilon\ln\varepsilon+\varepsilon). \tag{E54}$$

The last equality holds because

$$\lim_{\alpha\to1}\frac{\alpha B_\alpha}{\alpha-1} = \lim_{\alpha\to1}\frac{dB_\alpha}{d\alpha} \tag{E55}$$

$$= (\beta_c-\beta_h)\cdot\mathrm{var}(\hat{H}_\mathrm{c})_{\beta_c}, \tag{E56}$$

where Eq. (E55) is derived from L'Hôspital rule ($B_1 = 0$ follows from the definition, see Eq. (E26)), and Eq. (E56) comes by invoking Lemma 10. Thus noting that Eq. (E54) is simply the first two terms in Eq. (E56), we conclude that the small $g > 0$ and $\varepsilon > 0$ expansion of W_α for $\alpha > 0$ can be summarized as

$$W_\alpha = \tag{E57}$$

$$\begin{cases} \frac{1}{\beta_h(\alpha-1)}\left[\alpha ngB_\alpha - \varepsilon^\alpha + \alpha\varepsilon\right] + \Theta(g^2) + \Theta(\varepsilon^{2\alpha}) + \Theta(g\varepsilon^\alpha) + \Theta(\varepsilon^2) & \text{if } \alpha > 0,\ \alpha \neq 1 \\ \lim\limits_{\alpha\to1^+}\dfrac{1}{\beta_h(\alpha-1)}\left[\alpha ngB_\alpha - \varepsilon^\alpha + \alpha\varepsilon\right] + \Theta(\varepsilon g) + \Theta(\varepsilon^2\ln\varepsilon) + \Theta(\varepsilon^2) + \Theta(g^2) & \text{if } \alpha = 1. \end{cases} \tag{E58}$$

(C) For $\alpha = 0$

We will now investigate the $\alpha = 0$ case. This is also particularly important to understand the difference between perfect and near perfect work, since in Section E 1, the impossibility of extracting perfect work arises from evaluating the allowed values of W_ext under the $\alpha = 0$ constraint. We show that by allowing $\varepsilon > 0$, $W_\mathrm{ext} > 0$ is allowed once again. Recall $D_0(p\|q) = \lim_{\alpha\to0} D_\alpha(p\|q) = \sum_{i:p_i\neq0} q_i$. Thus from Eq. (E12)

$$D_0(\rho_\mathrm{W}^0\|\tau_\mathrm{W}) - D_0(\rho_\mathrm{W}^1\|\tau_\mathrm{W}) \geq nD_0(\tau_{\beta_f}\|\tau_{\beta_h}) - nD_0(\tau_{\beta_c}\|\tau_{\beta_h}) = 0. \tag{E59}$$

where the last equality follows from the fact that thermal states have full rank. This inequality is satisfied for any value of W_ext, since whenever $\varepsilon > 0$, ρ_W^1 is a full rank state, and $D_0(\rho_\mathrm{W}^1\|\tau_\mathrm{W}) = 0$. Furthermore, $D_0(\rho_\mathrm{W}^0\|\tau_\mathrm{W}) \geq 0$ because all Rényi divergences are non-negative. Therefore, taking into account Eqs. (E57) and (E59), for quasi-static heat engines which extract near perfect work, we only need to solve

$$W_\mathrm{ext} = \inf_{\alpha>0} W_\alpha, \tag{E60}$$

where W_α is given by Eq. (E57).

d. The choice of ε determines the infimum to evaluating W_{ext}

In this section, we will show that the infimum over all $\alpha > 0$ in Eq. (E60) can be simplified to taking the infimum over $\alpha > \bar{\kappa}$ instead, where the parameter $\bar{\kappa}$ determines how quickly ε goes to 0 w.r.t. the parameter g. We define κ in Lemma 11 and show its existence, for any function of $\varepsilon(g)$ such that $\lim_{g\to 0^+}\varepsilon(g) = 0$.

Lemma 11. *For every continuous function $\varepsilon(g) > 0$ satisfying $\lim_{g\to 0^+}\varepsilon(g) = 0$, $\exists\, \bar{\kappa} \in \mathbb{R}_{\geq 0}$ s.t.*

$$\delta(\kappa) = \lim_{g\to 0^+}\frac{\varepsilon^{\kappa}(g)}{g} = \begin{cases} 0 & \text{if } \kappa > \bar{\kappa} \\ \sigma \geq 0 & \text{if } \kappa = \bar{\kappa} \\ \infty & \text{if } \kappa < \bar{\kappa} \end{cases} \tag{E61}$$

where $\bar{\kappa} = +\infty$ is allowed (that is to say, $\lim_{g\to 0^+}\frac{\varepsilon^{\kappa}(g)}{g}$ diverges for every κ) and $\sigma = +\infty$ is also allowed.

Proof. The main idea in this proof is to divide the non-negative real line into an infinite sequence of intervals in an iterative process. We specify the ends of these intervals by constructing a sequence $\{\kappa_i\}_{i=1}^{\infty}$, and evaluating δ at these points. We then prove that according to our construction, there are only two possibilities:
1) κ_i forms a convergent sequence, where the limit $\lim_{n\to\infty}\kappa_n = \bar{\kappa}$, or
2) the ends of these intervals extend to infinity. In this case, $\bar{\kappa} = \infty$. The way to construct this interval is as follows: in the first round, pick some $\kappa_1 > 0$. The corresponding interval is $[0, \kappa_1]$. Evaluate $\delta(\kappa_1)$. If $\delta(\kappa_1) = \infty$, then proceed to look at the interval $[\kappa_1, \frac{3}{2}\kappa_1]$. Otherwise if $\delta(\kappa_1) < \infty$, choose $\kappa_2 = \frac{\kappa_1}{2}$ and evaluate $\delta(\kappa_2)$. Depending on whether $\delta(\kappa_2)$ goes to infinity, we pick one of the intervals $[0, \kappa_2]$ or $[\kappa_2, \kappa_1]$.

A general expression of choosing κ_n can be written: during the nth round, define the sets $\mathcal{S}_n^{(0)}, \mathcal{S}_n^{(\infty)}$ such that

$$\mathcal{S}_n^{(0)} = \{\kappa_i | 1 \leq i \leq n \text{ and } \delta(\kappa_i) = 0\}$$
$$\mathcal{S}_n^{(\infty)} = \{\kappa_i | 1 \leq i \leq n \text{ and } \delta(\kappa_i) = \infty\}.$$

Note that if we find $\delta(\kappa_i) = c \neq 0$ for some finite constant c, then our job is finished, i.e. $\bar{\kappa} = \kappa_i$ (We prove this later). Subsequently, define for $n \geq 1$,

$$\kappa_n^{(0)} = \min_{\kappa\in\mathcal{S}_n^{(0)}}\kappa \qquad \text{and} \qquad \kappa_n^{(\infty)} = \max_{\kappa\in\mathcal{S}_n^{(\infty)}}\kappa.$$

If either sets are empty, we use the convention that the corresponding minimization/maximization equals 0. Once these quantities are defined, we can choose the next interval by evaluating

$$\kappa_{n+1} = \kappa_n^{(\infty)} + \frac{|\kappa_n^{(\infty)} - \kappa_n^{(0)}|}{2}. \tag{E62}$$

In the n-th round, the corresponding interval is $[\kappa_n^{(\infty)}, \kappa_{n+1}]$.

Let us now analyze why we can use this scheme to find $\bar{\kappa}$. Firstly, consider the case where $\delta(\kappa_i)$ whenever evaluated, produces infinity. This means that in each round, $\kappa_n^{(\infty)} = \kappa_n$ increases with n (by the iterative scheme), and $\kappa_n^{(0)} = 0$ always stays at zero. Note that this scheme has been constructed in a way such that $\lim_{n\to\infty}\kappa_n = \infty$. Indeed, for all n, by using Eq. (E62),

$$\kappa_{n+1} = \frac{3}{2}\kappa_n = \left(\frac{3}{2}\right)^2\kappa_{n-1} = \cdots = \left(\frac{3}{2}\right)^n\kappa_1, \tag{E63}$$

which tends to infinity as n goes to infinity, whenever $\kappa_1 > 0$. Later we will prove a property of the function δ, which combined with this scenario means that $\delta(\kappa) = \infty$ for every $\kappa \geq 0$. Therefore, $\bar{\kappa} = \infty$.

Next, suppose that there exist an n-th round, such that $\delta(\kappa_n) = \infty$ and $\delta(\kappa_{n+1}) < \infty$, as illustrated in Fig 6. Note that the function $\delta(\kappa)$ has a peculiar property, i.e. we know that if $\delta(\kappa_n) = \infty$, then for any $\kappa < \kappa_n$,

$$\delta(\kappa) = \lim_{g\to 0^+}\underbrace{\varepsilon^{\kappa-\kappa_n}(g)}_{\to+\infty}\underbrace{\frac{\varepsilon^{\kappa_n}(g)}{g}}_{\to\infty} = +\infty. \tag{E64}$$

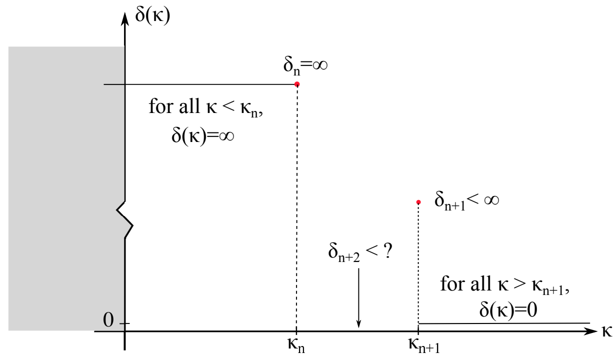

FIG. 6. Illustration of the scenario where $\delta(\kappa_n) = \infty$ and $\delta(\kappa_{n+1}) < \infty$.

On the other hand, if $\delta(\kappa_{n+1}) = 0$, then we know that for any $\kappa > \kappa_{n+1}$,

$$\delta(\kappa) = \lim_{g\to 0^+} \underbrace{\varepsilon^{\kappa-\kappa_{n+1}}(g)}_{\to 0} \underbrace{\frac{\varepsilon^{\kappa_{n+1}}(g)}{g}}_{\to 0} = 0. \tag{E65}$$

Moreover, if $\delta(\kappa_j) = c \neq 0$ for some positive, finite c, then following the same arguments, one can easily see that for all $\kappa < \kappa_j$, $\delta(\kappa) = \infty$ and for $\kappa > \kappa_j$, $\delta(\kappa) = 0$. In this case we find that $\bar{\kappa} = \kappa_j$. These observations are illustrated in Figure 6 for clarity.

One can now evaluate κ_{n+2} (which is the midpoint of κ_n and κ_{n+1}) and its corresponding value of $\delta(\kappa_{n+2})$. From this point on, in each iteration we either find $\bar{\kappa}$ exactly (whenever the function δ when evaluated produces a finite, non-zero number), or the length of the next interval gets halved, and goes to zero in the limit of $n \to \infty$. This, by Eq. (E62), also implies that $\lim_{n\to\infty} \kappa_n^{(\infty)} = \lim_{n\to\infty} \kappa_n^{(0)}$. We also know the following:
1) for all $\kappa < \kappa_n^{(\infty)}$, $\delta(\kappa) = \infty$,
2) for all $\kappa > \kappa_n^{(0)}$, $\delta(\kappa) = 0$.
Therefore, we see that $\bar{\kappa}$ exists and $\bar{\kappa} = \lim_{n\to\infty} \kappa_n^{(\infty)} = \lim_{n\to\infty} \kappa_n^{(0)}$. By this we conclude the proof. □

To provide some intuition about how $\bar{\kappa}$ compares the rate of convergence $\varepsilon, g \to 0$, let us look at the following examples:
1) Consider $\varepsilon_1(g) = \exp(-1/g)$. Then $\bar{\kappa} = 0$ with $\sigma = \infty$.
2) Consider $\varepsilon_2(g) = g \ln g$. Then $\bar{\kappa} = 1$ with $\sigma = \infty$.
3) Consider $\varepsilon_3(g) = c \cdot g^{1/k}$ for $k > 0$. Then $\bar{\kappa} = k$ with $\sigma = c$.

In the next lemma, we consider the scenario of near perfect work, given in Def. 2, and show that this imposes a finite range of values $\bar{\kappa}$ should take. Given a particular $\bar{\kappa}$, we also show that the minimization of Eq. (E60) changes with $\bar{\kappa}$.

Lemma 12. *Given any $\varepsilon(g) \in (0, 1]$ as a continuous function of g, where $g > 0$. If $\lim_{g\to 0^+} \varepsilon(g) = 0$ and $\lim_{g\to 0^+} \frac{\Delta S}{W_{\text{ext}}} = 0$, then the following holds:*

1. *The quantity $\bar{\kappa}$ (defined in Lemma 11) can only have any value in $\bar{\kappa} \in [0, 1]$, where $\lim_{g\to 0^+} \frac{\varepsilon \ln \varepsilon}{g} = 0$ has to hold if $\bar{\kappa} = 1$.*

2. *The extractable work can be written as*

$$W_{\text{ext}} = g \cdot \left[\inf_{\alpha \geq \bar{\kappa}} \frac{n\alpha B_\alpha}{\alpha - 1} + f(g) \right], \tag{E66}$$

where $\lim_{g\to 0^+} f(g) = 0$ and $\inf_{\alpha\geq\bar{\kappa}}$ can be exchanged for $\inf_{\alpha>\bar{\kappa}}$ if $\bar{\kappa} = 0$.

Proof. Firstly, let us use Eq. (E57) to simplify our expression for W_{ext}: $W_{\text{ext}} = \inf_{\alpha\geq 0} W_\alpha$, where

$$\beta_h W_\alpha = \begin{cases} g\tilde{W}_\alpha + \Theta(g^2) + \Theta(\varepsilon^{2\alpha}) + \Theta(g\varepsilon^\alpha) + \Theta(\varepsilon^2) & \text{if } \alpha \in (0, 1) \cup (1, \infty) \\ g\tilde{W}_1 + \Theta(\varepsilon g) + \Theta(\varepsilon^2 \ln \varepsilon) + \Theta(\varepsilon^2) + \Theta(g^2) & \text{if } \alpha = 1, \end{cases} \tag{E67}$$

and

$$\tilde{W}_\alpha := \frac{1}{\alpha-1}\left(\alpha n B_\alpha + \alpha\frac{\varepsilon}{g} - \frac{\varepsilon^\alpha}{g}\right), \tag{E68}$$

and for $\alpha = 1$

$$\tilde{W}_1 = \left(\lim_{\alpha\to 1}\frac{\alpha n B_\alpha}{\alpha-1}\right) + \frac{\varepsilon}{g} - \frac{\varepsilon}{g}\ln(\varepsilon). \tag{E69}$$

From now on, the order terms in Eq. (E67) can be neglected, since it can be checked that all of them are of higher order compared to the terms we grouped in $\tilde{W}_\alpha$, in the limit of vanishing g. Even then, we note that due to the complicated form of W_{ext}, it is not straightforward to begin our proof with the assumption $\lim_{g\to 0^+}\frac{\Delta S}{W_{\text{ext}}} = 0$.

Instead, we begin by noting that given a function $\varepsilon(g)$ that satisfies the conditions of the above lemma, then one can invoke Lemma 11, and therefore there exists a $\bar{\kappa}\in\mathbb{R}_{\geq 0}$ such that Eq. (E61) holds. We then, for all possible $\kappa\in\mathbb{R}_{\geq 0}$, evaluate all $\tilde{W}_\alpha$ to take the infimum and obtain W_{ext}. Given W_{ext}, we then evaluate the quantity $\lim_{g\to 0^+}\frac{\Delta S}{W_{\text{ext}}} = 0$.

The value of $\bar{\kappa}$ determines how the limits of quantities like $\frac{\varepsilon}{g}$, $\frac{\varepsilon^\alpha}{g}$ behave. Therefore, we need to split the analysis into three different regimes: $\bar{\kappa}\in[0,1)$, $\bar{\kappa}=1$, $\bar{\kappa}\in(1,\infty)$.

1) For $\bar{\kappa}\in[0,1)$

For this case, we know the following limits:

A. $\lim_{g\to 0^+}\frac{\varepsilon}{g} = 0$.

B. For $\alpha<\bar{\kappa}$, $\lim_{g\to 0^+}\frac{\varepsilon^\alpha}{g} = \infty$.

C. For $\alpha=\bar{\kappa}$, $\lim_{g\to 0^+}\frac{\varepsilon^\alpha}{g} = \sigma\geq 0$.

D. For $\alpha>\bar{\kappa}$, $\lim_{g\to 0^+}\frac{\varepsilon^\alpha}{g} = 0$.

E. Note that $\exists\, k_1>\bar{\kappa}$ such that $1-k_1>0$. Thus $\lim_{g\to 0^+}\frac{\varepsilon}{g}\ln\varepsilon = \lim_{g\to 0^+}\frac{\varepsilon^{k_1}}{g}\,\varepsilon^{1-k_1}\ln\varepsilon = 0$

Therefore, by using Eq. (E68) and (E69) (for $\alpha=1$ separately) we have

$$\tilde{W}_\alpha = \begin{cases} +\infty & \text{if } \alpha\in[0,\bar{\kappa}) \\ \dfrac{\alpha n B_\alpha}{\alpha-1} + \sigma + \Theta\left(\dfrac{\varepsilon}{g}\right) & \text{if } \alpha=\bar{\kappa} \\ \dfrac{\alpha n B_\alpha}{\alpha-1} + \Theta\left(\dfrac{\varepsilon^\alpha}{g}\right) & \text{if } \alpha\in(\bar{\kappa},1) \\ \dfrac{\alpha n B_\alpha}{\alpha-1} + \Theta\left(\dfrac{\varepsilon}{g}\right) & \text{if } \alpha\in(1,\infty) \\ \lim\limits_{\alpha\to 1}\dfrac{\alpha n B_\alpha}{\alpha-1} + \Theta\left(\dfrac{\varepsilon\ln\varepsilon}{g}\right) & \text{if } \alpha=1, \end{cases} \tag{E70}$$

where the expression in Eq.(E70) has been written as a leading order term, plus higher order terms that vanish in the limit $g\to 0$.

Therefore, we conclude that for $\bar{\kappa}\in[0,1)$ and any $\sigma\geq 0$, due to continuity in α of $\frac{\alpha n B_\alpha}{\alpha-1}$,

$$\beta_h W_{\text{ext}} = \inf_{\alpha>0} W_\alpha = g\cdot\left[\inf_{\alpha\geq\bar{\kappa}}\frac{\alpha n B_\alpha}{\alpha-1} + \Theta\left(f(g)\right)\right], \tag{E71}$$

where f satisfies $\lim_{g\to 0^+} f(g) = 0$ in the expression of Eq. (E70), Both functions vanish as g tends to zero. Note that if $\bar{\kappa}=0$, then $\inf_{\alpha\geq\bar{\kappa}}$ can be exchanged for $\inf_{\alpha>\bar{\kappa}}$ since in Eq. (E60) the point $\alpha=0$ was already excluded.

