Laws of thermodynamics beyond the von Neumann regime

Dario Egloff, 1 Oscar C. O. Dahlsten, 2, 3, 1 Renato Renner, 1 and Vlatko Vedral 2, 3

1 Institute for Theoretical Physics, ETH Zürich, 8093 Zurich, Switzerland
2 Center for Quantum Technologies, National University of Singapore, Republic of Singapore
3 Atomic and Laser Physics, Clarendon Laboratory, University of Oxford, Parks Road, Oxford OX1 3PU, United Kingdom

(Dated: May 2, 2014)

A recent development in information theory is the generalisation of quantum Shannon information theory to the operationally motivated smooth entropy information theory, which originates in quantum cryptography research. In a series of papers the first steps have been taken towards creating a statistical mechanics based on smooth entropy information theory. This approach turns out to allow us to answer questions one might not have thought were possible in statistical mechanics, such as how much work one can extract in a given realisation, as a function of the failure-probability. This is in contrast to the standard approach which makes statements about average work. Here we formulate the laws of thermodynamics that this new approach gives rise to. We show in particular that the Second Law needs to be tightened. The new laws are motivated by our main quantitative result which states how much work one can extract or must invest in order to affect a given state change with a given probability of success. For systems composed of very large numbers of identical and uncorrelated subsystems, which we call the von Neumann regime, we recover the standard von Neumann entropy statements.

General introduction.—The relationship between information, quantified by entropy, and work has been the centre of much intriguing and arguably very productive debate, c.f. Maxwell’s daemon [1–3], Szilard’s engine [4], Landauer’s erasure [5], and Bennett’s reversible measurements [6]. To our knowledge connecting smooth entropies [7, 8] with work was first considered in [9]. Particularly important for [9] and our considerations here is the notion of a Szilard engine [4] and Bennett’s extensions thereof [6]. A Szilard engine extracts work from a heat-bath at the cost of using up knowledge about the microstate of the working medium. It was shown in [9] that one should not use the Shannon/von Neumann entropy in general to quantify the extractable work in the Szilard engine, but rather the smooth entropies. This has been followed by several results. In [10] it was shown how to interpret negative conditional entropy in these settings ([9] does not deal with conditional entropy). Very recently, in [11] and independently [12] the non-conditional case is considered in a significantly more sophisticated and general manner than in [9]. These articles taken together indicate that a neat and greatly generalised statistical mechanics, tentatively dubbed single-shot statistical mechanics, is emerging. As mentioned above a key advantage with this approach is that one can answer questions such as “how much work can I extract in any given go (single shot extraction) with a probability x of success”. In standard thermodynamics one would ask what the potential average work output is. To see that in general these are two very different questions, consider a scenario where there is a threshold that needs to be exceeded. One may e.g. need to lift a weight onto a table of given height, or get an electron into the conduction band. Then if there is a significant variance around the average, as is common in nano-scale scenarios, one cannot infer the probability of reaching the threshold from the average alone. It is even possible for the average to be above the threshold whilst with a high probability the threshold is not exceeded.

In this Letter we take this approach much further. We give an expression for the extractable work in a single extraction which reduces to the expressions of [9] [11] [12] in the appropriate limits. We take the system constituting the working medium of the generalised Szilard engine to have a given but arbitrary set of energy levels before the extraction. Moreover there is an arbitrary probability distribution over these levels, representing our knowledge thereof. Similarly we take the post-extraction energy and probability distributions to also be fixed but arbitrary. (In [11, 12] the initial and final energy levels are taken to be the same, and one of the two states is taken to be the Gibbs’s state). Our finding is that the maximal work that can be extracted given these initial and final conditions is determined by a to our knowledge new measure of how much one distribution majorises another. The expression only reduces to the standard von Neumann entropy result in very specific limits, implying that the associated standard laws of thermodynamics need to be modified in general. We use our results to propose new laws of thermodynamics. In particular the second law requires a significant tightening.

Single-shot statistical mechanics-relevant key result.—We begin with briefly reviewing key results that we shall later recover as special cases of our expression. (This is thus not an exhaustive list of all previous results). The details of the models of work extraction in the different papers are not a priori identical, but we shall recover the same expressions within the model here.

In [9] an n-cylinder Szilard engine was considered and the following expression derived:

$$W^\varepsilon = (n - H^\varepsilon_{max}) kT \ln 2.$$ (1)
Here $W^\varepsilon$ is the work that can be extracted in a process with maximum probability of failure $\varepsilon$. $H_{\text{max}}$ is the smooth max entropy of the density matrix representing the work-extracting agent’s initial knowledge about the state of the working medium. This is defined as $H_{\text{max}}(\rho) = \log(\text{rank}(\rho))$, with $\text{rank}(\rho)$ the number of non-zero eigenvalues minimised over all states within $\varepsilon$ trace distance of $\rho$. (Actually there is an alternative definition as well but they are both known to coincide up to an additive $\log \frac{1}{2}$ term, so for simplicity we mention only one definition here.) $T$ is the temperature of the heat bath, and $k$ Boltzmann’s constant. $H_{\text{max}}(\rho)$ reduces to the von Neumann entropy in what we shall here call the von Neumann regime, where $\rho = \tau^\otimes n$, $n \to \infty$ and $\varepsilon \to 0$.

A key result obtained independently in the more recent papers [11, 12] is that given an initial state $\rho$ and a final thermal state $\rho_T$ over the same energy levels, the work that can be extracted with up to $\varepsilon$ failure probability is:

$$W^\varepsilon = kT \log(2) D_\rho(\rho || \rho_T),$$

where $\rho$ is taken to be diagonal in the energy eigenbasis, $\rho_T$ the corresponding thermal state on the same energy levels, and $D_\rho(\rho || \rho_T)$ is the $\varepsilon$-smooth relative entropy of order 0 (see [13]). This reduces to $W = kT \log(2) D(\rho || \rho_T)$ for the standard relative entropy in the von Neumann regime. That latter expression is well-established, see e.g. [14]. Eq. 2 reduces to Eq. 1 in the case of degenerate energy levels, as shown in [11]. In [12] an expression is also given for the inverse process of taking a thermal state to any diagonal state with the same energy spectrum.

The work extraction game.—Our work extraction model can be thought of as a game with simple but minimal rules. (It will nevertheless not be trivial toanalyse as there is a multitude of different strategies one may choose). There are three systems and a work-extraction agent. One system is the working medium, another is a heat bath of temperature $T$, and the last is the work reservoir. The agent wishes to transfer as much energy as possible into the work reservoir in a single extraction. We shall be concerned with quantifying how much energy can be transferred with a maximum failure probability $\varepsilon$, calling this the work $W^\varepsilon$.

The initial energy spectrum $\{E\}$ of the working medium is arbitrary. The initial density matrix $\rho$ of the system is diagonal in the energy basis. The final energy spectrum $\{F\}$ and diagonal density matrix $\sigma$ are also arbitrary. The agent has a few elementary processes it can combine in any way it chooses: (i) it may couple the working medium to the heat bath. This has the effect of changing the probabilities (not the energy levels) in such a way that they approach, by a stochastic matrix, the Gibbs thermal state for the given energy spectrum; (ii) it may change the energy levels (without altering the probabilities) by external intervention taking $\{E\}_j$ to $\{E\}_{j+1}$ where $j$ labels the time step. Here the energy must be accounted for by being taken from or given to the work reservoir. In a given realisation the system is in one, possibly unknown, energy eigenstate and only changes to that particular eigenstate cost or yield work. The combination of these elementary processes the agent chooses is called it’s strategy. Energy conservation is assumed throughout: any energy leaving or entering the system must enter or leave the heat bath and/or the work-extraction system.

**Relative mixedness.**—We now introduce a measure of how much more mixed one state $\rho$ is than another, $\sigma$, calling this the relative mixedness $M(\rho || \sigma)$. The definition of $M(\rho || \sigma)$ will later be justified by operational statements we will make.

**Definition 1** (Relative mixedness). The relative mixedness of two states $\rho$ and $\sigma$ with compact support and descending-order spectra $f(x_1)$, $g(x_2)$ respectively is given by

$$M(\rho || \sigma) := \max m : \int_0^l f(x_1) dx_1 \geq \int_0^{lm} g(x_2) dx_2 \forall l,$$

For states with discrete spectra $\{\lambda_i\}$ one evaluates $M$ for the associated step-function where the $i$-th ‘block’ has constant height $\lambda_i$ and all blocks have width 1.

If and only if $M \geq 1$ does (the spectrum of) $\rho$ majorise $\sigma$, $\rho \succ \sigma$. The actual number $M$ can be viewed as a measure of by how much.

**Gibbs-rescaling.**—We shall make use of a powerful insight from [15,17]; it is also used in [12]. The insight bridges a particular gap between information theory and statistical mechanics: the fact that the former does not care about energy. In information theory, the Shannon/von Neumann entropy of a state, $-\sum_i \lambda_i \log \lambda_i$ is independent of the energies of the states involved. As the extractable work is expected to depend on the energy levels involved it follows that it is not expected to be uniquely determined by an entropy.