We can now calculate $\lim_{g\to 0^+}\frac{\Delta S}{W_{\text{ext}}}$ for $\bar{\kappa}\in[0,1)$ and any $\sigma\geq 0$:

$$\lim_{g\to 0^+}\frac{\Delta S}{W_{\text{ext}}} = \lim_{g\to 0^+}\frac{-\varepsilon\ln\varepsilon-(1-\varepsilon)\ln(1-\varepsilon)}{\left(\inf_{\alpha\geq\bar{\kappa}}\frac{\alpha n B_\alpha}{\alpha-1}\right)g} = \lim_{g\to 0^+}\frac{1}{\inf_{\alpha\geq\bar{\kappa}}\frac{\alpha n B_\alpha}{\alpha-1}}\left(-\underbrace{\frac{\varepsilon\ln\varepsilon}{g}}_{\to 0\text{ (Item E)}} - \underbrace{\frac{\varepsilon+\Theta(\varepsilon^2)}{g}}_{\to 0\text{ (Item A)}}\right) = 0, \tag{E72}$$

where we have assumed that

$$\inf_{\alpha\geq\bar{\kappa}} \frac{\alpha n B_\alpha}{\alpha-1} > 0. \tag{E73}$$

As we will see later (see Eq. (E89)), Eq. (E73) holds if $\bar{\kappa} > 0$. However,

$$\frac{\alpha n B_\alpha}{\alpha-1} = 0, \tag{E74}$$

if $\alpha = 0$ and we need to use Eq. (E70) for the case $\alpha \in (\bar{\kappa}, 1)$ for $\bar{\kappa} = 0$. From which we conclude that

$$\beta_h W_{\text{ext}} = \inf_{\alpha>0} W_\alpha \geq \varepsilon^{1/2} = \Theta(f(g)), \tag{E75}$$

thus we have

$$\lim_{g\to 0^+} \frac{\Delta S}{W_{\text{ext}}} \leq \lim_{g\to 0^+} \frac{-\varepsilon\ln\varepsilon - (1-\varepsilon)\ln(1-\varepsilon)}{\varepsilon^{1/2}/g} = \lim_{g\to 0^+} g(-\varepsilon^{1/2}\ln\varepsilon - \varepsilon^{1/2}) = 0. \tag{E76}$$

thus from Eqs. (E74), (E75), and (E76), we conclude that Eq. Eqs. (E71) and (E72) are still valid when $\bar{\kappa} = 0$. To summarize, so far we have proven that whenever $\bar{\kappa} \in [0, 1)$, Eq. (E66) holds for some $f(g)$ which vanishes as g tends to zero, and furthermore $\lim_{g\to 0^+} \frac{\Delta S}{W_{\text{ext}}} = 0$.

2) For $\bar{\kappa} \in (1, \infty)$

In this regime, like the previous analysis, we can list out the following limits:
A. $\lim_{g\to 0^+} \frac{\varepsilon}{g} = 0$.
B. By definition of $\bar{\kappa}$, for $\alpha < 1$, $\lim_{g\to 0^+} \frac{\varepsilon^\alpha}{g} = \infty$.
C. $\lim_{g\to 0^+} \frac{\varepsilon\ln\varepsilon}{g} = \infty$ since both $\frac{\varepsilon}{g}$ and $\ln\varepsilon$ goes to infinity as $g \to 0$.
Therefore, by using Eq. (E68) and (E69) (for $\alpha = 1$ separately) we have

$$\tilde{W}_\alpha = \begin{cases} \frac{1}{g}\cdot\frac{1}{1-\alpha}\left[\varepsilon^\alpha + \Theta(\varepsilon) + \Theta(g)\right] & \text{if } \alpha \in [0,1) \\ \frac{1}{g}\cdot\left[-\varepsilon\ln\varepsilon + \Theta(\varepsilon) + \Theta(g)\right] & \text{if } \alpha = 1 \\ \frac{1}{g}\cdot\frac{1}{\alpha-1}\left[\alpha\varepsilon + \Theta(\varepsilon^\alpha) + \Theta(g)\right] & \text{if } \alpha \in (1,\infty). \end{cases} \tag{E77}$$

Note that for all of these expressions in Eq. (E77), $\tilde{W}_\alpha \to \infty$. Next we want to calculate W_{ext}, which is the infimum of W_α, taken over all $\alpha \geq 0$. Note that in the limit of vanishing g, ε also goes to zero. Therefore in Eq. (E77), the equation of $\tilde{W}_\alpha$ which vanishes most quickly in the limit $g \to 0$ happens when $\alpha \in (1, \infty)$. Therefore, we conclude that for $\bar{\kappa} \in (1, \infty)$ and any $\sigma \geq 0$,

$$\beta_h W_{\text{ext}} = \inf_{\alpha\geq 1} W_\alpha = g\cdot\left[\inf_{\alpha\geq 1}\frac{\alpha}{\alpha-1}\frac{\varepsilon}{g} + \Theta\left(f(g)\right)\right] = \varepsilon + g\cdot\Theta\left(f(g)\right) \tag{E78}$$

We can now calculate $\lim_{g\to 0^+} \frac{\Delta S}{W}$ for $\bar{\kappa} \in (1, \infty)$ and any $\sigma \geq 0$:

$$\lim_{g\to 0^+} \frac{\Delta S}{W} = \lim_{g\to 0^+} \frac{-\varepsilon\ln\varepsilon - (1-\varepsilon)\ln(1-\varepsilon)}{\varepsilon} = \lim_{g\to 0^+} -\underbrace{\frac{\varepsilon\ln\varepsilon}{\varepsilon}}_{\to\infty} - \underbrace{\frac{\varepsilon+\Theta(\varepsilon^2)}{\varepsilon}}_{\to 1} = +\infty. \tag{E79}$$

From this, we note that the whole regime of $\bar{\kappa} \in (1, \infty)$ does not contain any cases corresponding to our condition of interest: $\lim_{g\to 0^+} \frac{\Delta S}{W_{\text{ext}}} = 0$ never holds.

3) For $\bar{\kappa} = 1$

Similar to the first two cases, we again list out the relevant limits:
A. $\lim_{g\to 0^+} \frac{\varepsilon}{g} = \sigma$ for some $\sigma \geq 0$.
B. For $\alpha < 1$, $\lim_{g\to 0^+} \frac{\varepsilon^\alpha}{g} = \infty$.
C. For $\alpha > 1$, $\lim_{g\to 0^+} \frac{\varepsilon^\alpha}{g} = 0$.

Therefore, by using Eq. (E68) and (E69) (for $\alpha = 1$ separately) we have

$$\tilde{W}_\alpha = \begin{cases} \frac{1}{g} \cdot \frac{1}{1-\alpha} \left[\varepsilon^\alpha + \Theta(\varepsilon) + \Theta(g)\right] & \text{if } \alpha \in [0,1) \\ \frac{1}{g} \cdot \left[-\varepsilon \ln \varepsilon + \Theta(\varepsilon) + \Theta(g)\right] & \text{if } \alpha = 1 \ \ \&\& \ \ \sigma > 0 \\ n \lim_{\alpha \to 1} \dfrac{\alpha B_\alpha}{\alpha - 1} - \dfrac{\varepsilon \ln \varepsilon}{g} \geq n \lim_{\alpha \to 1} \dfrac{\alpha B_\alpha}{\alpha - 1} & \text{if } \alpha = 1 \ \ \&\& \ \ \sigma = 0 \\ \frac{1}{\alpha - 1} \left[\alpha n B_\alpha + \alpha\sigma - \Theta\left(\frac{\varepsilon^\alpha}{g}\right)\right] & \text{if } \alpha \in (1, \infty). \end{cases} \tag{E80}$$

Note that for $\alpha \in [0,1)$ and the case $\alpha = 1$ && $\sigma > 0$, $\tilde{W}_\alpha$ tends to infinity, while for the other cases $\tilde{W}_\alpha$ is finite.

Therefore, we can conclude that for $\bar{\kappa} = 1$,

$$\beta_h W_{\text{ext}} = g \cdot \left[\left(\inf_{\alpha \geq 1} \frac{\alpha}{\alpha - 1} \left(n B_\alpha + \sigma\right)\right) + \Theta\left(f(g)\right)\right], \tag{E81}$$

where $f(g) = \frac{\varepsilon^\alpha}{g}$ vanishes as g tends to zero.

Now, we evaluate the limit $\lim_{g \to 0^+} \frac{\Delta S}{W}$ for $\bar{\kappa} = 1$ and any $\sigma \geq 0$:

$$\lim_{g \to 0^+} \frac{\Delta S}{W} = \lim_{g \to 0^+} \frac{-\varepsilon \ln \varepsilon - (1-\varepsilon) \ln(1-\varepsilon)}{\left(\inf_{\alpha \geq 1} \frac{1}{\alpha - 1} \left(\alpha n F_\alpha + \alpha\sigma\right)\right) g} = \lim_{g \to 0^+} \frac{-\varepsilon \ln \varepsilon}{c \cdot g} - \underbrace{\frac{\varepsilon + \Theta(\varepsilon^2)}{c \cdot g}}_{\to 0}. \tag{E82}$$

This limit of interest can be zero if and only if $\lim_{g \to 0^+} \frac{\varepsilon \ln \varepsilon}{g} = 0$.

We have calculated the limits $\lim_{g \to 0^+} \Delta S / W_{\text{ext}}$ to leading order in g for all functions $\varepsilon(g) > 0$ satisfying $\lim_{g \to 0^+} \varepsilon = 0$. These are found in Eqs. (E72), (E79), and (E82). We have found that $\lim_{g \to 0^+} \Delta S / W_{\text{ext}} = 0$ occurs only in two cases:
i) $\bar{\kappa} \in [0,1)$, and
ii) $\bar{\kappa} = 1$ and $\lim_{g \to 0^+} \frac{\varepsilon \ln \varepsilon}{g} = 0$.
The amount of work, W_{ext} is found in Eq. (E71) and (E81) respectively. Indeed, they take the form of Eq. (E66), for different functions $f(g)$. With this, we conclude the proof of the lemma. □

e. *Solving the infimum for W_{ext}*

We have seen in Lemma 12 that the function $\frac{\alpha B_\alpha}{\alpha - 1}$ corresponds to the largest order term in W_{ext} w.r.t. small g (quasi-static heat engine). Our next objective is to find the infimum of $\frac{\alpha B_\alpha}{\alpha - 1}$ over $\alpha \in [\bar{\kappa}, \infty]$ appearing in Eq. (E66). Such an infimum is is not easy to evaluate, but whenever the cold bath consists of multiple identical qubits, we show that the derivative $\frac{d}{d\alpha} \frac{\alpha B_\alpha}{\alpha - 1}$ has some nice properties. Roughly speaking, we show that this derivative does not have many roots, which in turn means that $\frac{\alpha B_\alpha}{\alpha - 1}$ does not have many turning points. We have used this to prove in Lemma 14 that the infimum is either obtained at $\alpha = \bar{\kappa}$ or $\alpha \to \infty$.

The derivative of $\frac{\alpha B_\alpha}{\alpha - 1}$ w.r.t. α is given by

$$\frac{d}{d\alpha} \frac{\alpha B_\alpha}{\alpha - 1} = \frac{B_\alpha}{\alpha - 1} + \alpha \frac{B'_\alpha}{\alpha - 1} - \frac{\alpha B_\alpha}{(\alpha - 1)^2} = \frac{B'_\alpha}{(\alpha - 1)^2} \left[\alpha(\alpha - 1) - \frac{B_\alpha}{B'_\alpha}\right] = \frac{B'_\alpha}{(\alpha - 1)^2} G(\alpha), \tag{E83}$$

where

$$G(\alpha) := \alpha(\alpha - 1) - \frac{B_\alpha}{B'_\alpha}. \tag{E84}$$

Now, we shall evaluate the quantities B_α, B'_α, and $\frac{d}{d\alpha} \frac{B_\alpha}{B'_\alpha}$ for the case of qubits (see Assumption (A.5)), where the energy levels are $\{0, E\}$. By using Eq. (E18), we evaluate the quantity B_α defined by Eq. (E27) to obtain a simple expression:

$$B_\alpha = E \cdot \frac{e^{-\beta_c E}}{1+e^{-\beta_c E}} - E \cdot \frac{e^{-\alpha\beta_c E} e^{-(1-\alpha)\beta_h E}}{1+e^{-\alpha\beta_c E} e^{-(1-\alpha)\beta_h E}} \tag{E85}$$

$$= E \cdot \frac{1}{1+e^{\beta_c E}} - E \cdot \frac{e^{\alpha\beta_h E}}{e^{\alpha\beta_h E} + e^{(\beta_h+\alpha\beta_c)E}} \tag{E86}$$

$$= \frac{E}{1+e^{\beta_c E}} \cdot \left[1 - \frac{e^{\alpha\beta_h E}(1+e^{\beta_c E})}{e^{\alpha\beta_h E} + e^{(\beta_h+\alpha\beta_c)E}}\right] \tag{E87}$$

$$= \frac{E}{1+e^{\beta_c E}} \cdot \frac{e^{(\beta_h+\alpha\beta_c)E} - e^{(\beta_c+\alpha\beta_h)E}}{e^{\alpha\beta_h E} + e^{(\beta_h+\alpha\beta_c)E}}. \tag{E88}$$

We note that Eq. (E88) is zero only if $\alpha = 1$, and thus for $\alpha \neq 1$, $\alpha B_\alpha/(\alpha-1) \neq 0$. From Eq. (E56), we know that $\lim_{\alpha\to 1} \alpha B_\alpha/(\alpha-1) > 0$, thus due to continuity,

$$\frac{\alpha B_\alpha}{\alpha - 1} > 0 \quad \forall\, \alpha > 0. \tag{E89}$$

We also derive the first derivative of B_α w.r.t. α for the special case of qubits:

$$B'_\alpha = \frac{dB_\alpha}{d\alpha} = \frac{E^2(\beta_c - \beta_h)}{\left[e^{\alpha\beta_h E} + e^{(\beta_h+\alpha\beta_c)E}\right]^2} \cdot e^{(\beta_h+\alpha\beta_c+\alpha\beta_h)E}. \tag{E90}$$

Note that since $\beta_c > \beta_h$ by definition, therefore whenever $E > 0$, then $B'_\alpha > 0$ always holds. By further algebraic manipulation, we compute the first derivative of the function

$$\frac{d}{d\alpha}\frac{B_\alpha}{B'_\alpha} = \frac{\cosh[w(\beta_c,\beta_h,\alpha)E]}{\cosh(\beta_c E/2)}, \tag{E91}$$

where $w(\beta_c,\beta_h,\alpha) = (\beta_c - \beta_h)\alpha + \beta_h - \frac{\beta_c}{2}$.

We have written Eq. (E83) in this form, since for the special case of qubits, namely Eq. (E90), $B'_\alpha > 0$ is always true. Therefore, looking at the function $G(\alpha)$ whether it is positive or negative) will tell us whether $\frac{\alpha B_\alpha}{\alpha-1}$ (and therefore W_α) is increasing or decreasing in a particular interval.

In Lemma 13, we identify the conditions on the energy spacing E such that several different properties of $G(\alpha)$ hold.

Lemma 13. *Consider $G(\alpha) = \alpha(\alpha-1) - \frac{B_\alpha}{B'_\alpha}$, where B_α, B'_α is defined in Eq. (E88) and (E90). Then the following holds:*
1) If $E(\beta_c - \beta_h)\tanh(\beta_c E/2) > 2$,

$$\exists\, 0 < \tau < 1 \text{ s.t.} \quad G(\alpha) < 0 \quad \forall\alpha \in (\tau, 1) \cup (1, \infty) \tag{E92}$$

2) If $E(\beta_c - \beta_h)\tanh(\beta_c E/2) < 2$,

$$\begin{aligned} \exists\, \underline{\alpha} > 1 \text{ s.t.} \qquad & G(\alpha) > 0 \quad \forall\alpha \in (0,1) \cup (1, \underline{\alpha}) \\ & G(\alpha) < 0 \quad \forall\alpha \in (\underline{\alpha}, \infty). \end{aligned} \tag{E93}$$

3) If $E(\beta_c - \beta_h)\tanh(\beta_c E/2) = 2$,

$$\begin{aligned} & G(\alpha) > 0 \quad \forall\alpha \in (0,1) \\ & G(\alpha) < 0 \quad \forall\alpha \in (1, \infty). \end{aligned} \tag{E94}$$

Proof. First we note that since $B_1 = 0$, therefore $G(1) = 0$. Let us also compute the derivative of $G(\alpha)$ w.r.t. α:

$$G'(\alpha) = 2\alpha - 1 - \frac{\cosh\left((-\beta_c/2 + \beta_h + (\beta_c - \beta_h)\alpha)E\right)}{\cosh(\beta_c E/2)}. \tag{E95}$$

Before we continue, there are several properties of the function $G'(\alpha)$ which we shall make use of. Firstly, note that $G'(1) = 0$, in other words, G' has a root at $\alpha = 1$. Also, $G'(\infty) = -\infty$ for any value of $E > 0,\ \beta_h > 0,\ \beta_c > \beta_h$ [1]. Also, since $2\alpha - 1$ is linear (and hence both convex and concave), while the $-\cosh$ function is strictly concave [2] , therefore the function $G'(\alpha)$ is *strictly concave*. This implies that the second derivative $G''(\alpha) = \frac{d^2G(\alpha)}{d\alpha^2}$ is strictly decreasing w.r.t. α.

The properties of $G'(\alpha)$ indicate that we can fully analyze the function by considering 3 different cases:

1. G' has two roots at $\alpha = \{\underline{a}, 1\}$, wherewhere $\underline{a} \in (-\infty, 1)$. This corresponds to the case $G''(1) < 0$.

2. G' has two roots at $\alpha = \{1, \overline{a}\}$, where $\overline{a} \in (1, \infty)$. This corresponds to the case $G''(1) > 0$.

3. G' has a single root at $\alpha = 1$. This corresponds to the case $G''(1) = 0$.

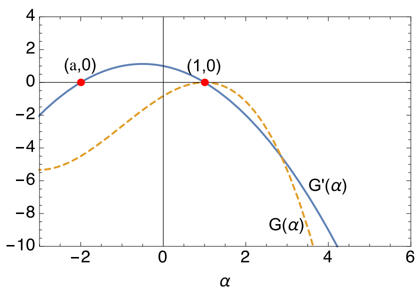

FIG. 7. $G''(1) < 0$.

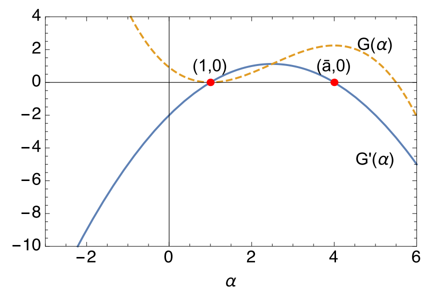

FIG. 8. $G''(1) > 0$.

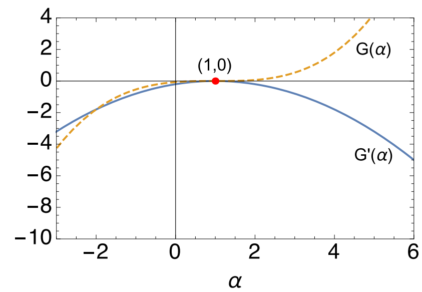

FIG. 9. $G''(1) = 0$.

FIG. 10. A convex function $G'(\alpha)$ and its corresponding $G(\alpha)$, for different values of $G''(\alpha)$.

We shall now consider these cases one by one. Suppose that

$$G''(1) = G''(\alpha)\Big|_{\alpha=1} = 2 - (\beta_c - \beta_h)E \tanh\left(\frac{\beta_c E}{2}\right) < 0, \tag{E96}$$

then $G''(\alpha) < 0$ for all $\alpha \in (1, \infty)$. Note that Eq. (E96) corresponds to the first condition in the lemma stated above.

This information about the second derivative $G''(\alpha)$ now allows us to conclude the following about $G(\alpha)$:

1. If for all $\alpha \in (1, \infty)$, $G''(\alpha) < 0$, then we know that $G'(\alpha) < 0$ holds for all $\alpha \in (1, \infty)$ too. Furthermore, this implies that $G(\alpha)$ is monotonically decreasing in the interval $(1, \infty)$ and therefore, $G(\alpha) < 0$ for all $\alpha \in (1, \infty)$.

2. $G''(1) < 0$ also implies that there exists an interval $(\tau, 1)$ such that $G'(1) > 0$ (See Fig. 7). And since $G(1) = 0$, this implies that within the interval $(\tau, 1)$, $G(\alpha) < 0$.

With this, we prove the first statement of the lemma.

Let us now analyze the second case, where $G''(1) > 0$. This implies that $G'(\alpha) > 0$ at least for some interval $\alpha \in (1, \overline{a})$, then $G'(\alpha)$ changes sign exactly once at $\alpha = \overline{a}$, and goes to $-\infty$. (Refer to Fig. 8). Also, recall that in the limit of $\alpha \to \infty$, G also goes to $-\infty$. Therefore, we conclude that there exists some $\overline{\alpha}$ such that

$$G(\alpha) \begin{cases} > 0 & \alpha \in (1, \overline{\alpha}) \\ < 0 & \alpha \in (\overline{\alpha}, \infty) \end{cases} \tag{E97}$$

With this, we prove the second statement of the lemma.

Finally, we look at the case where $G''(1) = 0$, and make the following observations:

1. Since the function $G'(\alpha)$ is concave, and since $G''(1) = 0$ implies that $\alpha = 1$ is an extremum point for the function $G'(\alpha)$, we know that it must also be the global maximum. Therefore, we know that for any $\alpha \neq 1$,$G'(\alpha) < 0$.

[1] This is due to the fact that 2α increases linearly w.r.t. α, while the cosh term increases exponentially.

[2] To be more precise; due to the concavity of $f(x) = -a\cosh(b + xc)$ for $a > 0$. This follows from the strict concavity of the cosh function, the invariancy of strict conc

2. Since for the interval $\alpha \in (-\infty, 1), G'(\alpha) < 0$ and we know that $G(1) = 0$, therefore we can deduce that for any $\alpha \in (-\infty, 1), G(\alpha) > 0$.
3. Since for the interval $\alpha \in (1, \infty), G'(\alpha) < 0$ and we know that $G(1) = 0$, therefore we can deduce that for any $\alpha \in (1, \infty), G(\alpha) < 0$.

With this, we prove the final statement of the lemma, and complete our proof. □

To summarize, in Lemma 13 we have identified conditions involving the energy gap of $\hat{H}_c$, and the temperatures β_h, β_c. Depending on whether these conditions are satisfied, we can describe the positivity/negativity of $G(\alpha)$ for different regimes of α. Comparing these different scenarios, we prove in Lemma 14 that for a quasi-static heat engine, the minimum of $\inf_{\alpha \geq \bar{\kappa}} \frac{\alpha B_\alpha}{\alpha - 1}$ is obtained only either at $\alpha = \bar{\kappa}$ or $\alpha = \infty$.

Lemma 14. *There exists some $0 \leq \nu < 1$ such that $\forall\, \kappa$ satisfying $\nu < \kappa < 1$, the following infimum is obtained at one of two points*

$$\inf_{\alpha \geq \kappa} \frac{\alpha B_\alpha}{\alpha - 1} = \inf \left\{ \lim_{\alpha \to \kappa} \frac{\alpha B_\alpha}{\alpha - 1}, \lim_{\alpha \to \infty} \frac{\alpha B_\alpha}{\alpha - 1} \right\} < \lim_{\alpha \to \beta} \frac{\alpha B_\alpha}{\alpha - 1} \quad \forall\, \beta \in (\kappa, \infty), \tag{E98}$$

where B_α is defined in Eq. (E88). Furthermore, if $E(\beta_c - \beta_h) \tanh(\beta_c E/2) \leq 2$, then we can set $\nu = 0$.

Proof. 1. If

$$\frac{d}{d\alpha} \frac{\alpha B_\alpha}{\alpha - 1} \begin{cases} > 0 \;\forall\, \alpha \in (0, 1) \cup (1, \overline{\alpha}) \text{ for some } \overline{\alpha} \geq 1 \\ < 0 \;\forall\, \alpha \in (\overline{\alpha}, \infty). \end{cases} \tag{E99}$$

then $\forall\, \kappa \in (0, 1)$,

$$\inf_{\alpha \geq \kappa} \frac{\alpha B_\alpha}{\alpha - 1} = \inf \left\{ \lim_{\alpha \to \kappa} \frac{\alpha B_\alpha}{\alpha - 1}, \lim_{\alpha \to \infty} \frac{\alpha B_\alpha}{\alpha - 1} \right\} < \lim_{\alpha \to \beta} \frac{\alpha B_\alpha}{\alpha - 1} \quad \forall\, \beta \in (\kappa, \infty). \tag{E100}$$

Recall from Eq. (E83) that

$$\frac{d}{d\alpha} \frac{\alpha B_\alpha}{\alpha - 1} = \frac{B'_\alpha}{(\alpha - 1)^2} G(\alpha), \tag{E101}$$

where $B'_\alpha > 0$, and we have derived some properties of $G(\alpha)$ in Lemma 13. In this proof, we apply Lemma 13 directly to consider the three scenarios detailed in Lemma 13.

First, consider the first statement of Lemma 13. If $E(\beta_c - \beta_h) \tanh(\beta_c E/2) > 2$, then $\exists\, 0 < t < 1$ s.t.

$$\frac{d}{d\alpha} \frac{\alpha B_\alpha}{\alpha - 1} < 0 \;\forall\, \alpha \in (t, 1) \cup (1, \infty) \tag{E102}$$

then by continuity of $\frac{\alpha B_\alpha}{\alpha - 1}$ in α, we conclude that $\forall\, \kappa$ satisfying $t < \kappa < 1$

$$\inf_{\alpha \geq \kappa} \frac{\alpha B_\alpha}{\alpha - 1} = \lim_{\alpha \to \infty} \frac{\alpha B_\alpha}{\alpha - 1} = \inf \left\{ \lim_{\alpha \to \kappa} \frac{\alpha B_\alpha}{\alpha - 1}, \lim_{\alpha \to \infty} \frac{\alpha B_\alpha}{\alpha - 1} \right\} < \lim_{\alpha \to \beta} \frac{\alpha B_\alpha}{\alpha - 1} \quad \forall\, \beta \in (\kappa, \infty). \tag{E103}$$

Next, consider the second and third statements of Lemma 13 jointly, where $E(\beta_c - \beta_h) \tanh(\beta_c E/2) \leq 2$. Note that both statements proved in Lemma 13 (namely, Eq. (E93) and (E94)) can be rewritten as the fact that there exists $\overline{\alpha} \geq 1$ s.t.

$$\frac{d}{d\alpha} \frac{\alpha B_\alpha}{\alpha - 1} \begin{cases} > 0 & \text{for } \alpha \in (0, 1) \cup (1, \overline{\alpha}) \\ < 0 & \text{for } \alpha \in (\overline{\alpha}, \infty). \end{cases} \tag{E104}$$

In fact, the third statement is simply a special case of the second, where $\overline{\alpha} = 1$. If Eq. (E104) holds, then $\forall\, \kappa \in (0, 1)$,

$$\inf_{\alpha \geq \kappa} \frac{\alpha B_\alpha}{\alpha - 1} = \inf \left\{ \lim_{\alpha \to \kappa} \frac{\alpha B_\alpha}{\alpha - 1}, \lim_{\alpha \to \infty} \frac{\alpha B_\alpha}{\alpha - 1} \right\} < \lim_{\alpha \to \beta} \frac{\alpha B_\alpha}{\alpha - 1} \quad \forall\, \beta \in (\kappa, \infty). \tag{E105}$$

By setting $\tau = 0$, we see that the statement of Lemma 14 is achieved.

Therefore, since we have analyzed all three cases stated in Lemma 13, we conclude that there always exists $\nu \in [0, 1)$ such that Eq. (E98) will always be satisfied $\forall\, \kappa \in (\nu, 1)$.

□

f. Main results: evaluating the efficiency

In this section, we derive the efficiency of quasi-static heat engines in the nano /quantum regime. We first need to define the quantity

$$\Omega := \min_{i\in\{1,\ldots,n\}} \frac{E_i(\beta_c-\beta_h)}{1+e^{-\beta_c E_i}}, \tag{E106}$$

where recall that E_i is the energy gap of the cold bath qubits, as described in Eq (E18) and the sentence right after it. Recall that n denotes the number of qubits in the cold bath, where $n\in\mathbb{Z}^+$ is any positive integer. Before stating the maximum efficiency, we will derive the efficiency as a function of $\bar{\kappa}$ defined in Lemma 11 (recall that this parameter is determined by the choice of ε). For simplicity, we will still consider the special case where $E_i = E$ for all i in Lemma 15, (i.e. all qubits of the cold bath are identical). Lemma 15 shows us that under the condition of extracting near perfect work, one can choose ε (and therefore $\bar{\kappa}$) such that a certain maximum efficiency value is achieved. The closer $\bar{\kappa}$ is to unity, the slower $\lim_{g\to 0^+}\Delta S/W$ converges to zero, and also the closer the efficiency is to the Carnot efficiency.

Using this lemma, we prove the achievability of the Carnot efficiency which depends on Ω. **This is the main result of our work, which is stated in Theorem 2.**

Lemma 15 (Quasi-static efficiencies as a function of $\bar{\kappa}$)**.** *For any $n\in\mathbb{Z}^+$ number of qubits, consider quasi-static heat engines (Def. 3) as a function of $\bar{\kappa}$ (defined in Lemma 11) which extract near perfect work (Def. 2). For any $\kappa\in(0,\infty)\backslash\{1\}$, define*

$$\gamma(\kappa) := \frac{\kappa B_\kappa}{\kappa-1} \tag{E107}$$

where B_κ is defined in Eq. (E88), while $\gamma(1)$ and $\gamma(\infty)$ are defined by taking the limits $\kappa\to 1,\infty$ respectively.
If $\Omega\leq 1$ (see Eq. (E106)):

1) There exists $\nu\in[0,1)$ such that for any $\bar{\kappa}\in(\nu,1]$ (and $\lim_{g\to 0^+}(\varepsilon\ln\varepsilon)/g = 0$ if $\bar{\kappa}=1$), the maximum efficiency is

$$\eta^{-1}(\bar{\kappa}) = 1+\frac{\beta_h}{\beta_c-\beta_h}\frac{\gamma(1)}{\gamma(\bar{\kappa})}+\Theta(f(g))+\Theta(g)+\Theta(\varepsilon), \tag{E108}$$

where $\gamma(1)\geq\gamma(\bar{\kappa})$ with equality iff $\bar{\kappa}=1$ and $\lim_{g\to 0+} f(g) = 0$.

2) The corresponding amount of work extracted is

$$W_{\rm ext}(\bar{\kappa}) = g\frac{n}{\beta_h}\left[\gamma(\bar{\kappa})+\Theta\left(f(g)\right)\right]. \tag{E109}$$

If $\Omega > 1$:

1) There exists $\nu'\in[0,1)$ such that for any $\bar{\kappa}\in(\nu',1]$ (and $\lim_{g\to 0^+}(\varepsilon\ln\varepsilon)/g = 0$ if $\bar{\kappa}=1$), the maximum efficiency is

$$\eta^{-1}(\bar{\kappa}) = 1+\frac{\beta_h}{\beta_c-\beta_h}\frac{\gamma(1)}{\gamma(\infty)}+\Theta(f(g))+\Theta(g)+\Theta(\varepsilon), \tag{E110}$$

where $\gamma(1)<\gamma(\infty)$.

2) The corresponding amount of work extracted is

$$W_{\rm ext}(\bar{\kappa}) = g\frac{n}{\beta_h}\left[\gamma(\infty)+\Theta\left(f(g)\right)\right] \tag{E111}$$

Proof. Firstly, let us begin by deriving the explicit form for $\gamma(1)$ and $\gamma(\infty)$:

$$\gamma(1) = \lim_{\alpha\to 1}\frac{\alpha}{\alpha-1}B_\alpha = \lim_{\alpha\to 1} B_\alpha+\alpha B'_\alpha = \frac{E^2(\beta_c-\beta_h)}{(1+e^{\beta_c E})^2}e^{\beta_c E}, \tag{E112}$$

where we have made use of the L'Hôspital rule. For $\alpha\to\infty$, since

$$\lim_{\alpha\to\infty} B_\alpha = \lim_{\alpha\to\infty}\frac{E}{1+e^{\beta_c E}}\frac{e^{\beta_h E}-e^{\beta_c E}e^{-\alpha(\beta_c-\beta_h)E}}{e^{\beta_h E}+e^{-\alpha(\beta_c-\beta_h)E}} = \frac{E}{1+e^{\beta_c E}},$$

therefore we have

$$\gamma(\infty) = \lim_{\alpha\to 1}\left(1+\frac{1}{\alpha-1}\right)\cdot B_\alpha = \frac{E}{1+e^{\beta_c E}}. \tag{E113}$$

By Lemma 14, we know that the infimum of $\gamma(\alpha)$ for $\alpha\in[\bar{\kappa},\infty)$ and $\bar{\kappa}\in(\nu,1]$ is either at $\alpha=\bar{\kappa}$ or $\alpha\to\infty$. Therefore, if we take the ratio of Eqs. (E112) and (E113) to be

$$\frac{\gamma(1)}{\gamma(\infty)} = \frac{E(\beta_c-\beta_h)}{1+e^{-\beta_c E}} = \Omega \leq 1, \tag{E114}$$

then $\gamma(\infty)\geq\gamma(1)>\gamma(\bar{\kappa})$, therefore the infimum of $\gamma(\alpha)$ for $\alpha\in[\bar{\kappa},\infty)$ and $\bar{\kappa}\in(\nu,1]$ has to be obtained at $\alpha=\bar{\kappa}$. Taking this into account and using the condition which is equivalent to that of near perfect work in Eq. (A11), we can use Lemma 14, to calculate the amount of work extracted:

$$W_{\text{ext}} = \inf_{\alpha\geq 0} W_\alpha = g\cdot\left[\inf_{\alpha>\bar{\kappa}}\frac{n}{\beta_h}\gamma(\bar{\kappa})+\Theta\left(f(g)\right)\right] = g\frac{n}{\beta_h}\left[\gamma(\bar{\kappa})+\Theta\left(f(g)\right)\right], \tag{E115}$$

where $\lim_{g\to 0^+} f(g)=0$. On the other hand, we can calculate ΔC, which is the change of average energy in the cold bath system, (recall this is done by Taylor expansion around $g=0$)

$$\Delta C = n\left(\langle E^2\rangle_{\beta_c}-\langle E\rangle_{\beta_c}^2\right)g+\Theta\left(g^2\right) = \frac{n\gamma(1)}{\beta_c-\beta_h}g+\Theta\left(g^2\right). \tag{E116}$$

Using Eq. (C15), we have $\Delta W=(1-\varepsilon)W_{\text{ext}}$. The (inverse) efficiency, according to the definition (C1), is thus

$$\begin{aligned}\eta^{-1}(\bar{\kappa}) &= 1+\frac{\Delta C}{W_{\text{ext}}}-\varepsilon = 1+\frac{n\gamma(1)/(\beta_c-\beta_h)g+\Theta\left(g^2\right)}{n\gamma(\bar{\kappa})g/\beta_h+\Theta\left(gf(g)\right)}-\varepsilon && \text{(E117)}\\ &= 1+\frac{\beta_h}{(\beta_c-\beta_h)}\frac{\gamma(1)}{\gamma(\bar{\kappa})}+\Theta(f(g))+\Theta(g)+\Theta(\varepsilon), && \text{(E118)}\end{aligned}$$

where we have used $\lim_{g\to 0^+} f(g)=0$ which is proven in Lemma 12. We will now investigate the efficiency when $\Omega>1$ is satisfied. Using $\Omega>1$ and Eq. (E114), we have that $\gamma(\infty)<\gamma(1)$. Thus from Lemma 14, due to continuity in $\bar{\kappa}$ of $\gamma(\bar{\kappa})$ we conclude that there exists a $\nu'\in[0,1)$ such that for any $\bar{\kappa}\in(\nu',1]$,

$$\inf_{\alpha\geq\bar{\kappa}}\gamma(\alpha)=\gamma(\infty). \tag{E119}$$

Therefore, since we are considering near perfect work, Eq. (A11) holds and we can use Lemma 12 to calculate the amount of work extracted

$$W_{\text{ext}} = \inf_{\alpha\geq 0} W_\alpha = g\cdot\left[\inf_{\alpha>\bar{\kappa}}\frac{n}{\beta_h}\gamma(\bar{\kappa})+f(g)\right] = g\frac{n}{\beta_h}\left[\gamma(\infty)+\frac{\beta_h}{n}f(g)\right], \tag{E120}$$

where $\lim_{g\to 0^+} f(g)=0$. Thus using the definition of inverse efficiency (Eq. (C1)), together with Eq. (E116), we have

$$\begin{aligned}\eta^{-1}(\bar{\kappa}) &= 1+\frac{\Delta C}{W_{\text{ext}}}-\varepsilon = 1+\frac{n\gamma(1)/(\beta_c-\beta_h)g+\Theta\left(g^2\right)}{n\gamma(\infty)g/\beta_h+\Theta\left(gf(g)\right)}-\varepsilon && \text{(E121)}\\ &= 1+\frac{\beta_h}{(\beta_c-\beta_h)}\frac{\gamma(1)}{\gamma(\infty)}+\Theta(f(g))+\Theta(g)+\Theta(\varepsilon), && \text{(E122)}\end{aligned}$$

where we have used $\lim_{g\to 0^+} f(g)=0$ which is proven in Lemma 12. □

We will now use Lemma 15 to conclude our main result of this letter.