A key way in which energy enters into statistical mechanics is that in a Gibbs state the probability of any given energy eigenstate state with energy $E$ is given by $p_T(E) = \exp(-\frac{E}{kT})/Z$, where $Z$ is the partition function. The insight we adapt from [15,17] is that we can take this bias into account by what essentially amounts to rescaling the density matrix’s eigenvalue distribution by $p_T(E)$. After the rescaling the occupation probabilities will turn out to uniquely determine our expression for the extractable work. More specifically, we shall be employing an operation we term Gibbs-rescaling to the eigenvalue spectrum. Consider states with discrete spectra $\{\lambda_i\}$. We firstly transform the spectrum into the associated step-function. Then we take each block, rescale its height as $\lambda_i \mapsto \lambda_i/\exp(-\frac{E_i}{kT})$, and its width $l = 1 \mapsto e^{-\frac{E_i}{kT}}$ such that the area of the new block is $\lambda_i$ as before. We write this operation applied to a density matrix $\rho$ as $G_T(\rho)$.

**The main theorem.**—Having defined the relative mixedness $M(\cdot || \cdot)$ and Gibbs-rescaling $G_T(\cdot)$ we can now give the main result.
Theorem 1. In the work extraction game defined above, consider an initial density matrix $\rho = \sum_i \lambda_i |e_i\rangle \langle e_i|$ and final density matrix $\sigma = \sum_j \nu_j |f_j\rangle \langle f_j|$ with $\{|e_i\rangle\},\{|f_j\rangle\}$ the respective energy eigenstates and both $\rho$ and $\sigma$ having finite rank. Let the work one can extract except with a probability $\varepsilon$ of failure using strategy $S$ be denoted $W^\varepsilon_S(\rho \rightarrow \sigma)$. For any strategy this respects $W^\varepsilon_S(\rho \rightarrow \sigma) \leq W^\varepsilon(\rho \rightarrow \sigma)$, where

$$W^\varepsilon(\rho \rightarrow \sigma) = kT \ln \left( \frac{M(G_T(\rho), G_T(\sigma))}{1 - \varepsilon} \right).$$

One may, under a certain proviso, construct a strategy that saturates the bound in the sense that $W^\varepsilon(\rho \rightarrow \sigma)$ is achieved except with probability of $\varepsilon$; the proviso is that the agent can access a single extra two-level system which is fixed to be in one of its energy eigenstates $|\xi\rangle$ both initially and finally, i.e. $\rho = \ldots \otimes |\xi\rangle \langle \xi| \ldots$ and $\sigma = \ldots \otimes |\xi\rangle \langle \xi| \ldots$.

The proofs are in the technical supplement. A rough intuition for the bound is that the Gibbs-rescaling is needed to take into account the bias imposed on the energy levels by the Gibbs statistics, and up to that bias, only the amount of majorisation matters as the work extraction process lowers the majorisation amount, the relative mixedness. See Figure 1 for a simple example of how to apply the theorem.

We now discuss the implications of the statement and develop the argumentation further. We focus on the laws of thermodynamics. There are two related reasons why some of the standard laws need modification: (i) they are about averages whereas we are making single-shot statements, (ii) they are based on the von Neumann entropy which we found is not the appropriate quantity away from the von Neumann regime.

0th and 1st laws of thermodynamics.—The 0th law can be stated as: There exists for every thermodynamic system in equilibrium a property called temperature. Equality of temperature is a necessary and sufficient condition for thermal equilibrium. This also holds after our generalisation. In particular we are still assuming heat baths that take the working medium closer to a Gibbs thermal state upon interaction.

The first law can be viewed as both asserting the conservation of energy as well as stating that it can be divided into two parts, work and heat, which are normally defined in the description accompanying the first law equation: $dU = dQ - dW$. $U = \text{tr}(\rho H)$ is the expected internal energy of the working medium with Hamiltonian $H$, $Q$ is ‘heat’, and $W$ ‘work’. The associated physical setting is that there is a system, the working medium, which can either exchange energy with another system in a thermal state dubbed a heat bath, or with a work reservoir system normally implicitly assumed to be in some energy eigenstate of its own Hamiltonian. Exchanges of energy with the heat bath are dubbed heat and those with the work reservoir work. This essentially carries over into our approach but with some important subtleties. We assume energy conservation (in every single extraction), as well as allowing for interactions with a heat bath and a work reservoir. Thus the following is respected: $dE_{\text{sys}} = -dE_{\text{bath}} - dE_{\text{reservoir}}$. We, more subtly, break $dE_{\text{reservoir}}$ into two parts: $dE_{\text{reservoir}} = dW^\varepsilon_S + dE_{\text{extra}}$. There is the energy transfer which is predictable (up to $\varepsilon$ probability of failure) in that it corresponds to $dW^\varepsilon_S(\rho \rightarrow \sigma)$ for the infinitesimal state change $\rho \rightarrow \sigma$ using strategy $S$. We view anything beyond that, given by $dE_{\text{extra}}$, as heat. The idea behind this is that only predicted energy transfer should count as work. One may for example imagine buckets lifting water out of a mine up to a certain height (or as a quantum example an electron excited into the conduction band). The height at which the buckets are tipped into a reservoir is specified in advance. If they go higher than this, the extra potential energy will be transferred to other degrees of freedom associated with the reservoir system, e.g. into movement of the water or heating of the semi-conductor. We may express the following first law for this approach: In any given extraction, with probability $p \geq 1 - \varepsilon$

$$dE_{\text{sys}} = -dE_{\text{bath}} - dW^\varepsilon_S - dE_{\text{extra}} \equiv dQ - dW^\varepsilon_S. \quad (4)$$

Second law.—Consider next the so-called Kelvin statement of the second law: No process is possible in which the sole result is the absorption of heat from a reservoir and its complete conversion into work. This does not say anything about processes with a non-zero probability of failure. We show in the appendix that
for given states of the working medium A and B respectively, \( W^\epsilon(A \rightarrow B) + W^\epsilon(B \rightarrow A) \leq W^{2\epsilon}(A \rightarrow A) \). We call this the ‘triangle inequality’. It implies together with the main theorem that all strategies in our game respect the following generalisation of Kelvin’s second law:

\[
\sum_{i=0}^{m-1} W_{S_i}^\epsilon(A_i \rightarrow A_{i+1}) \leq W_{m\epsilon}^\rho(A \rightarrow A) \quad \text{if} \quad A_m = A_1,
\]

where \( S_i \) is the choice of strategy in the \( i \)-th step of the cycle. Note that \( W^\rho(A \rightarrow A) = 0 \) (see main theorem), implying that deterministically no work can be extracted in such a cycle. One may still gain work in a single cycle at the cost of having \( \epsilon > 0 \) for one or more of the steps.

The second law is also closely related to entropy increasing with time and one may wonder what the corresponding generalisation of the statement is. A particular standard expression is that

\[
\Delta (S - \beta \langle E \rangle) \geq 0,
\]

where \( S \) and \( \langle E \rangle \) are the von Neumann entropy and expected energy of a system interacting with a heat-bath with inverse temperature \( \beta \). (\( \Delta \) indicates the change in these values during the interaction.) This actually still holds in our more general model; we show this in the technical supplement. However, crucially, it is not sufficient to guarantee that an evolution is possible. Instead it should be replaced by the statement that a state change \( \rho \rightarrow \rho' \) due to a thermalisation with a heat-bath at temperature \( T \) is possible if and only if

\[
W^\rho(\rho \rightarrow \rho') \geq 0.
\]

This is significant as there are processes that respect Eq. \([5]\) but violate Eq. \([6]\). A simple example is to consider degenerate energy levels, so that \( \Delta \langle E \rangle = 0 \), and three levels with probabilities \((1/2 \ 1/2 \ 0)^T \rightarrow (2/3 \ 1/6 \ 1/6)^T \). Then \( \Delta S \approx 0.25 \) but \( W^\rho \) is negative. Strikingly, such evolutions enable the deterministic violation of Kelvin’s second law (see technical supplement).

The inequivalence of entropy and majorisation has been noted previously in the context of the second law \([15, 16]\). Presumably this has not received more attention to date because in the von Neumann regime this inequivalence disappears. More precisely, if we consider a tensor product of \( n \) identical states each with von Neumann entropy \( S \) and let \( n \rightarrow \infty \), then with asymptotically small error we may approximate the spectrum as a uniform distribution with value \( 2^{-nS} \) in the range \([0.2^{-nS}] \). For such distributions the partial orders induced by \( S \) and majorisation respectively coincide.

**Recovering existing results.**—Eq. \([2]\) above from \([11, 12]\) and accordingly \([1]\) from \([9]\) are special cases of our main result—see the technical supplement. Eq. \([2]\) corresponds to the case where the final state \( \rho_T \) is demanded to have the same eigenspectrum and be a Gibbs state \( (\rho_T = \sum \rho_T(E_i) |e_i \rangle \langle e_i|) \). (It was also shown in Fig. 1 that a generalisation of Landauer’s principle is a special case).

**Outlook.**—The results could be made more general still by considering small heat baths, along the lines of \([18]\). We also note the striking similarities between what is discussed above and the question of quantifying entanglement in the non-asymptotic regime. It was shown in a seminal paper by Nielsen \([19]\) that majorisation is the central quantity there. We anticipate that many of the results from our work can be applied also in that context.

**Acknowledgements.**—We gratefully acknowledge discussions with J. Aaberg, J. Baez, B. Fong, J. Vicary and P. Perinotti, as well as support from the National Research Foundation (Singapore), the Ministry of Education (Singapore), the Swiss National Science Foundation (grant No.200020-135048) and the European Research Council (grant No.258392). This research was partly carried out in connection with DE’s Master’s thesis at ETH Zurich.