Lemma 16. *Consider the case of near perfect work (Def. (2)) and all cold bath qubits are identical (i.e. $E_i=E$ for $i=1,\ldots,n$), then:*

1) If $\Omega \leq 1$ (see Eq. (E106)) the optimal achievable efficiency η_{max} (see Eq. (C4)) is the Carnot efficiency:

$$\eta_{\text{max}} = \left(1 + \frac{\beta_h}{\beta_c - \beta_h}\right)^{-1} \tag{E123}$$

What is more, this efficiency is only *achieved for quasi-static heat engines, i.e. $\eta_{\text{max}} = \eta_{\text{max}}^{\text{stat}}$ (see Eq. (C8)).*

2) If $\Omega > 1$ and the heat engine is quasi-static, then the optimal achievable efficiency is (see Eq. (C8))

$$\eta_{\text{max}}^{\text{stat}} = \left(1 + \frac{\beta_h}{\beta_c - \beta_h}\Omega\right)^{-1}. \tag{E124}$$

3) If $\Omega > 1$ the maximum achievable efficiency η_{max} (see Eq. (C4)), is strictly less that the Carnot efficiency.

Proof. In Lemma 5, we found that the Carnot Efficiency is an upper bound for the efficiency when we are extracting near perfect work. We also found that Eq. (A11) is satisfied iff we are extracting near perfect work. In Lemma 15, we derived the optimal achievable efficiency for quasi-static heat engines as a function of $\bar{\kappa}$ when Eq. (A11) is satisfied. By choosing $\bar{\kappa} < 1$ arbitrarily close to one, if $\Omega \leq 1$ is satisfied, we will thus achieve an efficiency arbitrarily close to the Carnot efficiency. Thus since the upper bound is equal to the lower bound, we prove part 1) of the Theorem. Part 2) of the Theorem follows from setting $\bar{\kappa} = 1$ in Lemma 5 when $\Omega > 1$ is satisfied. □

By making use of Lemma 15, one can generalize Lemma 16 to consider the more general case stated in A5 (at the begging of section E 2 c) where the cold bath still consists of qubits, however the energy gaps of the qubits can be arbitrary. For convenience, we re-write the general cold bath Hamiltonian here: for a set of variables $E_1 > 0, \cdots, E_n > 0$,

$$\hat{H}_{\text{Cold}} = \sum_{k=1}^{n} \mathbb{1}^{\otimes(k-1)} \otimes \hat{H}_{\text{c}}^k \otimes \mathbb{1}^{\otimes(n-k)}, \quad \text{where} \quad \hat{H}_{\text{c}}^k = E_k |E\rangle\langle E|, \tag{E125}$$

Under the more general form of the cold bath Eq. (E125), we have the following theorem.

Theorem 2. *[Quantum/Nano heat engine efficiency] Consider the case of near perfect work (Def. (2)), when the cold bath consists of multiple qubits with energy gaps $\{E_i\}_{i=1}^n$.*

1) If $\Omega \leq 1$ (see Eq. (E106)) the optimal achievable efficiency η_{max} (see Eq. (C4)) is the Carnot efficiency:

$$\eta_{\text{max}} = \left(1 + \frac{\beta_h}{\beta_c - \beta_h}\right)^{-1} \tag{E126}$$

What is more, this efficiency is only *achieved for quasi-static heat engines, i.e. $\eta_{\text{max}} = \eta_{\text{max}}^{\text{stat}}$ (see Eq. (C8)).*

2) If $\Omega > 1$ and the heat engine is quasi-static, then the optimal achievable efficiency is (see Eq. (C8))

$$\eta_{\text{max}}^{\text{stat}} = \left(1 + \frac{\beta_h}{\beta_c - \beta_h}\Omega\right)^{-1}. \tag{E127}$$

3) If $\Omega > 1$ the maximum achievable efficiency η_{max} (see Eq. (C4)), is strictly less that the Carnot efficiency.

4) Allowing for correlations between the final state of the battery and cold bath cannot improve the efficiencies achieved in 1), 2) and 3) above.

Proof. 1) is relatively simple to prove: as long as there exists a qubit with energy E_i such that $\frac{E_i(\beta_c - \beta_h)}{1+e^{-\beta_h E_i}} \leq 1$, one way to achieve Carnot efficiency is to simply disregard the rest of the cold bath, and act only on such qubits. The result is a simple application of 1) in Lemma 16. This strategy might not be optimal in terms of work extracted, but it is sufficient for our proof.

For 2) and 3), suppose that $\Omega > 1$. Since Ω is a monotonic function of E, we conclude that for all E_i where $1 \leq i \leq n$, $\Omega_i := \frac{E_i(\beta_c - \beta_h)}{1+e^{-\beta_h E_i}} > 1$. By Lemma 15, we see that this implies that the work extractable for all the individual qubits (which is an optimization problem over all $\alpha \geq 0$) is obtained at $\alpha \to \infty$. In general, considering the qubits collectively does not mean that the collective W_{ext} is additive. This is because the minima of two functions is not necessarily the minima of these individual functions added together, as illustrated in the l.h.s. and middle diagrams of Figure. 11. However, (as illustrated on r.h.s. diagram of Figure. 11), when all the functions have their minima at the same value, then the collective minima is also obtained at that value.

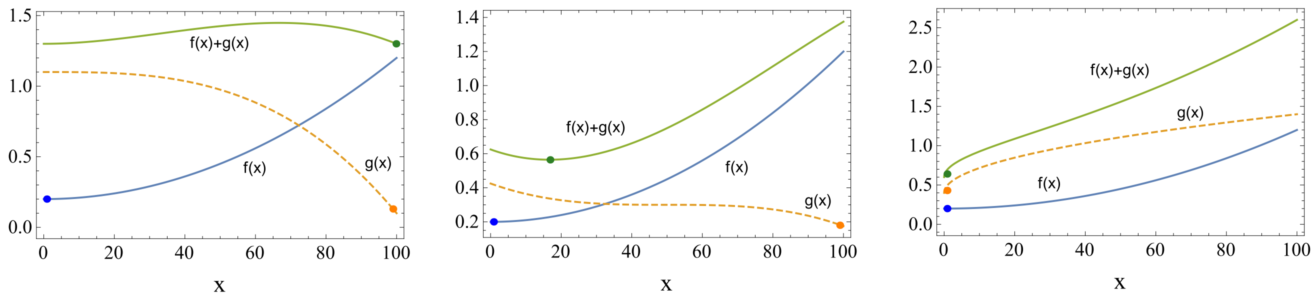

FIG. 11. Illustration of the minima of two individual functions $f(x), g(x)$ and minima of $f(x)+g(x)$.

Next, we show that no matter which subset of qubits $\mathcal{S}$ one picks, Carnot efficiency cannot be achieved. We begin by introducing the notation $\gamma_i(\alpha)$, where $\gamma_i(\alpha)$ is defined similarly with $\gamma(\alpha)$ in Eq. (E107) and (E88), and the index i indicates that E is substituted by E_i in Eq. (E88). Furthermore, recall that from Eq. (E114), $\Omega_i > 1$ is equivalent to $\gamma_i(1) > \gamma_i(\infty)$. Now, consider any subset of qubit indices $\mathcal{S}$, the amount of extractable work (as a function of g) is

$$W_{\text{ext}}^{\mathcal{S}} = \frac{g}{\beta_h}\left[\sum_{i\in\mathcal{S}}\gamma_i(\infty) + f(g)\right], \tag{E128}$$

where $\lim_{g\to 0+} f(g) = 0$.

On the other hand, we have that ΔC depends on the individual reduced qubit states, since there are no interaction terms in $\hat{H}_{\text{Cold}}$. Therefore, similar to Eq. (E116),

$$\Delta C^{\mathcal{S}} = \frac{g}{\beta_c - \beta_h}\sum_{i\in\mathcal{S}}\gamma_i(1) + \Theta\left(g^2\right). \tag{E129}$$

Following the same proof in Eq. (E121) Lemma 15,

$$\eta^{-1}(\bar{\kappa}) = 1 + \frac{\Delta C}{W_{\text{ext}}} - \varepsilon = 1 + \frac{\beta_h}{\beta_c - \beta_h}\frac{\sum_{i\in\mathcal{S}}\gamma_i(1)}{\sum_{i\in\mathcal{S}}\gamma_i(\infty)} + \Theta(g) + \Theta(\varepsilon). \tag{E130}$$

As we have observed before, the inverse of the Carnot efficiency $\eta_C^{-1} = 1 + \frac{\beta_h}{\beta_c - \beta_h}$. Furthermore, notice that by Eq. (E114), the condition $\Omega_i > 1$ implies that $\gamma_i(1) > \gamma_i(\infty)$. Since $\Omega_i > 1$ is true for all $1 \le i \le n$, therefore $\frac{\sum_{i\in\mathcal{S}}\gamma_i(1)}{\sum_{i\in\mathcal{S}}\gamma_i(\infty)} > 1$.

Lastly, part 4) is proven in Section F. □

Suppose n is large. Then since we have a spectrum which looks like a quasi-continuum: the full range of the spectrum is very large, compared to the individual energy gaps. One expects that in such a case, baths are of high temperature (small values of β), then the effects of quantization should give us the classical observations of being able to achieve Carnot always. This can be seen, that for $E_{\text{min}} = \min\limits_{i\in\{1,\cdots,n\}} E_i$, if the quantities $\beta_c E_{\text{min}}, \beta_h E_{\text{min}} \ll 1$, then

$$\Omega = \frac{E_{\text{min}}(\beta_c - \beta_h)}{1 + e^{-\beta_c E_{\text{min}}}} \le E_{\text{min}}(\beta_c - \beta_h) \ll 1. \tag{E131}$$

Whenever $\Omega \le 1$, we know that Carnot efficiency is achievable.

3. Running the heat engine for many cycles quasi-statically

We have so far proven that a heat engine can achieve the Carnot efficiency when $\Omega \le 1$. However, as like with macroscopic heat engines, this can only be achieved when the heat engine runs quasi-statically. Macroscopic heat engines can then extract a finite amount of work by running the heat engine over many cycles (in fact, over any infinite number of cycles if they want to obtain the Carnot efficiency in order to run quasi-statically). The following lemma, shows that when $\Omega \le 1$, a nano-scale heat engine with a machine that runs over infinitely many cycles can also achieve the Carnot efficiency, while extracting any finite amount of work W with vanishing entropy increase in the battery.

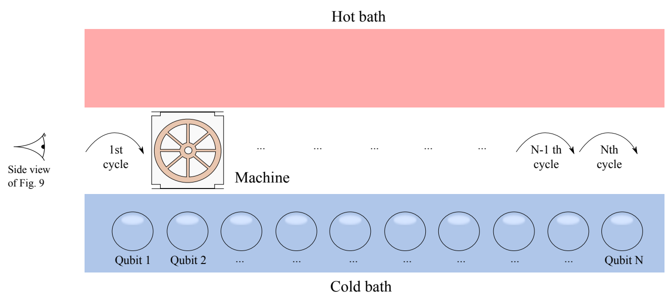

FIG. 12. Depiction (top view) of a heat engine comprising of a hot bath, a cold bath consisting of n identical qubits, a machine and a battery. In each cycle, the machine interacts specifically with one qubit from the cold bath, together with the hot bat and battery. After the end of one cycle, the machine is returned to its original state, and acts on a different qubit in the cold bath.

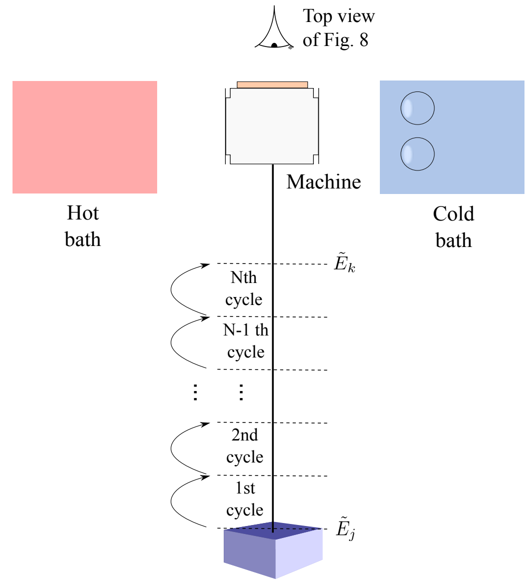

FIG. 13. Side view of the heat engine. After each cycle of the machine, the battery, depicted here as a weight moves upward by a small amount. After N machine cycles, it has been lifted from its original position $\left|\tilde{E}_j\right\rangle$ to a final state that has most of its weight on $\left|\tilde{E}_k\right\rangle$.

For simplicity, we will work with the case in which the quasi-continuum battery has a part of its spectrum equal to that of at least N qubits, each with an energy gap W_{ext}. We work within this subspace. We will run a heat engine between a hot bath, cold bath using a machine which performs N cyclic cycles. Let $\tilde{E}_j$ and $\tilde{E}_k$ be the smallest and largest energy eigenvalues within this subspace respectively. We let the initial state of the battery be

$$\rho_{\text{W}}^0 = |\tilde{E}_j\rangle\langle\tilde{E}_j|, \tag{E132}$$

$\hat{H}_{\text{W}}|\tilde{E}_j\rangle = \tilde{E}_j|\tilde{E}_j\rangle$ while we wish the final state of the battery to be of the form

$$\rho_{\text{W}}^1 = r|\tilde{E}_k\rangle\langle\tilde{E}_k| + (1-r)\,\rho_\psi, \tag{E133}$$

where $\hat{H}_{\text{W}}|\tilde{E}_k\rangle = \tilde{E}_k|\tilde{E}_k\rangle$, ρ_ψ is some orthogonal state to $|\tilde{E}_k\rangle$ and the value of the probability r is to be specified in the following lemma. We will define the amount of work extracted from the machine for N cycles

$$W_{\text{cyc}} := \tilde{E}_k - \tilde{E}_j. \tag{E134}$$

For simplicity, we will consider the case that the cold bath consists of n identical qubits with $\Omega \leq 1$, and during each cycle the machine interacts with one qubit from the cold bath. The running of the heat engine is depicted in Fig. 12 and 13.

Corollary 2. *[Many quasi-static heat engine cycles] Let W be the finite amount of work we wish to extract. Then for all $W > 0$ and $\delta > 0$ there exists an n identical qubit cold bath (with $\Omega \leq 1$) and an $N \in \mathbb{N}^+$ number of machine cycles with $n \geq N$ such that:*

1) $\eta_c \geq \eta \geq \eta_c - \delta$, *where the efficiency η is the efficiency per cycle and is defined by Eq. (C1), and $\eta_c = 1 - \beta_h/\beta_c$ is the Carnot efficiency,*

2) $W_{\text{cyc}} \geq W - \delta$,

3) $S(\rho_{\text{W}}^0) = 0$, $S(\rho_{\text{W}}^1) \leq \delta$, *and*

4) $r \geq 1-\delta$.

whats more, $\delta \to 0$ *as* $N \to +\infty$.

Proof. Since in the qubit subspace, the spectrum is that of at least N qubits, we can write the initial state in the form

$$\rho_{\mathrm{W}}^{0} = |E_j\rangle\langle E_j|^{\otimes N}, \tag{E135}$$

with $\hat{H}_{\mathrm{W}}\,|E_j\rangle^{\otimes N} = \tilde{E}_j\,|E_j\rangle^{\otimes N}$. We can now apply the heat engine results of Lemma 15 to the setup. Namely, we can apply the results of one cycle to each of the qubit subspaces of the battery in parallel. From Lemma 15 we conclude that this can be achieved with an efficiency given by Eq. (E108) and extract an amount of work per qubit/cycle given by Eq. (E109). For simplicity, we will run the heat engine using one qubit of the cold bath at a time. The final state of the battery is thus

$$\rho_{\mathrm{W}}^{1} = \left[(1-\varepsilon)\,|E_k\rangle\langle E_k| + \varepsilon\,|E_j\rangle\langle E_j|\right]^{\otimes N}. \tag{E136}$$

Noting that $|\tilde{E}_k\rangle\langle\tilde{E}_k| = |E_k\rangle\langle E_k|^{\otimes N}$ by definition, Eq. (E136) can be written as

$$\rho_{\mathrm{W}}^{1} = (1-\varepsilon)^N|\tilde{E}_k\rangle\langle\tilde{E}_k| + \left[1-(1-\varepsilon)^N\right]\,\rho_\psi. \tag{E137}$$

with ρ_ψ orthogonal to $|\tilde{E}_k\rangle$. From Eq. (E134) it follows,

$$W_{\mathrm{cyc}} = N W_{\mathrm{ext}} = \frac{Ng}{\beta_h}\left[\gamma(\bar{\kappa}) + \Theta\left(f(g)\right)\right], \tag{E138}$$

where in the last line we have used Eq. (E109). We now set

$$N = N(g) = \frac{\beta_h}{\gamma(\bar{\kappa})}\frac{W}{g} \tag{E139}$$

for all $g>0$ satisfying the constraint $N(g) \in \mathbb{N}^+$. For any positive constant $\frac{\beta_h W}{\gamma(\bar{\kappa})} > 0$, one can always consider the values of $\frac{\beta_h W}{\gamma(\bar{\kappa})} > g > 0$ so that $N(g)$ is large. This constraint imposes $g = \beta_h W/(\gamma(\bar{\kappa})\,N)$, where N has to be an integer. Therefore, g now belongs to a subset of the positive real line, rather than the positive real line itself as previously. However, since g monotonically decreases to zero as N increases to infinity, we can still take the limit $g \to 0^+$ as before. Thus achieving

$$W_{\mathrm{cyc}} = W + \Theta\left(f(g)\right). \tag{E140}$$

Since $\lim_{g\to 0^+} f(g) = 0$, we conclude part 2) of Corollary 2. For the entropy of the final state of the battery we have

$$S(\rho_{\mathrm{W}}^{1}) = NS\left((1-\varepsilon)\,|E_k\rangle\langle E_k| + \varepsilon\,|E_j\rangle\langle E_j|\right) = \frac{\beta_h W}{\gamma(\bar{\kappa})}\frac{(1-\varepsilon)\ln(1-\varepsilon)+\varepsilon\ln\varepsilon}{g} = \Theta\left(\frac{\varepsilon\ln\varepsilon}{g}\right). \tag{E141}$$

As stated above the efficiency is given by Eq. (E108), and thus we can always choose $\bar{\kappa} \in (0,1)$, and g (recall $\epsilon \to 0^+$ as $g \to 0^+$) such that 1) in Corollary 2 is satisfied. Furthermore, recall from the proof of Lemma 12 that

$$\lim_{g\to 0^+}\frac{\varepsilon\ln\varepsilon}{g} = 0, \tag{E142}$$

for all $\bar{\kappa} \in (0,1)$. Thus, from Eq. (E141) we conclude that 3) in Corollary 2. We will now prove part 4) of the Corollary. From Eq. (E137) and part 4) of the Corollary, we can identify $r = (1-\varepsilon)^N$. We thus study the limit

$$\lim_{g\to 0^+}(1-\varepsilon)^N = \left(\lim_{g\to 0^+}(1-\varepsilon)^{1/g}\right)^{\beta_h W/\gamma(\bar{\kappa})} = \left(\lim_{g\to 0^+}\left(\underbrace{(1-\varepsilon)^{1/\varepsilon}}_{\to\,\mathrm{e}}\right)^{\varepsilon/g}\right)^{\beta_h W/\gamma(\bar{\kappa})} = 1, \tag{E143}$$

where going to the last line, we have used that fact that Eq. (E142) implies that $\varepsilon/g \to 0$ as $g \to 0^+$. We thus conclude part 4) of the corollary. □

Thus by choosing $\delta > 0$ sufficiently small in Corollary 2, we can extract any finite amount of work with an arbitrarily small entropy contribution with an efficiency arbitrarily close to the Carnot efficiency as long as $\Omega \leq 1$.