[1] Maxwell, J. Theory of Heat (Longmans, Green, London, 1985), 4 edn.
[2] Leff, H. & Rex, A. F. Maxwell’s demon 2: Entropy, classical and quantum information, computing (CRC Press, Abingdon, 2003).
[3] Maruyama, K., Nori, F. & Vedral, V. The physics of Maxwell’s demon and information. Rev. Mod. Phys 81, 1–23 (2009).
[4] Szilárd, L. Über die Entropieverminderung in einem thermodynamischen System bei Eingriffen intelligenter Wesen. Z. Phys. 53, 840–856 (1929).
[5] Landauer, R. Irreversibility and heat generation in the computing process. IBM J. Res. Dev 5, 183 (1961).
[6] Bennett, C. H. Notes on Landauer’s principle, reversible computation, and Maxwell’s demon. Stud. Hist. Philos. Sci. B 34, 501–510 (2003).
[7] Renner, R. Security of Quantum Key Distribution. Ph.D. thesis, ETH Zürich (2005). Also available as quant-ph/0512258.
[8] Renner, R. & Wolf, S. Smooth Rényi entropy and applications. In Proc. IEEE Int. Symp. Info. Th., 233 (2004).
[9] Dahlsten, O. C. O., Renner, R., Rieper, E. & Vedral, V. Inadequacy of von Neumann entropy for characterizing extractable work. New. J. Phys. 13, 053015 (2011).
[10] Rio, L., Aberg, J., Renner, R., Dahlsten, O. & Vedral, V. The thermodynamic meaning of negative entropy. Nature 474, 61–63 (2011).
[11] Aberg, J. Truly work-like work extraction. pre-print arXiv:1110.6121 (2011).
[12] Horodecki, M. & Oppenheim, J. Fundamental limitations for quantum and nano thermodynamics. pre-print arXiv:1111.3834 (2011).
[13] Datta, N. Min- and max-relative entropies and a new entanglement monotone. Information Theory, IEEE Trans-
actions on 55, 2816–2826 (2009).

[14] Donald, M. Free energy and the relative entropy. *Journal of Statistical Physics* 49 (1987).

[15] Ruch, E. & Mead, A. The principle of increasing mixing character and some of its consequences. *Theor. Chim. Acta* 41, 95 (1976).

[16] Ruch, E. The diagram lattice as a structural principle. *Theor. Chim. Acta* 38, 167 (1975).

[17] Mead, A. Mixing character and its application to irreversible processes in macroscopic systems. *J. Chem. Phys.* 66, 95 (1977).

[18] A. E. Allahverdyan, R. Balian & Th. M. Nieuwenhuizen. Maximal work extraction from finite quantum systems. *Europhys. Lett.* 67, 565–571 (2004).

[19] Nielsen, M. Conditions for a class of entanglement transformations. *Phys. Rev. Lett.* 83 (1999).

[20] Hardy, G. H., Littlewood, I. E. & Polya, G. *Inequalities* (CUP, 1952).

[21] McDiarmid, C. *On the method of bounded differences* (London Mathematical Society Lecture Note Series 141, CUP, 1989).
Appendix A: Upper bounding $W^c$

In this section we define the setting more carefully and derive the upper bound on $W^c$. We first give certain definitions and lemmas which will be needed.

**Definition 2** (Gibbs rescaling). Consider a density matrix $\rho = \sum_{i=1}^{n} \lambda_i |e_i \rangle \langle e_i|$ with eigenvalues $\{\lambda_i\}_{i=1}^{n}$ and take the energy eigenstates of the system to be $\{|e_i\rangle\}_{i=1}^{n}$ with energies $\{E_i\}_{i=1}^{n}$ respectively. There is an associated step function for the spectrum, $\lambda(xn) = \lambda_{[xn]}$ where $x \in (0,1)$. Similarly there is an energy step function $E(xn) = E_{[xn]}$ where $x \in (0,1)$. The Gibbs rescaling associated with temperature $T$ combines $\lambda(xn)$ and $E(xn)$ to a new function $G_T(y)$ given by

$$G_T \left( \int_0^x e^{-\frac{E_{[xn]}}{kT}} dy \right) = \frac{\lambda_{[xn]}}{e^{-\frac{E_{[xn]}}{kT}}}.$$ 

It follows that $G_T(y)$ is defined on $(0,Z)$, with $Z = \sum_{j=1}^{n} \exp \left( -\frac{E_j}{kT} \right)$ the partition function. Moreover $G_T(y)$ is a probability distribution satisfying $\int_0^Z G_T(y) dy = 1$.

We now move towards defining the work extraction game. The essential idea is that the work extracting agent wants to extract work under the restriction that the pre- and post-extraction states are given, both in terms of the energy spectra and the occupation probabilities of the energy levels. The agent must choose a strategy before the extraction and then not intervene further during the extraction. A strategy is defined by the action the agent takes at each step, with the steps labelled as $j$. The actions it can choose from, which will be defined in detail later, are of two types: (i) thermalisations (changing occupation probabilities towards the Gibb’s state) or (ii) work extractions (changing a specified energy level but not energy eigenvalues). We will say there are $m$ work extractions in total (this may tend to infinity).

In a given realisation the system will be in a certain energy level to start with (though it is not in general know to the agent which one). It may hop between energy levels during thermalisations. When a level is changed it costs energy only if the system occupies that level.

**Definition 3** (Notation). $\bar{s} \in \{0,1\}^m$ : a vector with one entry for each of $m$ work extractions (subsequently called “steps”): $s_j = 1$ : system is in chosen states for work extraction $s_j = 0$: system is not in chosen states for work extraction $\bar{s}_j$ is the complement of $s_j$: $s_j = 1 \iff \bar{s}_j = 0$ and $s_j = 0 \iff \bar{s}_j = 1$ $w^z_j$: logarithmical work ($kT \ln(w^z_j) = W^z_j$) one extracted in the step $j$ on the path $\bar{s}$. $w^s_j$: The logarithmical work one extracts in step $j$ if the specified level is occupied. $w$: total logarithmical work demanded in order to call the total extraction successful. $G$ is the set of the successful paths, i.e. those yielding as much work as demanded:

$$G = \left\{ \bar{s} \mid \prod_{j=1}^{m} w^z_j \geq w \right\}$$

$\eta^s_j$: probability of doing step $j$ on the path $\bar{s}$. $P_S$: total probability of success: $P_S = \sum_{\bar{s} \in G} \prod_j \eta^s_j$. $\phi^s_j$: state of the system after step $j$ if the previous evolution of the system is given by the path $\bar{s}$. $\lambda^s_j$: eigenvalues of state $\phi^s_j$. $p^s_j = G \left( \phi^s_j \right)$: Gibbs rescaled probability distribution after step $j$ (before thermalizing) conditioned on the previous steps on path $\bar{s}$. $p^s_{j\perp}$: Gibbs rescaled probability distribution after step $j$ (after thermalizing) conditioned on the previous steps on path $\bar{s}$. $q$: final Gibbs rescaled probability distribution, conditioned on successful work extraction:

$$q = \sum_{\bar{s} \in G} \prod_{j} \eta^s_j \prod_{\bar{s}_j \perp} P_S$$

$B_j$: Bistochastic matrix one chooses after step $j$ by thermalizing the system (this has to be the same for all paths). $E^s_j(x)$: Energy of the level labelled by $x$ after step $j$. $\Theta_U(x)$: Step function on $U$:

$$\Theta_U(x) = \begin{cases} 1 & : \text{for } x \in U \\ 0 & : \text{else} \end{cases}$$

**Definition 4** (Block). For $a < b$ the interval $(a, b]$ is said to be a block corresponding to a level $k$, if $p^s_{j \perp}$ is constant on this interval $\forall \bar{s}$.

We now turn to how interactions with the heat bath, thermalisations, act on the state of the system. Roughly speaking these take the density matrix closer to the associated Gibb’s state. As already mentioned the thermalisation is taken to only change occupation probabilities and not energy eigenvalues. We take the thermalisation to act as a stochastic process on the energy eigenstates, in that the probability of occupying a given energy state, $P(i)$, becomes $P'(i) = \sum_j P(j \rightarrow i)P(j)$ where the summation is over all eigenstates, $P(j \rightarrow i)$ is a transition probability, and $P(j)$ an occupation probability (before the interaction with the heat bath). This can equivalently be written as $\hat{P}' = B\hat{P}$ where $B$ is a stochastic matrix (entries are probabilities and columns sum to 1).

Not any stochastic matrix $B$ is allowed however. The Gibb’s state (associated with temperature $T$) is taken to be invariant under a thermalisation. Consider the implications firstly for the fully degenerate case of all energies being the same. In this case the Gibb’s state is the uniform distribution. The only stochastic matrices that
leave the uniform distribution invariant are bistochastic ones (rows also sum to 1). Thus in the fully degenerate case $B$ must be bistochastic. We see no reason to impose further restrictions, so any such $B$ is allowed. (Choosing different allowed $B$’s is taken to correspond physically to choosing different times of coupling to the heat bath.)

Consider secondly the non-degenerate case. Here it is again convenient to use the Gibb’s rescaled distribution. Note that the Gibb’s state becomes uniform after the Gibb’s rescaling. Thus one may hope to model a thermalisation as a bistochastic matrix on the Gibb’s-rescaled distribution.