F. Extensions to the setup

Arguably, one may think that the inability to always achieve the Carnot efficiency in the nano regime is due to some subtlety of our setup (even though we have shown that according to the standard free energy one can always achieve the Carnot efficiency with our setup). For such reasons, in the next few sections we show that even under more general conditions than those laid out in Section A, one still cannot achieve the Carnot efficiency when $\Omega > 1$.

In Section F 1, we show that allowing for correlations between the final state of the battery and cold bath (and/or the finite dimensional machine) does not allow us to achieve the Carnot efficiency. The main result is Theorem 3.

In Section F 2, we show that allowing for the battery to be *any* state with trace distance ε from $|E_k\rangle\langle E_k|_{\mathrm{W}}$ cannot allow us to achieve the Carnot efficiency when $\Omega > 1$. This shows that whenever we are unable to achieve the Carnot efficiency, it is not a artificial defect from an overly specified battery model. The main result is Theorem 4.

1. Final correlations between battery, cold bath, and machine

In Section E 2 a, we stated that the final state of the heat engine after tracing out the hot bath was of the form

$$\mathrm{tr}_{\mathrm{Hot}}(\rho^1_{\mathrm{ColdHotMW}}) = \rho^1_{\mathrm{Cold}} \otimes \rho^1_{\mathrm{M}} \otimes \rho^1_{\mathrm{W}} \tag{F1}$$

where $\rho^1_{\mathrm{W}} = \varepsilon|E_j\rangle\langle E_j|_{\mathrm{W}} + (1-\varepsilon)|E_k\rangle\langle E_k|_{\mathrm{W}}$, i.e. the final state of the charged battery was a tensor product with the cold bath. We also demanded that the heat engine is cyclic i.e. that $\rho^1_{\mathrm{M}} = \rho^0_{\mathrm{M}}$. In this section, we show that if one allows for the final state of the battery, cold bath and machine to become correlated[3], one still cannot achieve the Carnot efficiency when $\Omega > 1$. That is to say, in this section we allow the final state to be

$$\mathrm{tr}_{\mathrm{Hot}}(\rho^1_{\mathrm{ColdHotMW}}) = \rho^1_{\mathrm{ColdMW}} \tag{F2}$$

with only two natural constrains, namely that our heat engine actually extracts work, i.e. that

$$\rho^1_{\mathrm{W}} = \varepsilon|E_j\rangle\langle E_j|_{\mathrm{W}} + (1-\varepsilon)|E_k\rangle\langle E_k|_{\mathrm{W}}, \tag{F3}$$

as before, and also that the heat engine is still cyclic, i.e.

$$\rho^1_{\mathrm{M}} = \rho^0_{\mathrm{M}}. \tag{F4}$$

Throughout this section, (unless stated otherwise) we will write ρ^1_{ColdMW} to refer to any generic tripartite quantum state on the cold bath, machine and battery satisfying Eqs. (F3) and (F4).

- In Section F 1 a, we first define the generalized efficiency where one is allowed to consider correlated final states. We see that although this may potentially affect the amount of extractable work W_{ext}, the amount of heat change in the bath remains the same, by making use of energy conservation and the fact that the global Hamiltonian $H_{\mathrm{ColdHotMW}}$ does not contain interaction terms between subsystems.
- In Section F 1 b, we make use of the generalized second law when $\alpha = 1$ (which is also the macroscopic second law), in order to show that final correlations still do not allow the surpassing of Carnot efficiency. This can be proven by noting that the von Neumann entropy is subadditive, and the result is summarized in Lemma 19. A proof sketch can be found in the beginning of Section F 1 b.
- In Section F 1 c, we turn to the case where $\Omega > 1$, where without final correlations it is shown in Theorem 2 that Carnot efficiency cannot be achieved.

a. Defining the generalized efficiency

Recall that before (see Section C 2), we have shown in Eq. (C17) that if the following assumptions hold:

(i) the final reduced state of the battery ρ^1_{W} is fixed by Eq. (A6),

---

[3] Recall that the final state of the cold bath, machine and battery are already allowed to become correlated with the hot bath

(ii) the state of the machine is preserved,

(iii) the final state is of tensor product form, i.e. $\rho^1_{\text{ColdMW}} = \rho^1_{\text{Cold}} \otimes \rho^1_{\text{M}} \otimes \rho^1_{\text{W}}$,

then the efficiency for a particular transformation $\rho^0_{\text{ColdHotMW}} \to \rho^1_{\text{ColdHotMW}}$ simplifies to being only an explicit function of ρ^1_{Cold} instead of the global final state. This simplified expression of the efficiency in Eq. (C17) is then used to evaluate, for example, $\eta^{\text{mac}}(\rho^1_{\text{Cold}})$ in Eq. (C5). Since we now drop Assumption (iii) for the final state being uncorrelated, the efficiency and the work extracted W_{ext} will now depend on the tripartite final state ρ^1_{ColdMW} instead.

Therefore, let us first write a generalized expression for the maximum efficiency corresponding to a transition $\rho^0_{\text{ColdHotMW}} \to \rho^1_{\text{ColdHotMW}}$ via the unitary operator $U(t)$ in this generalised setting:

$$\eta^{\text{qm}}(\rho^1_{\text{ColdMW}}) := \sup_{W_{\text{ext}}} \eta(\rho^1_{\text{Cold}}, W_{\text{ext}}) \quad \text{s.t.} \quad \text{tr}_{\text{Hot}}[U(t)\rho_{\text{ColdHotMW}}U(t)^\dagger] = \rho^1_{\text{ColdMW}}, \tag{F5}$$

$$[U(t), \hat{H}] = 0, \tag{F6}$$

$$\rho^1_{\text{W}} = \varepsilon|E_j\rangle\langle E_j|_{\text{W}} + (1-\varepsilon)|E_k\rangle\langle E_k|_{\text{W}}, \tag{F7}$$

$$\rho^1_{\text{M}} = \rho^0_{\text{M}}. \tag{F8}$$

See Fig (1) in main text for a definition of the other quantities appearing in Eq. (F5). Recall that the definition of η is given by $\eta = W_{\text{ext}}/\Delta H$ as in Eq. (C1). In Section C 2 we showed that this can be simplified to

$$\eta = (1 - \varepsilon + \Delta C/W_{\text{ext}})^{-1}, \tag{F9}$$

where $\Delta C = \Delta C(\rho^1_{\text{Cold}})$. This equation holds under Assumption (i) and (ii), together with the fact that the global Hamiltonian does not contain interaction terms between both baths, battery, and machine. Since the derivation of Eq. (F9) does not require Assumption (iii), it still holds for a general tripartite final state ρ^1_{ColdMW}. However, dropping Assumption (iii) may potentially allow for larger values of W_{ext}, and therefore subsequently might affect η^{qm}. For this reason we write $\eta^{\text{qm}} = \eta^{\text{qm}}(\rho^1_{\text{ColdMW}})$ to remind ourselves that it is a function of the entire final state ρ^1_{ColdMW}.

We have written $\eta = \eta(\rho^1_{\text{Cold}}, W_{\text{ext}})$ to explicitly show the W_{ext} dependency of η. Although not written explicitly in Eq. (F5), we should remember that $U(t), \rho^0_{\text{M}}, \hat{H}_{\text{Hot}}$ and $\hat{H}_{\text{M}}$ are arbitrary, other than satisfying condition **(A.4)** in Section A. As such, by maximizing η over W_{ext}, these quantities will accommodate their optimal values to maximize $\eta^{\text{qm}}(\rho^1_{\text{ColdMW}})$[4]. Throughout this section, we analyze Eq. (F5) only in the case of *near perfect work* (Def. (2)) since the proof that perfect work is not possible (see Lemma 8) also applies to Eq. (F5)[5].

For the purpose of our proofs, we need to define a new family of intermediate efficiencies. They provide the maximum possible efficiency, when considering only a particular instance $\alpha \geq 0$ of the generalized second laws. For any $\alpha \in [0, \infty)$, let us denote

$$\eta^{\text{qm}}_\alpha(\rho^1_{\text{ColdMW}}) = \sup_{W_{\text{ext}}} \eta(\rho^1_{\text{Cold}}, W_{\text{ext}}) \quad \text{s.t.} \quad F_\alpha(\tau^0_{\text{Cold}} \otimes \rho^0_{\text{M}} \otimes \rho^0_{\text{W}}, \tau^h_{\text{ColdMW}}) \geq F_\alpha(\rho^1_{\text{ColdMW}}, \tau^h_{\text{ColdMW}}), \tag{F10}$$

$$\text{tr}(\hat{H}_t \rho^0_{\text{ColdHotMW}}) = \text{tr}(\hat{H}_t \rho^1_{\text{ColdHotMW}}), \tag{F11}$$

$$\rho^1_{\text{W}} = \varepsilon|E_j\rangle\langle E_j|_{\text{W}} + (1-\varepsilon)|E_k\rangle\langle E_k|_{\text{W}}, \tag{F12}$$

$$\rho^1_{\text{M}} = \rho^0_{\text{M}}. \tag{F13}$$

See Eq. (B8) for definition of F_α. We denote $\eta^{\text{qm}}_\infty = \lim_{\alpha \to \infty} \eta^{\text{qm}}_\alpha$. The condition Eq. (F11), is always satisfied when all the second laws are satisfied. We add the condition as a constraint here, since we will need it in order to write the efficiency η in the form of Eq. (F9).

b. *Final correlations do not allow the surpassing of Carnot efficiency*

In this section, we first show that Carnot efficiency cannot be surpassed even when we allow arbitrary final correlations in the final state ρ^1_{ColdMW}. This can be done in the following steps:

---

[4] This is an advantage, since it rules out cases such as when the Hamiltonian does not support a thermal state (e.g. when the corresponding thermal state's partition function diverges). In this section we consider any cold bath Hamiltonian $\hat{H}_{\text{Cold}}$ that satisfies **(A.6)** in Section A (i.e. finite dimensional). As such it will always h

[5] For the sake

1. Using the definitions of generalized efficiency (allowing correlations) in Eq. (F5) and generalized intermediate efficiencies in Eq. (F10), we prove an inequality between $\eta^{\mathrm{qm}}(\rho^1_{\mathrm{ColdMW}})$ and $\eta^{\mathrm{qm}}_\alpha(\rho^1_{\mathrm{ColdMW}})$, for all $\alpha \geq 0$. This is done in Lemma 17. From this, we also conclude that $\eta^{\mathrm{qm}}(\rho^1_{\mathrm{ColdMW}}) \leq \eta^{\mathrm{qm}}_1(\rho^1_{\mathrm{ColdMW}})$.

2. On the other hand, we show that for any final state of the cold bath, machine and battery ρ^1_{ColdMW}, the generalized intermediate efficiency for $\alpha = 1$ only increases, if we consider the tensor product of the marginals ρ^1_{ColdMW}. In other words, $\eta^{\mathrm{qm}}_1(\rho^1_{\mathrm{ColdMW}}) \leq \eta^{\mathrm{qm}}_1(\rho^1_{\mathrm{Cold}} \otimes \rho^1_{\mathrm{W}} \otimes \rho^1_{\mathrm{M}})$. One can intuitively see why this is true: it comes from the fact that the von Neumann entropy is subadditive, therefore the final state $\rho^1_{\mathrm{Cold}} \otimes \rho^1_{\mathrm{W}} \otimes \rho^1_{\mathrm{M}}$ contains more entropy than ρ^1_{ColdMW}. Therefore according to the $\alpha = 1$ second law, one can potentially draw more work by going to the state $\rho^1_{\mathrm{Cold}} \otimes \rho^1_{\mathrm{W}} \otimes \rho^1_{\mathrm{M}}$ instead of a correlated state ρ^1_{ColdMW}.

3. Since the argument for $\eta^{\mathrm{qm}}_1(\rho^1_{\mathrm{Cold}} \otimes \rho^1_{\mathrm{W}} \otimes \rho^1_{\mathrm{M}})$ is of tensor product form, Assumption (iii) holds as before, and therefore the efficiency only depends on the final state of the cold bath ρ^1_{Cold}. This means that Eq. (F10) for $\alpha = 1$ reduces to Eq. (C5). Lastly, by using Lemma 22, this allows us to further show in Lemma 19 that even by allowing correlations in ρ^1_{ColdMW}, the efficiency $\eta^{\mathrm{qm}}(\rho^1_{\mathrm{ColdMW}})$ can never surpass the Carnot value.

Firstly, let us fix the following notation: for an R-partite state $\rho_{\mathrm{A}_1\mathrm{A}_2\ldots\mathrm{A}_R}$, define the uncorrelated counterpart

$$\underline{\rho_{\mathrm{A}_1\mathrm{A}_2\ldots\mathrm{A}_R}} := \bigotimes_{i=1}^{R} \rho_{\mathrm{A}_i}. \tag{F14}$$

Comparing $\rho_{\mathrm{A}_1\mathrm{A}_2\ldots\mathrm{A}_R}$ and $\underline{\rho_{\mathrm{A}_1\mathrm{A}_2\ldots\mathrm{A}_R}}$, one will see that each subsystem has the same reduced state, but the global state is different. Another useful thing is to note that if one is given a Hamiltonian which does not contain any interaction terms between each subsystem, i.e.

$$\hat{H}_{\mathrm{A}_1\mathrm{A}_2\ldots\mathrm{A}_R} = \sum_{i=1}^{R} \mathbb{1}_{\mathrm{A}_1} \otimes \cdots \hat{H}_{\mathrm{A}_i} \cdots \mathbb{1}_{\mathrm{A}_R}, \tag{F15}$$

then we may conclude that

$$\mathrm{tr}(\hat{H}_{\mathrm{A}_1\mathrm{A}_2\ldots\mathrm{A}_R}\rho_{\mathrm{A}_1\mathrm{A}_2\ldots\mathrm{A}_R}) = \sum_{i=1}^{R} \mathrm{tr}(\hat{H}_{\mathrm{A}_i}\rho_{\mathrm{A}_i}) = \sum_{i=1}^{R} \mathrm{tr}(\hat{H}_{\mathrm{A}_i}\underline{\rho_{\mathrm{A}_i}}) = \mathrm{tr}(\hat{H}_{\mathrm{A}_1\mathrm{A}_2\ldots\mathrm{A}_R}\underline{\rho_{\mathrm{A}_1\mathrm{A}_2\ldots\mathrm{A}_R}}). \tag{F16}$$

Lemma 17. *For all $\alpha \geq 0$ and all states $\rho^1_{\mathrm{ColdHotMW}}$,*

$$\eta^{\mathrm{qm}}(\rho^1_{\mathrm{ColdMW}}) \leq \eta^{\mathrm{qm}}_\alpha(\rho^1_{\mathrm{ColdMW}}), \tag{F17}$$

where η^{qm} and $\eta^{\mathrm{qm}}_\alpha$ are defined in Eqs. (F5) and (F10) respectively.

Proof. For every $\alpha \geq 0$, Eq. $F_\alpha(\tau^0_{\mathrm{Cold}} \otimes \rho^0_{\mathrm{M}} \otimes \rho^0_{\mathrm{W}}, \tau^h_{\mathrm{ColdMW}}) \geq F_\alpha(\rho^1_{\mathrm{ColdMW}}, \tau^h_{\mathrm{ColdMW}})$ in Eq. (F10) is a necessary condition for the transformation $\rho^0_{\mathrm{ColdMW}} \to \rho^1_{\mathrm{ColdMW}}$ to occur under an energy preserving unitary with the aid of a catalyst [5]. Energy preserving unitaries also preserve the average energy and thus the Eq. $\mathrm{tr}(\hat{H}_t\rho^0_{\mathrm{ColdHotMW}}) = \mathrm{tr}(\hat{H}_t\rho^1_{\mathrm{ColdHotMW}})$ in Eq. (F10) is also a necessary condition. If a unitary $U(t)$ satisfies the conditions in Eq. (F5), then by the second laws it satisfies Eq. (F10) for any particular $\alpha \geq 0$. As a consequence of these observations, the set of allowed unitaries $U(t)$ in Eq. (F5) is a subset of allowed unitaries facilitating the catalytic thermal operation which transforms ρ^0_{ColdMW} to ρ^1_{ColdMW} in Eq. (F10). $\square$

Lemma 18. *For any final state ρ^1_{ColdMW}, consider the quantity $\eta^{\mathrm{qm}}_1(\rho^1_{\mathrm{ColdMW}})$ defined in Eq. (F10). Consider the optimization problem*

$$a(\rho^1_{\mathrm{ColdMW}}) := \sup_{W_{\mathrm{ext}}} \eta(\rho^1_{\mathrm{Cold}}, W_{\mathrm{ext}}) \quad \text{s.t.} \;\; F_1(\tau^0_{\mathrm{Cold}} \otimes \rho^0_{\mathrm{M}} \otimes \rho^0_{\mathrm{W}}, \tau^h_{\mathrm{ColdMW}}) = F_1(\rho^1_{\mathrm{ColdMW}}, \tau^h_{\mathrm{ColdMW}}), \tag{F18}$$

$$\mathrm{tr}(\hat{H}_t\rho^0_{\mathrm{ColdHotMW}}) = \mathrm{tr}(\hat{H}_t\rho^1_{\mathrm{ColdHotMW}}), \tag{F19}$$

$$\rho^1_{\mathrm{W}} = \varepsilon|E_j\rangle\langle E_j|_{\mathrm{W}} + (1-\varepsilon)|E_k\rangle\langle E_k|_{\mathrm{W}}, \tag{F20}$$

$$\rho^1_{\mathrm{M}} = \rho^0_{\mathrm{M}}. \tag{F21}$$

Then, $\eta^{\mathrm{qm}}_1(\rho^1_{\mathrm{ColdMW}}) = a(\rho^1_{\mathrm{ColdMW}})$.

Proof. We begin by noting that the free energy F_1 can be written as

$$F_1(\rho,\tau^h) = \mathrm{tr}(\hat{H}\rho) - \frac{1}{\beta_h} S(\rho), \tag{F22}$$

where $\langle \hat{H}\rangle_\rho := \mathrm{tr}[\hat{H}\rho]$, and $S(\rho) = -\mathrm{tr}(\rho \ln \rho)$ is the von Neumann entropy, while τ^h is the thermal state at inverse temperature β_h for the Hamiltonian $\hat{H}$. Also, let us recall that $W_{\text{ext}} = E_k^{\text{W}} - E_j^{\text{W}} > 0$ where E_j^{W} is a constant.

Next, we consider the free energies $F_1(\tau^0_{\text{Cold}} \otimes \rho^0_{\text{M}} \otimes \rho^0_{\text{W}}, \tau^h_{\text{ColdMW}})$ and $F_1(\rho^1_{\text{ColdMW}}, \tau^h_{\text{ColdMW}})$ respectively, and how they relate to W_{ext}. First of all, note that the quantity $F_1(\tau^0_{\text{Cold}} \otimes \rho^0_{\text{M}} \otimes \rho^0_{\text{W}}, \tau^h_{\text{ColdMW}})$ is simply a constant that does not depend on W_{ext}. This is because

$$\begin{aligned} F_1(\tau^0_{\text{Cold}} \otimes \rho^0_{\text{M}} \otimes \rho^0_{\text{W}}, \tau^h_{\text{ColdMW}}) &= F_1(\tau^0_{\text{Cold}}, \tau^h_{\text{Cold}}) + F_1(\tau^0_{\text{M}}, \tau^h_{\text{M}}) + F_1(\tau^0_{\text{W}}, \tau^h_{\text{W}}) && \text{(F23)}\\ &= F_1(\tau^0_{\text{Cold}}, \tau^h_{\text{Cold}}) + F_1(\tau^0_{\text{M}}, \tau^h_{\text{M}}) + \mathrm{tr}(\hat{H}_{\text{W}}\rho^0_{\text{W}}) - \beta_h^{-1} S(\rho^0_{\text{W}}) && \text{(F24)}\\ &= F_1(\tau^0_{\text{Cold}}, \tau^h_{\text{Cold}}) + F_1(\tau^0_{\text{M}}, \tau^h_{\text{M}}) + E_j^{\text{W}}, && \text{(F25)} \end{aligned}$$

where the first two terms do not depend on the battery Hamiltonian at all, while in the last equality we have made use of the fact that $\rho^0_{\text{W}} = |E_j\rangle\langle E_j|_{\text{W}}$. On the other hand,

$$\begin{aligned} F_1(\rho^1_{\text{ColdMW}}, \tau^h_{\text{ColdMW}}) &= \mathrm{tr}(\hat{H}_{\text{ColdMW}}\rho^1_{\text{ColdMW}}) - \beta_h^{-1} S(\rho^1_{\text{ColdMW}}) && \text{(F26)}\\ &= \mathrm{tr}(\hat{H}_{\text{Cold}}\rho^1_{\text{Cold}}) + \mathrm{tr}(\hat{H}_{\text{M}}\rho^1_{\text{M}}) + \mathrm{tr}(\hat{H}_{\text{W}}\rho^1_{\text{W}}) - \beta_h^{-1} S(\rho^1_{\text{ColdMW}}) && \text{(F27)}\\ &= \mathrm{tr}(\hat{H}_{\text{Cold}}\rho^1_{\text{Cold}}) + \mathrm{tr}(\hat{H}_{\text{M}}\rho^1_{\text{M}}) - \beta_h^{-1} S(\rho^1_{\text{ColdMW}}) + \varepsilon E_j^{\text{W}} + (1-\varepsilon) E_k^{\text{W}}. && \text{(F28)} \end{aligned}$$

Note that again, $\mathrm{tr}(\hat{H}_{\text{Cold}}\rho^1_{\text{Cold}})$ and $\mathrm{tr}(\hat{H}_{\text{M}}\rho^1_{\text{M}})$ do not depend on the battery Hamiltonian and therefore do not depend on E_k^{W}. Similarly, $S(\rho^1_{\text{ColdMW}})$ depends only on the eigenvalues of the state, and is independent of E_k^{W}. Since $\varepsilon \in [0,1)$, we may conclude the following: *$F(\rho^1_{\text{ColdMW}}, \tau^h_{\text{ColdMW}})$ is a continuous function that strictly increases w.r.t. E_k^{W}, and therefore it also strictly increases w.r.t. W_{ext}.*

To prove this lemma, it suffices to show that the supremum over W_{ext} in Eq. (F10) for $\alpha = 1$ has to be achieved when $F_1(\tau^0_{\text{Cold}} \otimes \rho^0_{\text{M}} \otimes \rho^0_{\text{W}}, \tau^h_{\text{ColdMW}}) = F_1(\rho^1_{\text{ColdMW}}, \tau^h_{\text{ColdMW}})$. We prove this by contradiction. Suppose that $\hat{W}_{\text{ext}}$ achieves the supremum for η_1^{qm}, and for this value of $\hat{W}_{\text{ext}}$, $F_1(\tau^0_{\text{Cold}} \otimes \rho^0_{\text{M}} \otimes \rho^0_{\text{W}}, \tau^h_{\text{ColdMW}}) > F_1(\rho^1_{\text{ColdMW}}, \tau^h_{\text{ColdMW}})$. Since we know that $F(\rho^1_{\text{ColdMW}}, \tau^h_{\text{ColdMW}})$ strictly increases w.r.t. W_{ext}, there must exist an $W'_{\text{ext}} > \hat{W}_{\text{ext}}$ such that $F_1(\tau^0_{\text{Cold}} \otimes \rho^0_{\text{M}} \otimes \rho^0_{\text{W}}, \tau^h_{\text{ColdMW}}) \geq F_1(\rho^1_{\text{ColdMW}}, \tau^h_{\text{ColdMW}})$. Furthermore, since by Eq. (F9) we know that the efficiency is monotonically increasing w.r.t. W_{ext} as well, it follows that W'_{ext} achieves a higher value of efficiency compared to $\hat{W}_{\text{ext}}$ while satisfying the required constraints at the same time. This is a contradiction, and therefore we conclude that the optimization for η_1^{qm} can be simplified to $a(\rho^1_{\text{ColdMW}})$, where the constraint on F_1 holds with equality. $\square$

Lemma 19. *For any final state $\rho^1_{\text{ColdHotMW}}$, and any Hamiltonian of the form in Eq. (A1), then for perfect or near perfect work extraction (see Defs. 1 and 2), we have*

$$\eta^{\text{qm}}\left(\rho^1_{\text{ColdMW}}\right) \overset{(1)}{\leq} \eta_1^{\text{qm}}\left(\rho^1_{\text{ColdMW}}\right) \overset{(2)}{\leq} \eta_1^{\text{qm}}\left(\underline{\rho^1_{\text{ColdMW}}}\right) \overset{(3)}{=} \eta^{\text{mac}}\left(\underline{\rho^1_{\text{Cold}}}\right) \overset{(4)}{\leq} 1 - \frac{\beta_h}{\beta_c}, \tag{F29}$$

with equality in (2) *iff $\rho^1_{\text{ColdMW}} = \underline{\rho^1_{\text{ColdMW}}}$. The quantities η_1^{qm} and η^{mac} are defined in Eq.* (F10) *and Eq.* (C5) *respectively.*

Proof. Note that inequality (1) is a direct consequence of Lemma 17, while inequality (4) holds because of Lemma 6. It remains to prove inequalities (2) and (3).
Proof of inequality (2): Using the definition in Eq. (F10) together with Lemma 18, let us compare the quantities

$$\begin{aligned} \eta_1^{\text{qm}}(\rho^1_{\text{ColdMW}}) = \sup_{W_{\text{ext}}} \eta(\rho^1_{\text{Cold}}, W_{\text{ext}}) \quad \text{s.t.} \quad & F_1(\tau^0_{\text{Cold}} \otimes \rho^0_{\text{M}} \otimes \rho^0_{\text{W}}, \tau^h_{\text{ColdMW}}) = F_1(\rho^1_{\text{ColdMW}}, \tau^h_{\text{ColdMW}}), && \text{(F30)}\\ & \mathrm{tr}(\hat{H}_t \rho^0_{\text{ColdHotMW}}) = \mathrm{tr}(\hat{H}_t \rho^1_{\text{ColdHotMW}}), && \text{(F31)}\\ & \rho^1_{\text{W}} = \varepsilon |E_j\rangle\langle E_j|_{\text{W}} + (1-\varepsilon)|E_k\rangle\langle E_k|_{\text{W}}, && \text{(F32)}\\ & \rho^1_{\text{M}} = \rho^0_{\text{M}}, && \text{(F33)} \end{aligned}$$

and

$$\eta_1^{\text{qm}}(\underline{\rho_{\text{ColdMW}}^1}) = \sup_{W_{\text{ext}}} \eta(\underline{\rho_{\text{Cold}}^1}, W_{\text{ext}}) \quad \text{s.t. } F_1(\tau_{\text{Cold}}^0 \otimes \rho_{\text{M}}^0 \otimes \rho_{\text{W}}^0, \tau_{\text{ColdMW}}^h) = F_1(\underline{\rho_{\text{ColdMW}}^1}, \tau_{\text{ColdMW}}^h), \tag{F34}$$

$$\text{tr}(\hat{H}_t \rho_{\text{ColdHotMW}}^0) = \text{tr}(\hat{H}_t \rho_{\text{ColdHotMW}}^1), \tag{F35}$$

$$\rho_{\text{W}}^1 = \varepsilon |E_j\rangle\langle E_j|_{\text{W}} + (1-\varepsilon)|E_k\rangle\langle E_k|_{\text{W}}, \tag{F36}$$

$$\rho_{\text{M}}^1 = \rho_{\text{M}}^0. \tag{F37}$$

We first make the following observations:

- By our definition of $\underline{\rho_{\text{ColdMW}}^1}$, we have that $\rho_{\text{Cold}}^1 = \underline{\rho_{\text{Cold}}^1}$. Therefore, the term ΔC in Eq. (F9) which is only a function of the reduced state on the cold bath is the same for both efficiencies in Eq. (F30) and Eq. (F34). Therefore, to compare the efficiencies, we need only to compare the value of W_{ext} that satisfies the free energy constraint in both optimization problems.

- In [30] (pg. 395) it has been proven that the von Neumann entropy is subadditive

$$S(\rho_{AB}) \leq S\left(\underline{\rho_{AB}}\right), \tag{F38}$$

with equality iff $\rho_{AB} = \underline{\rho_{AB}}$. Furthermore, since $\hat{H}_{\text{ColdMW}}$ does not contain interaction terms, as we have demonstrated earlier in Eq. (F16),

$$\text{tr}(\hat{H}_{\text{ColdMW}} \rho_{\text{ColdMW}}^1) = \text{tr}(\hat{H}_{\text{ColdMW}} \underline{\rho_{\text{ColdMW}}^1}). \tag{F39}$$

Thus, by Eq. (F22) we conclude that

$$F_1(\underline{\rho_{\text{ColdMW}}^1}) \leq F_1(\rho_{\text{ColdMW}}^1), \tag{F40}$$

with equality iff $\rho_{\text{ColdMW}}^1 = \underline{\rho_{\text{ColdMW}}^1}$.

- For any final state ρ_{ColdMW}^1 where $\rho_{\text{W}}^1 = \varepsilon |E_j\rangle\langle E_j|_{\text{W}} + (1-\varepsilon)|E_k\rangle\langle E_k|_{\text{W}}$, we have seen in the proof of Lemma 18 that $F_1(\rho_{\text{ColdMW}}^1, \tau_{\text{ColdMW}}^h)$ is a continuous function that strictly increases with W_{ext}.

With these three observations we can now prove inequality (2). Note that when $\rho_{\text{ColdMW}}^1 = \underline{\rho_{\text{ColdMW}}^1}$, equality holds trivially. Therefore, let us consider the case where $\rho_{\text{ColdMW}}^1 \neq \underline{\rho_{\text{ColdMW}}^1}$. Suppose $\hat{W}_{\text{ext}}$ achieves the supremum in $\eta_1^{\text{qm}}(\rho_{\text{ColdMW}}^1)$, and for such a value of $\hat{W}_{\text{ext}}$,

$$F_1(\tau_{\text{Cold}}^0 \otimes \rho_{\text{M}}^0 \otimes \rho_{\text{W}}^0, \tau_{\text{ColdMW}}^h) = F_1(\rho_{\text{ColdMW}}^1, \tau_{\text{ColdMW}}^h) > F_1(\underline{\rho_{\text{ColdMW}}^1}, \tau_{\text{ColdMW}}^h). \tag{F41}$$

We note also that since $F_1(\underline{\rho_{\text{ColdMW}}^1}, \tau_{\text{ColdMW}}^h)$ strictly increases with W_{ext}, and therefore there exists some $W'_{\text{ext}} > \hat{W}_{\text{ext}}$ such that $F_1(\tau_{\text{Cold}}^0 \otimes \rho_{\text{M}}^0 \otimes \rho_{\text{W}}^0, \tau_{\text{ColdMW}}^h) = F_1(\underline{\rho_{\text{ColdMW}}^1}, \tau_{\text{ColdMW}}^h)$. Therefore, W'_{ext} is a feasible solution for Eq. (F34), i.e. it satisfies the constraints in the optimization problem. In conclusion, we have

$$\eta_1^{\text{qm}}(\rho_{\text{ColdMW}}^1) = \left[1 - \varepsilon + \frac{\Delta C}{\hat{W}_{\text{ext}}}\right]^{-1} \leq \left[1 - \varepsilon + \frac{\Delta C}{W'_{\text{ext}}}\right]^{-1} \leq \eta_1^{\text{qm}}(\underline{\rho_{\text{ColdMW}}^1}). \tag{F42}$$

Proof of equality (3): Consider the quantity $\eta_1^{\text{qm}}(\underline{\rho_{\text{ColdMW}}^1})$. Since the state $\underline{\rho_{\text{ColdMW}}^1}$ takes on a product structure form between all the subsystems now, Assumption (iii) in the beginning of Section F 1 a holds again. By Eqns. (F32) and (F33), we know that Assumptions (i) and (ii) also hold. Therefore, we know that under these assumptions the efficiency does not depend anymore on the global state $\underline{\rho_{\text{ColdMW}}^1}$, but only ρ_{Cold}^1. Again comparing the conditions of $\eta^{\text{mac}}(\underline{\rho_{\text{Cold}}^1})$ and $\eta_1^{\text{qm}}(\underline{\rho_{\text{ColdMW}}^1})$, we see that they are exactly the same quantity. □

Therefore, Lemma 19 tells us that correlations between the final states of the cold bath, machine and battery cannot allow you to achieve an efficiency greater than the Carnot efficiency.

c. Achievability of Carnot efficiency still depends on more than temperature

Earlier in Section F 1 b, we proved in Lemma 19 that Carnot efficiency gives an upper bound to the efficiency of any arbitrary final state ρ^1_{ColdMW}. In this section, we want to prove that when $\Omega > 1$ holds, one cannot achieve the Carnot efficiency even when allowing correlations between the final states of the battery and the cold bath. This can be done in the following steps:

- According to Lemma 19, Carnot efficiency can be attained only when all the inequalities in Eq. (F29) are satisfied with equalities. We use this to prove in Lemma 20 that in order to achieve the Carnot efficiency, we may only consider the limit where correlations in the final state vanish. Not only so, the magnitude of these correlations also have to vanish quickly enough in order for Carnot efficiency to be achieved. In particular, we define a parameter k which quantifies the amount of correlations, and show that k has to vanish faster than the quasi-static parameter g, in order to achieve the Carnot efficiency η_C.
- Next, in Lemma 21, we show that if the parameter k vanishes faster than the quasi-static parameter g, then whenever $\Omega > 1$, one can derive an upper bound for the intermediate efficiency $\eta^{\text{qm}}_\infty(\rho^1_{\text{ColdMW}})$ which considers the amount of work extractable by invoking only the generalized second law of $\alpha \to \infty$. Combining Lemma 20 and Lemma 21, we conclude in Corollary 3 that when $\Omega > 1$, $\eta^{\text{qm}} \leq \eta^{\text{qm}}_\infty \leq \eta_C$ is strictly upper bounded away from the Carnot efficiency.

Before we begin, let us note that by definition, the initial state ρ^0_{ColdW} is diagonal in its energy eigenbasis. Furthermore, the state ρ^0_{ColdMW} is of the form $\rho^0_{\text{Cold}} \otimes \rho^0_{\text{M}} \otimes \rho^0_{\text{W}}$. Since w.l.o.g. we can assume that $\hat{H}_{\text{M}}$ is proportional to the identity (or called the *trivial Hamiltonian*, see [5]), ρ^0_{M} can always be written as a diagonal state in an energy eigenbasis of its Hamiltonian. Therefore the state ρ^0_{ColdMW} is always diagonal in the energy eigenbasis of the Hamiltonian $\hat{H}_{\text{ColdMW}} := \hat{H}_{\text{Cold}} + \hat{H}_{\text{M}} + \hat{H}_{\text{W}}$. Since catalytic thermal operations cannot create coherences [5], ρ^1_{ColdMW} has to be also diagonal in the energy eigenbasis of $\hat{H}_{\text{ColdMW}}$.