Consider dividing the Gibb’s-rescaled distribution into fine blocks such that all fine blocks have the same width $w$. Let $N$ be the number of fine blocks. (As the total support is given by the partition function $Z$ we have $w = Z/N$.) Let $N_k$ be the number of fine grained blocks associated with level $k$, such that $\sum_{k=1}^{N} N_k = N$. Each energy level is associated with one block only labelled by $k$. Each $l$-th fine block is associated with a level $k_l$.

Fine blocks associated with the same energy level $k$ must all have the same height, given by $P(k_l)/e(k_l)$. Let $\bar{f}$ contain the $N$ heights of the fine blocks, with $P(k_l)/e(k_l) = P(k_l)N/(ZN_k)$ as its $l$-th entry. Now when the occupation probabilities transform under $B$, $\bar{f}$ undergoes an associated transform. We will argue it is given by a matrix $F$ whose entry in the $l$-th row and $m$-th column is given by

$$F_{lm} := \frac{B_{k_l k_m}}{N_k}.$$ 

To see this note firstly that $P_{l} = \sum_j B_{ij} P_j = \sum_j \frac{N_j}{N} B_{ij} P_j$, and recall that $f_{l} = P_{l}N/(ZN_k)$. Thus

$$f_{l} = \sum_{j=1}^{N} \frac{N_j}{N} B_{kj} P_j \frac{N}{ZN_k} \quad (A1)$$

$$= \sum_{j=1}^{N} \frac{N_j}{N} B_{kj} P_j \frac{N}{ZN_k} \quad (A2)$$

$$= \sum_{j=1}^{N} \frac{B_{kj}}{N_k} f_m \quad (A3)$$

$$= \sum_{m=1}^{N} \frac{B_{km}}{N_k} f_m. \quad (A4)$$

As $B_{ij}$ and $N$ are nonnegative real numbers $F$ has non-negative real entries only. To see that the columns sum to 1 so that $F$ is a stochastic matrix, note that the column sums are the same as for $B$ which is stochastic. Moreover as $B$ must leave the Gibb’s state invariant, and this is a uniform distribution after the Gibb’s rescaling, $F$ must leave the uniform distribution (or anything proportional to it) invariant. Then for any row $i$, $\sum_j F_{ij}(1/N) = 1/N$ so each row of $F$ must sum to 1. Therefore $F$ is a bistochastic matrix. Note that $F$ is additionally restricted, through being defined via $B$, to keep the heights of fine blocks the same whenever these are associated with the same level.

Accordingly we define interactions with the heat-baths, thermalizations, to act in the following way on the system.

**Definition 5** (Thermalization). If after step $j$ one chooses to do a thermalization the energy levels are invariant. The occupation probabilities transform under a stochastic matrix $B_j$. $B_j$ leaves the Gibb’s state invariant. This implies as shown above that any fine-grained Gibb’s-rescaled distribution where all blocks have same width transforms under a bistochastic matrix $F$ defined above. $F$ cannot introduce differences between heights of fine blocks associated with the same energy level.

The only assumptions we have used in the above to arrive at the above definition are:

1. Only the occupation probabilities of the levels change.

2. The occupation probabilities of the energy levels transform according to a stochastic matrix.

3. The Gibb’s state (of temperature $T$) is invariant (when heat bath has temperature $T$).

Note that a permutation matrix is bistochastic so one may permute the levels in the thermalization step. We use this to simplify the notation of the work extraction definition. We take the levels lowered or raised to form the first $l$ levels.

**Definition 6** (Work extraction). To do a work extraction in step $j$ one first defines an interval $(0,a] = \bigcup_{k=1}^{l} (a_{k-1}, a_k)$, where $(a_{k-1}, a_k)$ correspond to levels $\{1, \ldots, l\}$, on which one wants to change the energy by $\Delta E = -kT \ln (w_{\eta_1}^{(s_j)} = 1)$. The remaining levels remain untouched.

For $x \in (0,1]$, the eigenvalues of the levels after step $j$, conditioned on the previous state are given by:

In the case $s_j = 1$ (state of the system is found to be in the levels corresponding to $(0,a)$):

$$\lambda_{\eta_{1,s_j=1}}^{j}([x,n]) = \Theta_{(0,a)}(Z_{j-1} x) \frac{\lambda_{x}^{j-1}([x,n])}{\eta_{\eta_{1,s_j=1}}^{j}}$$

where $\eta_{\eta_{1,s_j=1}}^{j} = \int_{0}^{a} \lambda_{x}^{j-1} \left( \frac{[x,n]}{Z_j} \right) dx$.

In the case $s_j = 0$ (state of the system is not in the levels corresponding to $(0,a)$):

$$\lambda_{\eta_{1,s_j=0}}^{j}([x,n]) = \Theta_{(a,Z_{j-1})}(Z_{j-1} x) \frac{\lambda_{x}^{j-1}([x,n])}{\eta_{\eta_{1,s_j=0}}^{j}}$$

where $\eta_{\eta_{1,s_j=0}}^{j} = \int_{a}^{Z_{j-1}} \lambda_{x}^{j-1} \left( \frac{[x,n]}{Z_j} \right) dx$. In this case there is no change in the work reservoir: $w_{\eta_{1,s_j=0}}^{j} = 1$. 

Definition 7 (Our work extraction game). There are three systems and a work-extraction agent. One system is the working medium, another is a heat bath of temperature $T$, and the last is the work reservoir.

The initial energy spectrum $\{E\}$ of the working medium is arbitrary but given. The initial density matrix $\rho$ of the same is diagonal in the energy basis. The final energy spectrum $\{F\}$ and diagonal density matrix $\sigma$ are also arbitrary but given.

The agent can combine thermalization (defined above) and work extraction(also defined above) in any sequence. This sequence, together with the specifications for each step is called its strategy.

A given work extraction will transfer some energy $\nu$ to the work extraction reservoir. Before the extraction the agent must specify $W$. If $\nu \geq W$ the work extraction is termed successful. The probability of success is called $1 - \varepsilon$.

We shall be interested in bounding $W$ given $\varepsilon$ and the initial and final conditions. We break the calculation into several lemmas which will later be combined to prove the main theorem.

We next consider how much work extraction flattens the Gibbs rescaled distribution.

Lemma 2. For $x \in (0, Z_j]$ (with $Z_j$ the partition function after step $j$), by doing a work extraction in step $j$, the two possible Gibbs rescaled probability distribution after the step, conditioned on the previous steps on path $\vec{s}$ are given by the following. In the case where $s_j = 1$:

$$p^j_{\vec{s}|s_j=1}(x) = \Theta_{(0, aw^j s_j=1)}(x) \frac{p^{j-1}_x (\frac{x}{w^{j-1} s_j=1} + a)}{w^{j-1} s_j=1 \eta^{j-1}_s s_j=1}$$

In the case where $s_j = 0$:

$$p^j_{\vec{s}|s_j=0}(x) = \Theta_{(aw^{j-1} s_j=1, Z_j)}(x) \frac{p^{j-1}_x (x - aw^{j-1} s_j=1 + a)}{\eta^{j-1}_s s_j=0}$$

Proof. Case $s_j = 1$:

Let $w = w^j, p^j = \eta^j, p^{j-1} = \eta^{j-1}, \lambda^j = \lambda^j_s$ and $\eta^j = \eta^j_s$

Let $x \in (0, aw] \cap (0, Z_j]$ and $b \in (0, \infty)$ such that

$$\int_0^b \exp \left( - \frac{E^{j-1}_w}{kT} \right) \, dy = x.$$

$$p^j(x) = p^j \left( \int_0^b \exp \left( - \frac{E^{j-1}_w}{kT} \right) \, dy \right)$$

$$= p^j \left( \int_0^b \exp \left( - \frac{E^{j-1}_w}{kT} - kT \ln(w) \right) \frac{1}{w} \, dy \right)$$

$$= p^j \left( \int_0^{b/w} \exp \left( - \frac{E^{j-1}_w}{kT} - kT \ln(w) \right) \, dz \right)$$

$$= \frac{\lambda^j_w}{\exp \left( - \left( \frac{E^{j-1}_w}{kT} - kT \ln(w) \right) \right)}$$

$$= \exp \left( - \left( \frac{E^{j-1}_w}{kT} - kT \ln(w) \right) \right) wp^j$$

$$= \frac{1}{wp^j} \left( \frac{b}{w} \exp \left( - \left( \frac{E^{j-1}_w}{kT} \right) \right) \right)$$

$$= \frac{1}{wp^j} \left( \frac{b}{w} \right)$$

One easily sees that $p^j(x) = 0$ for $x \geq \frac{a}{w}$.

The proof for the case $s_j = 0$ is analogous. 

The next lemma is in contrast concerned with how much wider the Gibbs rescaled distribution is after the work extraction (recall that the distribution has support on $(0, Z]$ where $Z$ is the partition function).

Lemma 3. The partition function $Z_j$ immediately after step $j$ is given by:

$$Z_j = Z_{j-1} + a(w_{j-1} - 1)$$

Proof. Let $(0, a]$ be an interval consisting of blocks corresponding to the levels $\{1, \ldots, l\}$ and $n$ such that it can
split the interval \((a, Z_j]\) into \(n - l\) blocks.

\[
Z_j = \sum_{k} e^{-\frac{E_s}{kT} + \frac{\eta}{kT}} = \sum_{k=1}^{l} e^{-\frac{E_s}{kT}} \sum_{k-l+1}^{n} e^{-\frac{E_s}{kT}} = w_1 \sum_{k=1}^{l} e^{-\frac{E_s}{kT}} + \sum_{k-l+1}^{n} e^{-\frac{E_s}{kT}} = w_1 a - a + \sum_{k=1}^{l} e^{-\frac{E_s}{kT}} + \sum_{k-l+1}^{n} e^{-\frac{E_s}{kT}} = Z_{j-1} + aw_1 - a
\]

out of which the lemma follows.

We now combine the two previous lemmas to gain another relation between the Gibbs rescaled distribution and steps \(j\) and \(j - 1\). We shall use this later in an iterative manner to relate the very first and final Gibbs rescaled distributions.

**Lemma 4.** The Gibbs rescaled probability distributions at steps \(j\) and \(j - 1\) respectively satisfy the relation

\[
p_{s_{l},t}^{j-1}(x) = \sum_{k=0,1} w_s^j n_{s_{j},s_{j}} = k \eta_{s_{j},s_{j}} = k | \eta_{s_{j},s_{j}} = k | \eta_{s_{j},s_{j}} = k | \eta_{s_{j},s_{j}} = k | \eta_{s_{j},s_{j}} = k | \eta_{s_{j},s_{j}} = k | \eta_{s_{j},s_{j}} = k | \eta_{s_{j},s_{j}} = k | \eta_{s_{j},s_{j}} = k | \eta_{s_{j},s_{j}} = k
\]

with constants \(c_{s_{j},s_{j}} = 0\) and \(c_{s_{j},s_{j}} = aw_j - a\).

**Proof.** Let \(w_k = w_s^j n_{s_{j},s_{j}} = k\), \(p_k^j = p_{s_{j},s_{j}} = k\), \(p^j_{s_{j},s_{j}} = p^j_{s_{j},s_{j}} = k\), \(\eta_k = \eta_{s_{j},s_{j}} = k\).
Let \(c_0 = aw_1 - a\) and \(c_1 = 0\). Then:

\[
\eta_0 w_0 p_0^j(xw_0 + c_0) + \eta_1 w_1 p_1^j(xw_1 + c_1) = \eta_0 w_0 p_0^j(xw_1 - a) + \eta_1 w_1 p_1^j(xw_1) = \Theta(a,w_1)(xw_1 - a) / p_0^j(xw_1) p_1^j(xw_1) = \Theta(a,\eta_0-w_0+ax)(x)p_0^j(xw_1) + \Theta(a,\eta_1)(x)p_1^j(xw_1) = \Theta(a,\eta_0)(x)p_0^j(xw_1) + \Theta(a,\eta_1)(x)p_1^j(xw_1)
\]

\[
\sum_{\sigma \in \{0,1\}^j} \int_0^l \tau_t \circ p_{s_{l},t}(x) \, d\sigma
\]

where \(\tau_t\) is the permutation, which maximizes the left hand side, while \(\tau_t^{j+1}\) is the one which maximizes the right hand side.

**Proof.** Let \(p_1 = p_{s_{l},t}^{j+1}, p_0 = p_{s_{l},t}^{j+1}, \eta_1 = \eta_{s_{l},t}^{j+1}, \eta_0 = \eta_{s_{l},t}^{j+1}, w = u_{l}^{j+1}\)

\[
\sum_{\sigma \in \{0,1\}^j} \int_0^l \tau_t \circ p_{s_{l},t}(x) \, d\sigma
\]

where \(\tau_t\) reorders \(\sum_{\sigma \in \{0,1\}^j} p_1 \in (aw, Z_j]\) in descending order in \((aw, Z_j]\). This is possible since \(p_1\) and \(p_0\) have disjoint support, also for different \(s\), since a in definition 6 has to be chosen independently of the path. (See lemma 2).

\[
l_1 \in (0, \min[a, t])\]

is a value which maximizes the right hand side of the last line. Using the same argument backwards after changing variables, we get:

\[
\sum_{\sigma \in \{0,1\}^j} \int_0^l \tau_t \circ p_{s_{l},t}(x) \, d\sigma
\]

Applying any bistochastic matrix \(B\) on the probabilities \(p_0\) and \(p_1\) and reordering in descending order with \(\tau_t^{j+1}\)
afterwards, we get:

\[
\sum_{\vec{s} \in \{0,1\}^j} \int_0^t \tau^j_t \circ p^j_{\vec{s}_0,t}(x) \, dx \\
\geq \int_0^t \tau^{j+1}_t \circ B \circ (\tau^{j+1})^{-1} \circ \\
\sum_{\vec{s} \in \{0,1\}^j} \left( \eta_1 w_t^{j+1} \circ p_1(xw) + \eta_0 \tau^{j+1}_t \circ p_0(x) \right) \, dx \\
= \sum_{\vec{s} \in \{0,1\}^j} \int_0^t \left( w_1 \eta_1^{j+1} \circ B \circ p_1(xw) + \eta_0 \tau^{j+1}_t \circ B \circ p_0(x) \right) \, dx \\
= \sum_{\vec{s} \in \{0,1\}^j} \int_0^t \left( w_1 \eta_1^{j+1} \circ p_1(xw) + \eta_0 \tau^{j+1}_t \circ p_0(x) \right) \, dx
\]

Where the inequality follows out of the inequality \( Bp \succeq p \) for any bistochastic matrix \( B \) and vector \( p \), which is proved in [20].

Therefore (with \( P_S = 1 - \varepsilon \)):

\[
W^\varepsilon = kT \ln(w) \\
\leq kT \ln \left( \max \left\{ m \int_0^t p(x_1) dx_1 \geq \int_0^t (1 - \varepsilon) q(x_2) dx_2 \forall t \right\} \right) \\
= kT \ln \left( M \left( \frac{G(p)}{1 - \varepsilon} \right) \right).
\]

This proves the first part of the main theorem.

\( \Box \)

**Theorem** (First part of Theorem 1 in main body, giving the bound). In the work extraction game defined above, if one is given an initial density matrix \( \rho = \sum \lambda_i |e_i\rangle \langle e_i| \) and final density matrix \( \sigma = \sum j |w_j| f_j \rangle \langle f_j| \) with \( \{|e_i|\} \) and \( \{|f_j|\} \) the respective energy eigenstates and both \( \rho \) and \( \sigma \) having finite rank, then the work \( W^\varepsilon \) one can extract with certainty except with \( \varepsilon \) probability respects

\[
W^\varepsilon \leq kT \ln \left( M \left( \frac{G(p)}{1 - \varepsilon} \right) \right).
\]

**Proof.** Define \( p^0_\vec{S} = p \). W.l.o.g. \( \vec{s} = \{0, \ldots, 0\} \) (the first probability distribution is independent of the path afterwards). Inductively using lemma 5 one gets:

\[
\int_0^t p(x) \, dx \\
= \int_0^t p^0_\vec{S}(x) \, dx \\
\geq \sum_{\vec{s} \in \{0,1\}^m} \int_0^t \left( \prod_{j=1}^m \eta_{j}^l(xw) \right) \tau^m_t \circ p_\vec{S}(x) \\
= \sum_{\vec{s} \in \{0,1\}^m} \int_0^t \left( \prod_{j=1}^m \eta_{j}^l(xw) \right) \tau^m_t \circ p_\vec{S}(x) \\
\geq \sum_{\vec{s} \in \{0,1\}^m} \int_0^t \left( \prod_{j=1}^m \eta_{j}^l \right) \tau^m_t \circ p_\vec{S}(x) \\
= P_S \int_0^t q(x) \, dx
\]

where \( P_S = 1 - \varepsilon \):

\[
W^\varepsilon = kT \ln(w) \\
\leq kT \ln \left( \max \left\{ m \int_0^t p(x_1) dx_1 \geq \int_0^t (1 - \varepsilon) q(x_2) dx_2 \forall t \right\} \right) \\
= kT \ln \left( M \left( \frac{G(p)}{1 - \varepsilon} \right) \right).
\]

This proves the first part of the main theorem. 

(\( \Box \))
Appendix B: \(W^e\) is achievable

This section concerns the second statement of the main theorem (theorem 1). We specify a protocol that achieves the bound given in theorem 1, i.e., it extracts \(W^e\) of work with a failure probability no greater than \(\varepsilon\). The protocol is within the rules of the game (defined in section 3). The protocol works for the initial \((\rho)\) and final \((\sigma)\) states taking the form \(\rho = \ldots \otimes |\xi\rangle \langle \xi|\) and \(\sigma = \ldots \otimes |\xi\rangle \langle \xi|\), where \(|\xi\rangle\) is one of the energy eigenstates of a system with two energy eigenstates in total. This is a small restriction. It amounts to allowing the agent an extra two-level system in a known state, working as a catalyst in the sense that it aids the process but is ultimately unchanged by it.