We observe that any ρ^1_{ColdMW} can always be written as

$$\rho^1_{\text{ColdMW}} = (1-k^*)\underline{\rho^1_{\text{ColdMW}}} + k^* \rho^{\text{corr}}_{\text{ColdMW}}, \tag{F43}$$

where $k^* = \min\{k \in [0,1] | \rho^1_{\text{ColdMW}} = (1-k)\underline{\rho^1_{\text{ColdMW}}} + kQ, Q \geq \mathbf{0}\}$. This means that ρ^1_{ColdMW} can be written as a convex combination of two states: one being $\underline{\rho^1_{\text{ColdMW}}}$, and the other $\rho^{\text{corr}}_{\text{ColdMW}}$ containing all other correlations. Note that such a k^* always exists, in particular, $k = 1$ is always a feasible solution.

We now define a particular parametrization of the final states,

$$\rho^1_{\text{ColdMW}}(k, \rho^{\text{no corr}}_{\text{ColdMW}}, \rho^{\text{corr}}_{\text{ColdMW}}) := (1-k)\rho^{\text{no corr}}_{\text{ColdMW}} + k\rho^{\text{corr}}_{\text{ColdMW}}, \quad k \in [0, k^*] \tag{F44}$$

where the following holds:

$$\text{(i)} \quad \rho^{\text{no corr}}_{\text{ColdMW}} = \underline{\rho^1_{\text{ColdMW}}}, \tag{F45}$$

$$\text{(ii)} \quad \rho^{\text{corr}}_{\text{ColdMW}} \neq \rho^{\text{no corr}}_{\text{ColdMW}}, \tag{F46}$$

$$\text{(iii)} \quad \rho^1_{\text{M}} = (1-k)\rho^{\text{no corr}}_{\text{M}} + k\rho^{\text{corr}}_{\text{M}} = \rho^0_{\text{M}}. \tag{F47}$$

Since in our heat engine, the initial state has no coherences, it suffices to consider ρ^1_{ColdMW} which is diagonal in the energy eigenbasis. This implies that $\rho^{\text{no corr}}_{\text{ColdMW}} = \underline{\rho^1_{\text{ColdMW}}}$ is also diagonal in the energy eigenbasis, and therefore the same holds for $\rho^{\text{corr}}_{\text{ColdMW}}$ due to Eq. (F44). All correlations between the individual systems of cold bath, machine and battery are contained only in $\rho^{\text{corr}}_{\text{ColdMW}}$. Therefore, $\rho^1_{\text{ColdMW}}(\cdot,\cdot,\cdot)$ parametrizes every possible quantum state on $\mathcal{H}_{\text{ColdMW}}$ which is diagonal in the global energy eigenbasis and that returns the machine locally to its initial state after one cycle of the heat engine. In Eq. (F47), ρ^1_{M} is the final state of the machine, since the heat engine is cyclic, recall from Section A that we require $\rho^1_{\text{M}} = \rho^0_{\text{M}}$.

Lemma 20. *For every family of states $\rho^1_{\text{ColdMW}}(k, \rho^{\text{no corr}}_{\text{ColdMW}}, \rho^{\text{corr}}_{\text{ColdMW}})$ parametrized by k, (see Eqs. (F44)-(F47)), if the quantum efficiency η^{qm}_1 defined in Eq. (F10) achieves the Carnot efficiency*

$$\eta^{\text{qm}}_1(\rho^1_{\text{ColdMW}}) = 1 - \frac{\beta_h}{\beta_c}, \tag{F48}$$

then the following conditions are satisfied:

1) The state $\underline{\rho^1_{\text{ColdMW}}}$ is the final state of a quasi-static heat engine (see Def. 3)

$$\underline{\rho^1_{\text{ColdMW}}} = \tau(g) \otimes \rho^0_{\text{M}}(g) \otimes \rho^1_{\text{W}} \quad \textit{with } g \to 0^+. \tag{F49}$$

2) The correlations must vanish sufficiently quickly. That is to say, the parameter k in Eq. (F44) vanishes more quickly compared to g, i.e.

$$\lim_{g\to 0^+}\frac{k}{g}=0. \tag{F50}$$

Proof. Firstly, suppose that Carnot efficiency is achieved, i.e. $\eta^{\rm qm}(\rho^1_{\rm ColdMW})=1-\frac{\beta_h}{\beta_c}$. Then according to Lemma 19, all inequalities in Eq. (F29) should be satisfied with equality, in particular inequality (4). We have established in Lemma 6 that this equality is only achieved in the quasi-static limit, i.e. $\rho^1_{\rm Cold}=\tau_{\rm Cold}(g)$ where $g\to 0^+$. This implies Condition 1) in the statement of the lemma.

The proof for Condition 2) consists of calculating $W_{\rm ext}$ for $\alpha=1$ in Eq. (F10) to leading order in g and k. This $W_{\rm ext}$ quantity can be later used to evaluate $\eta_1^{\rm qm}$. We will show that we can write the expression for $\eta_1^{\rm qm}$ into two terms: one term describes the efficiency when there are no final correlations, and the other term is a strictly negative contribution which must vanish in order to achieve the Carnot efficiency. This latter constraint will give us Eq. (F50).

Let us denote $W'_{\rm ext}$ as the value of battery energy gap $W_{\rm ext}=E^{\rm W}_k-E^{\rm W}_j$ that solves the equation

$$F_1(\tau^0_{\rm Cold}\otimes\rho^0_{\rm M}(g)\otimes\rho^0_{\rm W},\tau^h_{\rm ColdMW})=F_1(\rho^1_{\rm ColdMW}(k,\rho^{\rm no\ corr}_{\rm ColdMW},\rho^{\rm corr}_{\rm ColdMW}),\tau^h_{\rm ColdMW})^{6}, \tag{F51}$$

while $\hat{W}_{\rm ext}$ as the value that solves the case where $k=0$, i.e.

$$F_1(\tau^0_{\rm Cold}\otimes\rho^0_{\rm M}(g)\otimes\rho^0_{\rm W},\tau^h_{\rm ColdMW})=F_1(\rho^{\rm no\ corr}_{\rm ColdMW},\tau^h_{\rm ColdMW}). \tag{F52}$$

Since $\rho^{\rm no\ corr}_{\rm ColdMW}=\underline{\rho^1_{\rm ColdMW}}=\rho^1_{\rm Cold}\otimes\rho^0_{\rm M}(g)\otimes\rho^1_{\rm W}$ contains no correlations, $\hat{W}_{\rm ext}$ was given by Eq. (D51). According to Lemma 18, we know that $W'_{\rm ext}$ and $\hat{W}_{\rm ext}$ are the values of $W_{\rm ext}$ which solve $\sup\limits_{W_{\rm ext}}\eta_1^{\rm qm}(\rho^1_{\rm ColdMW},W_{\rm ext})$ and $\sup\limits_{W_{\rm ext}}\eta_1^{\rm qm}(\underline{\rho^1_{\rm ColdMW}},W_{\rm ext})$ respectively. Solving Eq. (F52) for $W'_{\rm ext}$ with the aid of Eq. (F22), we find

$$W'_{\rm ext}=\hat{W}_{\rm ext}-\chi, \tag{F53}$$

where $W_{\rm ext}$ is the solution to Eq. (F52) when $k=0$, given by Eq. (D51), while

$$\chi:=\frac{1}{\beta_h}\frac{1}{1-\varepsilon}\left[S(\rho^{\rm no\ corr}_{\rm ColdMW})-S\left(\rho^1_{\rm ColdMW}(k,\rho^{\rm no\ corr}_{\rm ColdMW},\rho^{\rm corr}_{\rm ColdMW})\right)\right]. \tag{F54}$$

Let us first note some properties of χ, which we will later use:

- Since $S(\cdot)$ is subadditive, due to the parametrization of $\rho^1_{\rm ColdMW}(\cdot,\cdot,\cdot)$ in Eq. (F44), we have

$$\chi\geq 0 \tag{F55}$$

  with equality iff $\underline{\rho^1_{\rm ColdMW}}=\rho^1_{\rm ColdMW}$ i.e. iff $k=0$. Therefore, we may conclude that $\frac{\hat{W}_{\rm ext}}{W'_{\rm ext}}\geq 1$.

- We have that

$$\frac{d}{dk}\chi(k,\rho^{\rm no\ corr}_{\rm ColdMW},\rho^{\rm corr}_{\rm ColdMW})\Big|_{k=k_0}=0 \tag{F56}$$

  if and only if

$$\rho^1_{\rm ColdMW}(k_0,\rho^{\rm no\ corr}_{\rm ColdMW},\rho^{\rm corr}_{\rm ColdMW})=\mathbb{1}_{\rm ColdMW}/N. \tag{F57}$$

  Eqs. (F56) and (F57) are direct consequences of the observations:
  1) Entropy is strictly concave, i.e. $S\left(\rho^1_{\rm ColdMW}(k,\rho^{\rm no\ corr}_{\rm ColdMW},\rho^{\rm corr}_{\rm ColdMW})\right)$ is strictly concave in $k\in[0,1]$. Therefore, by Eq. (F57) χ is strictly convex in $k\in[0,1]$. When the first derivative of the convex function $\frac{d\chi}{dk}=0$, this must be the global minimum [**?**].
  2) However, we know that the entropy is uniquely maximized (and therefore χ is uniquely minimized) for the maximally mixed state.

[6] We denote $\rho^0_{\rm M}(g)$ because for different values of g, we are allowed to choose different initial machine states, as long as $\rho^1_{\rm M}(g)=\rho^0_{\rm M}(g)$.

Returning to evaluate the efficiency, we may use Eq. (F9) to calculate the inverse efficiency,

$$[\eta_1^{\text{qm}}(\rho^1_{\text{ColdMW}})]^{-1} = 1-\varepsilon+\frac{\Delta C(\rho^1_{\text{Cold}})}{W'_{\text{ext}}} \tag{F58}$$

$$= 1-\varepsilon+\frac{\Delta C(\rho^1_{\text{Cold}})}{\hat{W}_{\text{ext}}}\frac{\hat{W}_{\text{ext}}}{W'_{\text{ext}}} \tag{F59}$$

$$\geq 1-\varepsilon+\frac{\Delta C(\rho^1_{\text{Cold}})}{\hat{W}_{\text{ext}}}. \tag{F60}$$

The last term in Eq. (F59) is non-negative because we know the terms $\Delta C(\rho^1_{\text{Cold}})$, $\hat{W}_{\text{ext}}$ and W'_{ext} are all non-negative.

With Condition 1), we now know that

$$1-\varepsilon+\frac{\Delta C(\rho^1_{\text{Cold}})}{W_{\text{ext}}} = 1-\frac{\beta_h}{\beta_c}, \tag{F61}$$

in the quasi-static limit, and therefore a necessary condition to achieve the Carnot efficiency is that $\lim_{g\to 0}\frac{\hat{W}_{\text{ext}}}{W'_{\text{ext}}}=1$ also in the quasi-static limit. Using the relation $W'_{\text{ext}}=\hat{W}_{\text{ext}}+\chi$, we have the requirement that

$$\lim_{g\to 0^+}\frac{\chi(k,\rho^{\text{no corr}}_{\text{ColdMW}}(g),\rho^{\text{corr}}_{\text{ColdMW}})}{\hat{W}_{\text{ext}}(\rho^{\text{no corr}}_{\text{ColdMW}}(g))}=0. \tag{F62}$$

First, let us observe that $\hat{W}_{\text{ext}}(\rho^{\text{no corr}}_{\text{ColdMW}}(g)) = W_{\text{ext}}(\beta_c-g)$ given by Eq. (D37). The leading order term of $W_{\text{ext}}(\beta_c-g)=\Theta(g)$ as $g\to 0^+$. Therefore, in order to satisfy Eq. (F62), we must firstly have $\lim_{g\to 0}\chi=0$. From Eqs. (F44), (F54), this implies that we need $k\to 0$ for all $\rho^{\text{no corr}}_{\text{ColdMW}}$.

Since the numerator and denominator of Eq. (F62) both go to zero, by L'Hospital rule, to evaluate the limit we need to take the derivative of both terms w.r.t. g. Therefore, we expand χ to first order in k and g. From Eq. (F54) it follows

$$\chi(k,\rho^{\text{no corr}}_{\text{ColdMW}}(g),\rho^{\text{corr}}_{\text{ColdMW}}) = \frac{d}{dk}\chi(k,\rho^{\text{no corr}}_{\text{ColdMW}}(0),\rho^{\text{corr}}_{\text{ColdMW}})\Big|_{k=0} k + \frac{d}{dg}\chi(0,\rho^{\text{no corr}}_{\text{ColdMW}}(g),\rho^{\text{corr}}_{\text{ColdMW}})\Big|_{g=0} g + o(gk)+o(k^2)+o(g^2) \tag{F63}$$

$$= \frac{d}{dk}\chi(k,\rho^{\text{no corr}}_{\text{ColdMW}}(0),\rho^{\text{corr}}_{\text{ColdMW}})\Big|_{k=0} k + o(gk)+o(k^2)+o(g^2). \tag{F64}$$

The term $\frac{d}{dg}\chi(0,\rho^{\text{no corr}}_{\text{ColdMW}}(g),\rho^{\text{corr}}_{\text{ColdMW}})\Big|_{g=0}=0$ since when $k=0$, χ will be constant for all g. Next, we note that since Eqs. (F55) holds, it must be that $\frac{d}{dk}\chi(k,\rho^{\text{no corr}}_{\text{ColdMW}}(0),\rho^{\text{corr}}_{\text{ColdMW}})\Big|_{k=0}\geq 0$. Furthermore, from Eq. (F56), we have that

$$\frac{d}{dk}\chi(k,\rho^{\text{no corr}}_{\text{ColdMW}}(0),\rho^{\text{corr}}_{\text{ColdMW}})\Big|_{k=0}\neq 0, \tag{F65}$$

for all $\rho^{\text{corr}}_{\text{ColdMW}}$ since by definition $\rho^1_{\text{ColdMW}}(0,\rho^{\text{no corr}}_{\text{ColdMW}}(0),\rho^{\text{corr}}_{\text{ColdMW}})\neq \mathbb{1}_{\text{ColdMW}}/N$. We can infer that ρ^1_{ColdMW} is not maximally mixed from a few observations, for example: this is true because we have required that the reduced state on the battery is not maximally mixed since we consider near perfect work extraction.

Thus, taking into account $W_{\text{ext}}(\beta_c-g)=\Theta(g)$, Eq. (F62) implies Eq. (F50). □

By now, we have established a constraint on how quickly the correlations have to vanish w.r.t. g, for the possibility of achieving Carnot efficiency. In the next Lemma 21, we will show that the constraints given by Eq. (F50) can be used to derive an upper bound for η^{qm}_∞.

Lemma 21. *If Eqs.* (F49) *and* (F50) *are satisfied, then the quantity* $\eta_\infty^{\mathrm{qm}}$ *can be upper bounded by*

$$\eta_\infty^{\mathrm{qm}}(\rho^1_{\mathrm{ColdMW}}(k,\rho^{\text{no corr}}_{\mathrm{ColdMW}}(g),\rho^{\mathrm{corr}}_{\mathrm{ColdMW}})) \tag{F66}$$
$$\leq \left[1+\frac{\beta_h}{\beta_c-\beta_h}\frac{\gamma(1)}{\gamma(\infty)}\right]^{-1} + \Theta(f(g)) + \Theta(k/g) + \Theta(g) + \Theta(\varepsilon), \tag{F67}$$

with $\lim_{g\to 0^+} f(g) = 0$.

Proof. The main idea of our proof is as follows: we show that if Eqns. (F49) and (F50) hold, then we can upper bound W_{ext} while considering only the F_∞ condition. This bound differs from the value given when no correlations are present by only a small amount. Substituting this into the expression for $\eta_\infty^{\mathrm{qm}}$, we obtain Eq. (F67).

Let us begin by analyzing the difference in eigenvalues of the states ρ^1_{ColdMW} and $\underline{\rho^1_{\mathrm{ColdMW}}}$. Recall that

$$\rho^1_{\mathrm{ColdMW}}(k,\rho^{\text{no corr}}_{\mathrm{ColdMW}},\rho^{\mathrm{corr}}_{\mathrm{ColdMW}}) = (1-k)\rho^{\text{no corr}}_{\mathrm{ColdMW}} + k\rho^{\mathrm{corr}}_{\mathrm{ColdMW}} \tag{F68}$$

where $\rho^{\text{no corr}}_{\mathrm{ColdMW}}, \rho^{\mathrm{corr}}_{\mathrm{ColdMW}}$ are both diagonal in the energy eigenbasis. Since ρ^1_{ColdMW} is a mixture of two energy-diagonal states, it is also diagonal. Let us denote its eigenvalues as $[\rho^1_{\mathrm{ColdMW}}]_i$.

As for $\underline{\rho^1_{\mathrm{ColdMW}}}$, Eqn. (F49) gives the explicit form of the state,

$$\underline{\rho^1_{\mathrm{ColdMW}}} = \rho^1_{\mathrm{Cold}}\otimes\rho^1_{\mathrm{M}}\otimes\rho^1_{\mathrm{W}} = \tau(g)\otimes\rho^0_{\mathrm{M}}(g)\otimes\rho^1_{\mathrm{W}}. \tag{F69}$$

Let us denote its eigenvalues as $[\underline{\rho^1_{\mathrm{ColdMW}}}]_i$.

We first observe two properties involving trace distance $d(\cdot,\cdot)$:

(P.i) Consider two states σ_1,σ_2 diagonal in the same eigenbasis. Then if $\rho=(1-k)\sigma_1+k\sigma_2$ for some $k\in[0,1]$, then one can conclude that the distance

$$d(\rho,\sigma_1)\leq k. \tag{F70}$$

(P.ii) For any two states ρ,σ diagonal in the same basis, with eigenvalues p_i, q_i, if their trace distance

$$d(\rho,\sigma)=\frac{1}{2}\|\rho-\sigma\|_1\leq\varepsilon, \tag{F71}$$

then this implies that their eigenvalues cannot differ by more than ε, i.e. $\forall i, |p_i-q_i|\leq\varepsilon$. By using this fact, we may first calculate the trace distance between ρ^1_{ColdMW} and $\underline{\rho^1_{\mathrm{ColdMW}}}$, then bound the difference of their eigenvalues.