Before giving the general protocol it is instructive to consider an example. We begin with a state \(\phi\) with energy-eigenvalues \(\lambda_i(j)\) and energy eigenvalues \(E_i(j)\) and \(A_i\) defined by \(A_i(j) = \exp\left(-\frac{E_i(j)}{kT}\right)\). These are given by:

\[
\lambda_i = \left(\frac{2}{3}, \frac{1}{3}, 0\right) \\
A_i = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)
\]

and therefore:

\[
p_i(x) = \begin{cases} 
2, & x \in (0, \frac{1}{3}) \\
1, & x \in \left(\frac{1}{3}, \frac{2}{3}\right) \\
0, & x \in \left(\frac{2}{3}, 1\right)
\end{cases} \quad (B1)
\]

The final state we want to reach is defined through:

\[
\lambda_f = \left(\frac{1}{2}, \frac{1}{2}, 0\right) \\
A_f = \left(\frac{1}{2}, \frac{1}{2}, 0\right)
\]

and therefore:

\[
p_f(x) = \begin{cases} 
3, & x \in (0, \frac{1}{2}) \\
\frac{3}{2}, & x \in \left(\frac{1}{2}, \frac{3}{2}\right)
\end{cases} \quad (B2)
\]

With a risk \(\varepsilon = \frac{1}{2}\), the work for this game is limited by \(W = kT \ln \left(M \left(\frac{p_i}{1-\varepsilon} \parallel p_f\right)\right) = kT \ln \left(\frac{4}{3}\right)\). In this example we show how this amount of work can be extracted.

We first want to raise as many levels as we can to infinite energy, such that if we succeed, we start with a more known state. Unfortunately the sum of the occupation probabilities of the lowest levels will never yield exactly \(\varepsilon\), so we need to change this first.

We start by raising the empty level to infinite energy, such that even if one mixes it completely with any other level it will stay empty. Then we lower the energy of the empty level, while constantly mixing this level with the first one. At the same time we enhance the energy of the first level, such that in total the energy of the work reservoir is unchanged with probability 1 (the details of this action can be found below in definition 4 and the following lemma). We then have:

\[
\lambda_1 = \left(\frac{1}{2}, \frac{1}{2}, \frac{1}{6}\right) \\
A_1 = \left(\frac{1}{4}, \frac{1}{3}, \frac{1}{3}\right)
\]

\[
p_1(x) = \begin{cases} 
2, & x \in \left(0, \frac{1}{3}\right) \\
1, & x \in \left(\frac{1}{3}, \frac{2}{3}\right)
\end{cases}
\]

The lowest two eigenvalues now sum up to \(\varepsilon\). We enhance the energy of these two levels by doing a work extraction changing the energy of their states by \(\infty\). With probability \(1 - \varepsilon = \frac{1}{2}\) we get the work 0 and the state:

\[
\lambda_2 = (1, 0, 0) \\
A_2 = \left(\frac{1}{4}, 0, 0\right)
\]

\[
p_2(x) = 4, x \in (0, \frac{1}{2})
\]

which in this case is a pure state (the state would not have been pure if we had chosen \(\varepsilon\) to be smaller than \(\frac{1}{3}\)). With probability \(\frac{1}{2}\) we get the work –\(\infty\), so the above is the only step to be considered.

Now we extract the work \(W = kT \ln \left(\frac{4}{3}\right)\) on all the levels. This succeeds with probability 1. The state afterwards is given by:

\[
\lambda_3 = (1, 0, 0) \\
A_3 = \left(\frac{1}{3}, 0, 0\right)
\]

\[
p_3(x) = 3, x \in (0, \frac{1}{3})
\]

Again we need two levels where we only have one. Acting again as defined in definition 4 on the first two levels we can get:

\[
\lambda_4 = \left(\frac{1}{2}, \frac{1}{2}, 0\right) \\
A_4 = \left(\frac{1}{2}, \frac{1}{6}, 0\right)
\]

\[
p_4(x) = 3, x \in (0, \frac{1}{3})
\]

The energy of the second level is now too high and we need to lower it by \(kT \ln(2)\):

\[
\lambda_5 = \begin{cases} 
(1, 0, 0), & \text{with probability } \frac{1}{2} \\
(0, 1, 0), & \text{with probability } \frac{1}{2}
\end{cases} \\
A_5 = \left(\frac{1}{6}, \frac{1}{3}, 0\right)
\]
The work extracted in this step is in both cases at least 0. So by measuring whether the energy in the work-reservoir has been enhanced by at least \( W = kT \ln \left( \frac{4}{3} \right) \), we get a "yes" and the wanted final state with probability \( \frac{1}{2} \).

As seen in the above example, we need some algorithm which allows us to shift some probability from one level to the other, if they are in thermal equilibrium. We only want to change these two levels (say \( j \) and \( k \)) by isothermal shift of the boundary between them, which allows us to shift some probability from one level to another.

Definition 8 (Isothermal shift of boundary). Let \( A(j) = \exp \left( -\frac{E_j}{k T} \right) \), where \( E_j \) is the energy eigenvalue of the \( j \)th level.

Let the levels \( j, k = j + 1 \) have the same Gibbs rescaled probability.

We call the limit \( n \to \infty \) of the following process an isothermal shift of the boundary between \( j \) and \( k \) by \( w \in \left( -\frac{A(j)}{A(j) + A(k)}, \frac{A(k)}{A(j) + A(k)} \right) \) in direction \( k \):

1. Do a permutation, which brings the level \( j \) in front and level \( k \) as second.

2. Do a work extraction on \( (0, A(j)] \) by:
   \[
   w_1 = 1 - \frac{w A(j) + A(k)}{n A(j)}
   \]

3. Do a permutation, which brings the level \( k \) in front and level \( j \) second.

4. Do a work extraction on \( (0, A(k)] \) by:
   \[
   w_2 = 1 - \frac{w A(j) + A(k)}{n A(k)}
   \]

5. Do a thermalization totally mixing the two levels \( j \) and \( k \) and letting all others untouched (i.e. the matrix with entries \( 1/2 \) in \( (1, 1), (1, 2), (2, 1) \) and \( (2, 2) \) and \( \delta_{n,1} \) everywhere else, such that the first entry of the vector it is applied on, is the probability of the level \( j \) after work extraction and the second is the probability of the level \( k \)).

6. Restart with 1. \( n \) times in total, redefining \( A(j) \) and \( A(k) \) as above for the probabilities after this process.

7. Do a permutation, which brings back the levels \( j \) and \( k = j + 1 \) at their position at the beginning (we show below, that this is possible).

Instead of the first four actions, we could have simply said we do extract the work \( w_1 \) on the level \( j \) and the work \( w_2 \) on the level \( k \). Then we would have had to continue with doing the total mixing also between these levels (instead of at the first and second position of the matrix) and so on. What we mean here with doing a work extraction on the level \( j \) is the action: do a permutation bringing the block \( (a, b) \) corresponding to the level \( j \) in front, extract work, permute the level back. This yields the same result, as if we just defined the work extraction generally on any blocks \( (a, b) \) (i.e. squeezing of the block by \( w \), keeping the corresponding area constant), instead of only on blocks \( (0, c) \) (the proof is trivial). In later definitions we will make use of this. Here we do not, since the algebra would get slightly more complicated.

The following Lemma shows that the above process costs no work with probability 1 and that it can indeed be seen as a shift of the separation between the levels.

Lemma 6 (Action of the isothermal shift of boundary). Let \( A(j) = \exp \left( -\frac{E(j)}{k T} \right) \), where \( E(j) \) is the energy eigenvalue of the \( j \)th level.

Let the levels \( j, k = j + 1 \) have the same Gibbs rescaled probability.

After an isothermal shift of the boundary between \( j \) and \( k \) by \( w \in \left( -\frac{A(j)}{A(j) + A(k)}, \frac{A(k)}{A(j) + A(k)} \right) \) in direction \( k \):

1. \( (a) \) the energy eigenvalues of all levels but \( j \) and \( k \) remain constant.

   \( (b) \) At the end \( A_f(j) = \exp \left( -\frac{E_f(j)}{k T} \right) \) is given by \( A_f(j) = A(j) + w(A(j) + A(k)) \) and for the level \( k \): \( A_f(k) = A(k) - w(A(j) + A(k)) \) (\( E_f(j) \) is the energy of the eigenvalue \( j \) after the shift).

2. with probability \( 1 - (\lambda(j) + \lambda(k)) \), the eigenvalues of the final state are given by
   \[
   \lambda(l) = \frac{\lambda(l)}{1 - (\lambda(j) + \lambda(k))} \quad \text{for} \quad l \neq j, k
   \]
   and 0 for \( l = j, k \).
3. With probability $\lambda(j) + \lambda(k)$, the eigenvalues of the final state are given by $A_j^k(l)$ for $l = j, k$ and 0 else.

4. With probability 1 the energy in the work reservoir is changed by $W = 0$.

Proof. 1.(a) just follows out of the algorithm, since we did not do any work extraction on any levels and this is the only way we can change energies in our game. For 1.(b) we need to look at how the energy eigenvalues of the $j$th and $k$th level change each of the $n$ times one goes through the algorithm in definition \[\$\] directly from the algorithm we get, that in the first time one goes through it $A(j)$ changes to $A_1(j) = \exp\left(-E(j) + \frac{T_B\ln(w_1)}{k}\right)$ and we get $A_1(j) = w_1 A(j) = A(j) + \frac{w_1}{n}(A(j) + A(k))$ and by the same argument $A_1(k) = w_1 A(k) = A(k) + \frac{w_1}{n}(A(j) + A(k))$. Since $A_1(j) + A_1(k) = A(j) + A(k)$ we see, that after one times one goes through the algorithm, one ends up with: $A_1(j) = A(j) + \frac{(l-1)n}{w_1}(A(j) + A(k)) + \frac{w_1}{n}(A(j) + A(k)) = A(j) + l \frac{w_1}{n}(A(j) + A(k))$ and $A_1(k) = A(k) - \frac{w_1}{n}(A(j) + A(k))$. With $l = n$ we get what is stated in 1. (b).

In order to derive 2. and 3. we need to have a closer look at how the eigenvalue changes each of the $n$ times we go through the algorithm. The eigenvalues are given by the Gibbs rescaled probabilities multiplied with the corresponding $A(l)$.

Let $q$ be the Gibbs rescaled probability distribution after step 1. of the $l$th time one goes through the algorithm in definition \[\$\]. After step 2. we have:

$$ q(x) \Rightarrow \begin{cases} q(x)_{(A_k, A_j)}(x) \frac{\eta(q)}{w_1 \eta(q)} & , \text{with prob. } \eta(q) \\ q(x-A_1 w_1 A_j \eta(A_{j+1}, x)) \frac{\eta(q)}{1-\eta(q)} & , \text{with prob. } 1 - \eta(q) \\ \end{cases}$$

where $\eta(q) = \int q(x) d x$ and $Z(q)$ is the partition function of $q$.

After step 4. we thus have:

$$ \begin{align*} q(x) & = \frac{q(x)_{(A_k, A_j)}(x)}{w_2 \eta(q)} , \text{w. prob. } \eta(q) \\ q(x-A_1 w_1 A_k w_2 A_j \eta(A_{j+1}, x)) & = \frac{q(x)_{(A_k, A_j)}(x)}{w_2 \eta(q)} , \text{w. prob. } \eta(q) \\ q(x-A_1 w_1 A_k w_2 A_j \eta(A_{j+1}, x)) & = \frac{q(x-A_1 w_1 A_k w_2 A_j \eta(A_{j+1}, x))}{1-\eta(q)} , 1 - \eta(q) - \eta(q) \end{align*}$$

Noting that $q(x) = q(x/w_2)$ for $x \in [0, A_k]$ and similarly for $x \in \{A_j, A_k + A_j\}$ and $x - A_j w_1 A_k w_2 A_j + A_j = x$, we can rewrite this as:

$$ \begin{align*} q(x)_{(A_k, A_j)}(x) & = \frac{q(x)_{(A_k, A_j)}(x)}{w_2 \eta(q)} , \text{w. prob. } \eta(q) \\ q(x)_{(A_k, A_j)}(x) & = \frac{q(x)_{(A_k, A_j)}(x)}{w_2 \eta(q)} , \text{w. prob. } \eta(q) \\ q(x-A_1 w_1 A_k w_2 A_j \eta(A_{j+1}, x)) & = \frac{q(x-A_1 w_1 A_k w_2 A_j \eta(A_{j+1}, x))}{1-\eta(q)} , 1 - \eta(q) - \eta(q) \end{align*}$$

Which means that after step 5. we get:

$$ \begin{cases} q(x)_{(A_k, A_j)}(x) \frac{w_2 \eta(q)}{w_2 \eta(q)} , \text{w. prob. } \eta(q) \\ q(x)_{(A_k, A_j)}(x) \frac{w_2 \eta(q)}{w_2 \eta(q)} , \text{w. prob. } \eta(q) \\ q(x-A_1 w_1 A_k w_2 A_j \eta(A_{j+1}, x)) \frac{w_2 \eta(q)}{1-\eta(q)} , 1 - \eta(q) - \eta(q) \end{cases}$$

For 2. note that with probability $1 - (\lambda(j) + \lambda(k))$ we get after the first time one goes through the algorithm: $q_j = q_k = 0$ (which just means, that the state is measured to be orthogonal to $j$ and $k$). And therefore in the subsequent steps we have $\eta(q_j) = \eta(q_k) = 0$. So we get with probability $1 - (\lambda(j) + \lambda(k))$, the final probability distribution:

$$ p(x) = \frac{\Theta(A_1 + A_k, x)}{1 - (\lambda(j) + \lambda(k))} $$

Since the energy eigenvalues of these levels are unchanged, we get $\frac{\lambda(j) + \lambda(k)}{\lambda(j) + \lambda(k)}$ for $l \neq j, k$ and 0 for $l = j, k$ for the eigenvalues, which proves 2. The final Gibbs rescaled probabilities of the levels $j$ and $k$ have the same value (since we completely mix them in step 5.). Their integral ($\int q^j + q^k dx$) after the first time one goes through the algorithm keeps 1 (with probability $\lambda(j) + \lambda(k)$). As noticed before, $A(j) + A(k) = A(j) + A(k)$. Thus we get that with probability $\lambda(j) + \lambda(k)$ the eigenvalues of the levels are given by: $\frac{A_1^k(l)}{A_{(j+k)}}$ for $l = j, k$ and 0 else. Which proves 3.

Suppose in the first time one goes through the algorithm the state is orthogonal to the levels $j, k$: then the energy in the work reservoir is unchanged throughout the whole $n$ times one goes through the algorithm and for this case, 4. follows trivially.

We now look at the other case (the case where the state is projected onto the levels $j, k$ the first time one goes through the algorithm).

Let $s \in \{1, 2\}$. Define $\sigma(2) = 1$ and $\sigma(1) = -1$. Define $\alpha_1 \equiv \frac{\lambda(l)}{\lambda_j + \lambda(k)}$ and $\alpha_2 \equiv 1 - \alpha_1$.

In the $l$th time one goes through the algorithm one either gets the logarithmical work

$$ w_l(1) = 1 + \frac{w_j A(j) + A_k(k)}{n A(j)} $$

$$ = 1 + \frac{w_j A(j) + A_k(k)}{n A(j) + (l - 1) \frac{w_j A(j) + A_k(k)}{n}} $$

$$ = \frac{\alpha_1 + \frac{l w}{n}}{\alpha_1 + \frac{(l - 1) w}{n}}$$

or the similarly derivable value for $w_l(2)$ ($A_l$ is defined in the proof of 1.(b)). Thus we can write:

$$ w_l(s_l) = \frac{\alpha_{s_l} + \sigma(s_l) l \frac{w}{n}}{\alpha_{s_l} + \sigma(s_l) (l - 1) \frac{w}{n}}$$

In total we get the logarithmical work:

$$ w_{tot} = \prod_{l=1}^{n} \frac{\alpha_{s_l} + \sigma(s_l) l \frac{w}{n}}{\alpha_{s_l} + \sigma(s_l) (l - 1) \frac{w}{n}}$$
with probability (given, that we have the case where the state is projected onto the levels \(j, k\) the first time one goes through the algorithm):

\[
P(\vec{s}_j \land k) = \prod_{l=1}^{n} \left( \alpha_{s_l} + \sigma(s_l)(l-1)\frac{w}{n} \right)
\]

The expectation value of \(w_{tot}\) can be computed as follows (for \(n < \infty\)):

\[
E(w_{tot}) = \sum_{\vec{s}} P(\vec{s}_j \land k) w_{tot}(\vec{s})
\]

\[
= \sum_{\vec{s}} \prod_{l=1}^{n} \left( \alpha_{s_l} + \sigma(s_l)\frac{w}{n} \right)
\]

\[
= \prod_{l=1}^{n} \left( \sum_{\vec{s}} \alpha_{s_l} + \sigma(s_l)\frac{w}{n} \right)
\]

\[
= \prod_{l=1}^{n} \left( \alpha_1 + \alpha_2 + (\sigma(l) + \sigma(2))\frac{w}{n} \right)
\]

\[
= 1
\]

We now look at how much the work \(W = \ln(w_{tot})\) changes, if in step \(l\) one replaces \(s_l\) by \(\hat{s}_l\) (remember that \(\hat{s}_j = 1\) and vice versa):

\[
W(s_1, \ldots, s_n) - W(s_1, \ldots, \hat{s}_l, \ldots, s_n)
\]

\[
= \ln \left( \frac{\alpha_{s_l} + \sigma(s_l)\frac{w}{n}}{\alpha_{\hat{s}_l} + \sigma(\hat{s}_l)(l-1)\frac{w}{n}} \right) - \ln \left( \frac{\alpha_{s_l} + \sigma(s_l)\frac{w}{n}}{\alpha_{\hat{s}_l} + \sigma(\hat{s}_l)(l-1)\frac{w}{n}} \right)
\]

with \(c = w\sigma(s_l)\) (and therefore \(w\sigma(\hat{s}_l) = -c\)), \(a = \alpha_{s_l}\) (and \(\alpha_{\hat{s}_l} = 1-a\)), \(x = a + c\frac{w}{n}\) and \(y = 1 - a - c\frac{w}{n}\) we get:

\[
|W(s_1, \ldots, s_n) - W(s_1, \ldots, \hat{s}_l, \ldots, s_n)|
\]

\[
= \ln \left( \frac{x (y + \frac{w}{n})}{(x - \frac{w}{n}) y} \right)
\]

\[
= \ln \left( 1 + \frac{z}{n} \right)
\]

\[
\leq \frac{z}{n} =: q_l.
\]

Using the McDiarmid inequality we get that the probability that \(W\) differs from its expectation value is bounded by:

\[
P(|W(\vec{s}) - E(W)| \geq \delta) \leq 2 \exp \left( -\frac{2\delta^2}{\sum_{l} q_l^2} \right) \leq 2 \exp \left( -\frac{2\delta^2}{\frac{1}{n} |\vec{s}|^2} \right)
\]

which tends to 0 for any \(\delta > 0\). Therefore we get that the work in this process is given by 0 with probability 1, which proves 4.