We find that

$$d(\rho^1_{\mathrm{ColdMW}},\underline{\rho^1_{\mathrm{ColdMW}}}) \leq d(\rho^1_{\mathrm{ColdMW}},\rho^{\text{no corr}}_{\mathrm{ColdMW}}) + d(\rho^{\text{no corr}}_{\mathrm{ColdMW}},\underline{\rho^1_{\mathrm{ColdMW}}}) \tag{F72}$$
$$\leq k + d(\rho^{\text{no corr}}_{\mathrm{Cold}},\rho^1_{\mathrm{Cold}}) + d(\rho^{\text{no corr}}_{\mathrm{M}},\rho^1_{\mathrm{M}}) + d(\rho^{\text{no corr}}_{\mathrm{W}},\rho^1_{\mathrm{W}}) \tag{F73}$$
$$\leq 4k. \tag{F74}$$

The first inequality is a triangle inequality that holds for all states. The second inequality holds because of (P.i), and because trace distance is subadditive under tensor product (note that both $\rho^{\text{no corr}}_{\mathrm{ColdMW}}$ and $\underline{\rho^1_{\mathrm{ColdMW}}}$ are tensor product states). The third inequality holds because we know $d(\rho^1_{\mathrm{ColdMW}},\rho^{\text{no corr}}_{\mathrm{ColdMW}})\leq k$ and that trace distance decreases under partial trace. By (P.ii), Eq. (F74) tells us that $\forall i$,

$$[\rho^1_{\mathrm{ColdMW}}]_i = [\underline{\rho^1_{\mathrm{ColdMW}}}]_i + o(k). \tag{F75}$$

With Eq.(F75), we may relate the F_∞ quantities for the states ρ^1_{ColdMW} and $\underline{\rho^1_{\text{ColdMW}}}$. From Eq. (B8), we have

$$F_\infty\left(\rho^1_{\text{ColdMW}}(k, \rho^{\text{no corr}}_{\text{ColdMW}}(g), \rho^{\text{corr}}_{\text{ColdMW}}), \tau^h_{\text{ColdMW}}\right) \tag{F76}$$

$$= \ln \max_i \left\{ \frac{[\rho^1_{\text{ColdMW}}]_i}{\tau_i} \right\}, \tag{F77}$$

$$= \ln \max_i \left\{ \frac{[\underline{\rho^1_{\text{ColdMW}}}]_i}{\tau_i} \right\} + o(k), \tag{F78}$$

$$= F_\infty\left(\tau(g) \otimes \rho^0_{\text{M}}(g) \otimes \rho^1_{\text{W}}, \tau^h_{\text{ColdMW}}\right) + o(k), \tag{F79}$$

where we used Eq. (B8) in the last line.
The next step is to evaluate the restriction on W_{ext} that satisfies

$$F_\infty(\tau^0_{\text{Cold}} \otimes \rho^0_{\text{M}} \otimes \rho^0_{\text{W}}, \tau^h_{\text{ColdMW}}) \geq F_\infty\left(\rho^1_{\text{ColdMW}}(k, \rho^{\text{no corr}}_{\text{ColdMW}}(g), \rho^{\text{corr}}_{\text{ColdMW}}), \tau^h_{\text{ColdMW}}\right) \tag{F80}$$

$$= F_\infty\left(\tau(g) \otimes \rho^0_{\text{M}}(g) \otimes \rho^1_{\text{W}}, \tau^h_{\text{ColdMW}}\right) + o(k), \tag{F81}$$

for W_{ext} up to order $o(k)$. Taking into account the additivity of F_∞ under tensor product, we can rearrange Eq. (F81) to provide an upper bound on W_{ext},

$$W_{\text{ext}} \leq \frac{ng}{\beta_h}\left[\gamma(\infty) + \Theta(f(g)) + o(k/g)\right], \tag{F82}$$

where $\lim_{g\to 0^+} f(g) = 0$, $\gamma(\infty)$ is given by Eq. (E113). The bound in (F82) is achievable since the F_∞ conditions imposed by Eq. (F80) are achievable with equality.

Lastly, by using the expression for efficiency in Eqs. (F9), and substituting W_{ext} from Eq. (F82) (with equality for the maximum possible W_{ext}) followed by ΔC from Eq. (E116), we have

$$\sup_{W_{\text{ext}}>0} \eta(\rho^1_{\text{Cold}}, W_{\text{ext}}) = \sup_{W_{\text{ext}}>0} \left(1 - \varepsilon + \frac{\Delta C}{W_{\text{ext}}}\right)^{-1} \tag{F83}$$

$$= \left[1 + \frac{\beta_h}{(\beta_c - \beta_h)}\frac{\gamma(1)}{\gamma(\infty)}\right]^{-1} + \Theta(f(g)) + o(k/g) + \Theta(g) + \Theta(\varepsilon). \tag{F84}$$

Hence using Eqs. (F10), (F84) we find Eq. (F67). Note that in Eq (F67) we have an inequality, this is due to the fact that in the optimisation problem Eq. (F10), there is an additional constraint (namely mean energy conservation) which is not taken into account in the derivation of Eq. (F84). □

Finally, the above lemmas allow us to conclude that allowing further correlations in the final state cannot allow us to achieve the Carnot efficiency when $\Omega > 1$.

Theorem 3. *[Correlations do not improve efficiency] Suppose that $\Omega > 1$. Parametrizing the final state of the heat engine by Eq. (F44)-(F47), the quantum efficiency η^{qm} defined in Eq. (F5) is strictly upper bounded by the Carnot efficiency,*

$$\sup_{k\in[0,1],\ \rho^{\text{no corr}}_{\text{ColdMW}}} \eta^{\text{qm}}\left(\rho^1_{\text{ColdMW}}(k, \rho^{\text{no corr}}_{\text{ColdMW}}, \rho^{\text{corr}}_{\text{ColdMW}})\right) < 1 - \frac{\beta_h}{\beta_c}. \tag{F85}$$

Proof. From Lemma 17, we have that both $\eta^{\text{qm}} \leq \eta^{\text{qm}}_1$ and $\eta^{\text{qm}} \leq \eta^{\text{qm}}_\infty$ hold. Thus a necessary condition to achieve the Carnot efficiency for a particular ρ^1_{ColdMW}, is that both η^{qm}_1 and η^{qm}_∞ are equal to or greater than the Carnot efficiency.

Lemma 20 proves that Eqs. (F49) and (F50) are necessary conditions for η^{qm}_1 to achieve the Carnot efficiency. However, when Eqs. (F49), (F50) are satisfied, then Lemma 21 provides an upper bound on the efficiency η^{qm}_∞ in Eq. (F67).

Now, suppose $\Omega > 1$. Since it is shown in Eq. (E114) that $\gamma(1)/\gamma(\infty) = \Omega$, plugging this into the leading term appearing in Eq. (F67)

$$\left[1 + \frac{\beta_h}{(\beta_c - \beta_h)}\frac{\gamma(1)}{\gamma(\infty)}\right]^{-1}, \tag{F86}$$

we have that the quantity η^{qm}_∞ (and therefore also η^{qm}) is strictly less than the Carnot efficiency $1 - \beta_h/\beta_c$. □

2. A more general final battery state

For the simplicity of our analysis, we have assumed that the battery is left in the specific final state described in Eq. (A5), i.e. an amount of work $W_{\rm ext} = E_k - E_j$ is extracted, except with failure probability ε that the battery remains in the initial state $|E_j\rangle\langle E_j|_{\rm W}$. In this section, we show that this is a simplification which can be removed in general, i.e. the final battery state is allowed to be any state within the ε-ball of $|E_k\rangle\langle E_k|_{\rm W}$. In particular, our result that the Carnot Efficiency cannot be achieved when $\Omega > 1$ still holds.

In Lemma 22, we show that for any final state of the cold bath $\rho^1_{\rm Cold}$, allowing a more general final battery state does not affect the amount of *work* bounded by the F_∞ condition. We then use this to prove in Theorem 4 that when $\Omega > 1$, Carnot cannot be achieved even if we allow a more general battery final state.

Lemma 22. *For any given $\rho^0_{\rm Cold}, \rho^1_{\rm Cold}$, with $\rho^0_{\rm W} = |E_j\rangle\langle E_j|_{\rm W}$, consider the maximum $W^1_\infty := E_{k_1} - E_j$ such that $\rho^0_{\rm Cold}\otimes\rho^0_{\rm W} \to \rho^1_{\rm Cold}\otimes\rho^1_{\rm W}$ is allowed by the non-increasing F_∞ condition (Eq. (B7)) i.e.*

$$D_\infty(\rho^0_{\rm Cold}\|\tau^{\beta_h}_{\rm Cold}) + D_\infty(\rho^0_{\rm W}\|\tau^{\beta_h}_{\rm W}) \geq D_\infty(\rho^1_{\rm Cold}\|\tau^{\beta_h}_{\rm Cold}) + D_\infty(\rho^1_{\rm W}\|\tau^{\beta_h}_{\rm W}), \tag{F87}$$

with

$$\rho^1_{\rm W} = (1-\varepsilon)\,|E_{k_1}\rangle\langle E_{k_1}|_{\rm W} + \varepsilon\,|E_j\rangle\langle E_j|_{\rm W}\,. \tag{F88}$$

On the other hand, consider any battery final state

$$\rho^2_{\rm W} = (1-\varepsilon)\,|E_{k_2}\rangle\langle E_{k_2}|_{\rm W} + \varepsilon\rho^{\rm junk}_{\rm W}, \tag{F89}$$

where $\rho^{\rm junk}_{\rm W}$ is an energy-diagonal state orthogonal to $|E_{k_2}\rangle\langle E_{k_2}|_{\rm W}$ which may depend on ε, i.e. $\rho^{\rm junk}_{\rm W} = \sum_i p_i\,|E_i\rangle\langle E_i|_{\rm W}$ with $p_{k_2} = 0$ and $\sum_i p_i = 1$. Define $W^2_\infty := E_{k_2} - E_j$ such that $\rho^0_{\rm Cold}\otimes\rho^0_{\rm W} \to \rho^1_{\rm Cold}\otimes\rho^2_{\rm W}$ is allowed by the non-increasing F_∞ condition, i.e.

$$D_\infty(\rho^0_{\rm Cold}\|\tau^{\beta_h}_{\rm Cold}) + D_\infty(\rho^0_{\rm W}\|\tau^{\beta_h}_{\rm W}) \geq D_\infty(\rho^1_{\rm Cold}\|\tau^{\beta_h}_{\rm Cold}) + D_\infty(\rho^2_{\rm W}\|\tau^{\beta_h}_{\rm W}). \tag{F90}$$

Then for all $0 < \varepsilon \leq \hat\varepsilon = \left[1 + e^{\beta_h(E_{\max}-E_j)}\right]^{-1}$, we have $W^1_\infty = W^2_\infty$.

Proof. Firstly, note that any energy-diagonal state $\rho^2_{\rm W}$ with trace distance $d(\rho^2_{\rm W}, |E_{k_2}\rangle\langle E_{k_2}|_{\rm W}) = \varepsilon$ can be written in the form of Eq. (F89). Rearranging the terms in Eq. (F87),

$$D_\infty(\rho^1_{\rm W}\|\tau^{\beta_h}_{\rm W}) \leq D_\infty(\rho^0_{\rm W}\|\tau^{\beta_h}_{\rm W}) + D_\infty(\rho^0_{\rm Cold}\|\tau^{\beta_h}_{\rm Cold}) - D_\infty(\rho^1_{\rm Cold}\|\tau^{\beta_h}_{\rm Cold}) =: A. \tag{F91}$$

One can use the definition of D_∞ in Eq. (B12) to expand the L.H.S. of Eq. (F91), obtaining

$$\log\max\{(1-\varepsilon)e^{\beta_h E_{k_1}}, \varepsilon e^{\beta_h E_j}\} \leq A - \log Z^{\beta_h}_{\rm W}. \tag{F92}$$

We know that since near perfect work is extracted, ε is arbitrarily small. This implies that for ε small enough, $\max\{(1-\varepsilon)e^{\beta_h E_{k_1}}, \varepsilon e^{\beta_h E_j}\} = (1-\varepsilon)e^{\beta_h E_{k_1}}$.

Similarly, one can evaluate Eq. (F87) to obtain

$$\log\max\{(1-\varepsilon)e^{\beta_h E_{k_2}}, \{\varepsilon p_i e^{\beta_h E_i}\}_{i\neq k_2}\} \leq A - \log Z^{\beta_h}_{\rm W}. \tag{F93}$$

Note that the maximization in Eq. (F93) only picks out the maximum value. In particular, denoting $E_{\max}$ to be the largest energy eigenvalue of the battery, then whenever

$$(1-\varepsilon)e^{\beta_h E_{k_2}} \geq \varepsilon e^{\beta_h E_{\max}}, \tag{F94}$$

or equivalently

$$\varepsilon \leq \left[1 + e^{\beta_h(E_{\max}-E_{k_2})}\right]^{-1}, \tag{F95}$$

then $\max\{(1-\varepsilon)e^{\beta_h E_{k_2}}, \{\varepsilon p_i e^{\beta_h E_i}\}_{i\neq k_2}\} = (1-\varepsilon)e^{\beta_h E_{k_2}}$. In other words, as long as ε is upper bounded by Eq. (F95), we know which terms attains the maximization in Eq. (F92). However, we also want an upper bound that is independent of any limit involving the final state $\rho^1_{\rm ColdMW}$ we wish to take, or any amount of work extracted (and therefore, we want the bound to be independent of E_{k_2}). As such, let us construct the following upper bound $\varepsilon \leq \hat{\varepsilon}$ where,

$$\hat{\varepsilon} := \inf_{\substack{E_{k_2} \\ W_\infty^2 > 0}} \left[1 + e^{\beta_h (E_{\max} - E_{k_2})}\right]^{-1} = \left[1 + e^{\beta_h (E_{\max} - E_j)}\right]^{-1} \tag{F96}$$

Now, we see that E_{k_1} and E_{k_2} correspond to the solutions for Eq. (F92) and Eq. (F93), which for $\varepsilon \leq \hat{\varepsilon}$ reduce to exactly the same equation. Therefore, $E_{k_1} = E_{k_2}$ and hence $W_\infty^1 = W_\infty^2$. □

We will use Lemma 22 to prove Theorem 4. But before we proceed, let us fix some notation: we define the efficiency as a function of $\alpha \geq 0$:

$$\eta_\alpha^J(\rho_{\text{Cold}}^1) = \sup_{E_{k_J} - E_j > 0} \eta(\rho_{\text{Cold}}^1) \quad \text{subject to} \quad F_\alpha(\rho_{\text{W}}^0 \otimes \tau_{\text{Cold}}^0, \tau_{\text{ColdW}}^h) \geq F_\alpha(\rho_{\text{W}}^J \otimes \rho_{\text{Cold}}^1, \tau_{\text{ColdW}}^h), \tag{F97}$$

$$\text{and} \quad \text{tr}(\hat{H}_t \rho_{\text{ColdHotMW}}^0) = \text{tr}(\hat{H}_t \rho_{\text{ColdHotMW}}^{1,J}). \tag{F98}$$

with $J = 1, 2$ denoting the final battery state ρ_{W}^J. We also define an α independent efficiency:

$$\eta^J(\rho_{\text{Cold}}^1) = \sup_{E_{k_J} - E_j > 0} \eta(\rho_{\text{Cold}}^1) \quad \text{subject to} \tag{F99}$$

$$F_\alpha(\rho_{\text{W}}^0 \otimes \tau_{\text{Cold}}^0, \tau_{\text{ColdW}}^h) \geq F_\alpha(\rho_{\text{W}}^J \otimes \rho_{\text{Cold}}^1, \tau_{\text{ColdW}}^h) \, \forall \alpha \geq 0. \tag{F100}$$

For any $\alpha \geq 0$, and any state ρ_{Cold}^1, $\eta_\alpha^J(\rho_{\text{Cold}}^1) \geq \eta^J(\rho_{\text{Cold}}^1)$ holds.

We already know that when $\Omega > 1$, for any final cold bath state ρ_{Cold}^1, the efficiency $\eta^1(\rho_{\text{Cold}}^1)$ is strictly less than the Carnot efficiency. Theorem 4 shows that this is also true for $\eta^2(\rho_{\text{Cold}}^1)$, i.e. when allowing a more general battery final state.

Theorem 4. *[General battery states do not improve efficiency] Consider a heat engine with a cold bath consisting of n qubits, and consider the case where $\Omega > 1$ (recall the definition of Ω in Eq. (E106)). Then for any final cold bath state ρ_{Cold}^1, the efficiency $\eta^2(\rho_{\text{Cold}}^1)$ is strictly less than the Carnot efficiency.*

Proof. Firstly, suppose that $\Omega > 1$. By Lemma 15 we know that the infimum is obtained at $\alpha = \infty$, and by Lemma 16 we know that the efficiency for quasi-static heat engine is strictly less than the Carnot value:

$$\lim_{g \to 0} \eta^1(\tau_{\beta_f}) = \lim_{g \to 0} \eta_\infty^1(\tau_{\beta_f}) < \eta_C. \tag{F101}$$

On the other hand, we also know from Lemma 6 that $\eta^2(\rho_{\text{Cold}}^1)$ can only possibly achieve Carnot efficiency in the quasi-static limit. In other words, for all other final states ρ_{Cold}^1 we know that Carnot efficiency cannot be achieved. Therefore, it suffices to see that in the quasi-static limit,

$$\lim_{g \to 0} \eta^2(\tau_{\beta_f}) \leq \lim_{g \to 0} \eta_\infty^2(\tau_{\beta_f}) = \lim_{g \to 0} \eta_\infty^1(\tau_{\beta_f}) = \lim_{g \to 0} \eta^1(\tau_{\beta_f}) < \eta_C. \tag{F102}$$

The second equality is obtained by noting that for any state $\tilde{\rho}_{\text{Cold}}^1$ (and therefore for τ_{β_f}):

1. ΔC is the same for both expressions of efficiency $\eta_\infty^1(\rho_{\text{Cold}}^1)$ and $\eta_\infty^2(\rho_{\text{Cold}}^1)$.

2. By Lemma 22, for all $0 < \varepsilon < \left[1 + e^{\beta_h (E_{\max} - E_j)}\right]^{-1}$, $W_\infty^1(\tilde{\rho}_{\text{Cold}}^1) = W_\infty^2(\tilde{\rho}_{\text{Cold}}^1)$.

Hence, from Items 1 and 2, one concludes that $\eta_\infty^1(\tilde{\rho}_{\text{Cold}}^1) = \eta_\infty^2(\tilde{\rho}_{\text{Cold}}^1)$. The third equality in Eq. (F102) comes directly from Eq. (F101). □