Next we need an algorithm which makes it possible to get the end state \(\sigma\) out of the initial state \(\rho\), if \(p \succ p_f\) (the generalization of the step 4 \(\rightarrow 5\) in the example). To write down the algorithm, we first need a definition and a technical lemma simplifying the notation:

**Definition 9** (Generalized sum). If \(c \in \mathbb{R}, c \geq 1\), we define \(\sum_{i=l}^c d_i := \sum_{i=1}^{|c|} d_i + (c - |c|)d_{|c|}\). If \(c \in \mathbb{R}, c < 1\), we define \(\sum_{i=l}^c d_i := 0\).

(Note that the above definition reduces to the usual sum if \(c \in \mathbb{N}\).

**Lemma 7.** Let \(p\) and \(p_f\) be Gibbs rescaled probability distributions of two states \(\rho\) and \(\sigma\), with the same number of levels \(n\).

Let \(Z = Z_f\) where \(Z\) is the partition function of \(\rho\) and likewise for \(Z_f\).

Let \(p \succ p_f\). Then:

\[\exists \theta = a_0 \leq \ldots \leq a_n \leq n\]

\[0 = b_0 \leq \ldots \leq b_n \leq n\]

\[\text{with } b_k \leq a_k \forall k\]

\[\text{s.t. :} \]

\[\sum_{i=1}^{a_k} \lambda(i) - \sum_{i=1}^{k-1} \lambda_f(i) = \frac{\lambda_f(k)}{A_f(k)} \]

\[\sum_{i=1}^{a_k} A(i) - \sum_{i=1}^{k-1} A_f(i) = \frac{\lambda_f(k)}{A_f(k)}\]

and:

\[\sum_{i=1}^{b_k} \lambda(i) = \sum_{i=1}^{k} \lambda_f(k)\]

**Proof.** Let \(A(j) = \exp \left( -\frac{E(j)}{kT} \right)\), where \(E(j)\) is the energy eigenvalue of the level \(j\) of the state \(\rho\) and likewise for \(A_f(j)\) for the end state \(\sigma\).

We first prove that such \(a_k\) exist:

For \(k = 1\) we have:

\[\frac{\lambda(1)}{A(1)} \geq \frac{\lambda_f(1)}{A_f(1)} \geq \sum_{i=1}^{n} \frac{\lambda_f(i)}{A_f(i)} = \sum_{i=1}^{n} \frac{\lambda(i)}{A(i)}\]

Since \(\frac{\lambda(i)}{A(i)}\) is continuous in \(a_1\) it follows that there is an \(a_1\) for which:

\[\sum_{i=1}^{a_1} \frac{\lambda(i)}{A(i)} = \frac{\lambda_f(1)}{A_f(1)}\]

For \(k > 1:\)

\[\frac{\sum_{i=1}^{a_{k-1}} \lambda(i) - \sum_{i=1}^{k-1} \lambda_f(i)}{\sum_{i=1}^{a_{k-1}} A(i) - \sum_{i=1}^{k-1} A_f(i)} = \frac{\lambda_f(k-1)}{A_f(k-1)} \geq \frac{\lambda_f(k)}{A_f(k)}\]
Let $\rho$ and $\sigma$ be Gibbs rescaled probability distributions of two states $\rho$ and $\sigma$, with the same number of levels $2n + 1$.

Let $\rho$, $\sigma$ have at least $n + 1$ levels with eigenvalues $e(i) = \lambda(i + n) = 0$ and energy eigenvalues $E = \infty$.

Let $p \succ p_f$.

We call the following algorithm an assimilation of $\rho$ to $\sigma$:

1. Do a work extraction on the $n + 1$'th empty level s.t $Z = Z_f$.

2. Choose $a_k$ and $b_k$ for the $2n + 1$ levels as in the above lemma.

3. For $k = 1$ to $n$:
   
   (a) Do an isothermal shift on the level $e_k$ in direction $[a_k]$ by $a_k - [a_k]$.
   
   (b) Totally mix $e_k$ with $[a_k]$ and do an isothermal shift on the level $[a_k]$ in direction $e_k$ by 1.
   
   (c) Totally mix $[b_{k-1}], \ldots, [b_k]$.
   
   (d) Do an isothermal shift on the level $e_k$ in direction $[b_k]$ by $b_k - [b_k]$.
   
   (e) Do an isothermal shift on the level $e_k$ in direction $[b_{k-1}, \ldots, [b_k]$ by 1.

![Figure 3: Assimilation](image)

**Definition 10** (Assimilation of $\rho$ to $\sigma$). Let $p$ and $p_f$ be Gibbs rescaled probability distributions of two states $\rho$ and $\sigma$, with the same number of levels $2n + 1$.

Let $\rho$, $\sigma$ have at least $n + 1$ levels with eigenvalues $e(i) = \lambda(i + n) = 0$ and energy eigenvalues $E = \infty$.

Let $p \succ p_f$.

We call the following algorithm an assimilation of $\rho$ to $\sigma$:

1. Do a work extraction on the $n + 1$'th empty level s.t $Z = Z_f$.

2. Choose $a_k$ and $b_k$ for the $2n + 1$ levels as in the above lemma.

3. For $k = 1$ to $n$:
   
   (a) Do an isothermal shift on the level $e_k$ in direction $[a_k]$ by $a_k - [a_k]$.
   
   (b) Totally mix $e_k$ with $[a_k]$ and do an isothermal shift on the level $[a_k]$ in direction $e_k$ by 1.
   
   (c) Totally mix $[b_{k-1}], \ldots, [b_k]$.
   
   (d) Do an isothermal shift on the level $e_k$ in direction $[b_k]$ by $b_k - [b_k]$.

4. After step 3.(d) of the $k$-1 time one goes through the algorithm the eigenvalue of the level $\lambda_{k-1,d}(b_{k-1})$ (the index stands for the $k - 1, d$ time one goes through the

**Lemma 8** (Action of the assimilation $\rho$ to $\sigma$). Let $p$ and $p_f$ be Gibbs rescaled probability distributions of two states $\rho$ and $\sigma$, with the same number of levels $2n + 1$.

Let $\rho$, $\sigma$ have at least $n + 1$ levels with eigenvalues $\lambda_e = 0$ and energy eigenvalues $E_e = \infty$.

Let $p \succ p_f$.

After an assimilation of $\rho$ to $\sigma$ the final state is given by $\sigma$ with probability 1. The extracted work is 0.

**Proof.** Note that $z < Z_f$ since $p \succ p_f$. Therefore one can lower the energy of one empty level, such that we get $Z = Z_f$. This costs no work and we still have $p \succ p_f$ afterwards. Thus after step (1) we can apply Lemma 7.

After step 3.(d) of the $k - 1$ time one goes through the algorithm the eigenvalue of the level $\lambda_{k-1,d}(b_{k-1})$ (the index stands for the $k - 1, d$ time one goes through the
algorithm after step d) is given by:

\[ \lambda_{k-1,d}(b_{k-1}) = \lambda_{k-1,c}(b_{k-1})(1 - (b_{k-1} - b_{k-1})) \]
\[ = \lambda_{k-1,c}(b_{k-1})(b_{k-1} - b_{k-1}) \]

and likewise \( A_{k-1,d}(b_{k-1}) = A_{k-1,c}(b_{k-1})(b_{k-1} - b_{k-1}) \); (see lemma 6). After steps 3.(a),(b):

\[ \lambda_{k-1,b}(a_k) = \lambda_{k-1,c}(a_k) + \lambda_{k-1,c}(a_k)(a_k - a_k) \]

and likewise for \( A_{k-1,b}(a_k) \).

Assume that \( \sum_{k=1}^{n} k_{i+1}(i + n) = \sum_{i=1}^{k} A_f(i) (i + n) \) is the first part of the proof of lemma 6). After step 3.(c) we get for any level \( j \)

\[ \sum_{i=1}^{n} \lambda(i) = 1. \]

After steps 3.(d),(e) we therefore have, that \( e, k, e \) we end up with levels \( e, k, e \) with \( A, n, e \) as required.

We can now define an algorithm which is more physical. We assume here that we have at least \( n/2 \) levels, with 0 probability, but make sure that in the end these levels have again 0 probability (note, that this does not change the upper bound for the work). We assume that the levels are ordered in descending order of their Gibbs rescaled probability.

**Definition 11 (Work extraction algorithm).** Let \( p \) and \( p_f \) be Gibbs rescaled probability distributions of two states \( \rho \) and \( \sigma \), with the same number of levels \( 2n + 1 \). Let \( \rho, \sigma \) have at least \( n + 1 \) levels with eigenvalues \( \lambda_e = 0 \) and energy eigenvalues \( E_e = \infty \). Define \( W = kT \ln(M^c(p,q)) \).

1. First we lift all the empty levels to infinite energy, which costs no work.

2. If there is no \( k \) for which \( 1 - \epsilon = \sum_{i=1}^{k} \lambda(i) \),
   Take an empty level \( e \) (the level \( n + 1 \)) and a level \( k \) for which \( \sum_{i=1}^{k} \lambda(i) < 1 - \epsilon < \sum_{i=1}^{k} \lambda(i) \).
   Make an isothermal shift of the boundary between \( e \) and \( k \) by \( 1 - \frac{1 - \epsilon - \sum_{i=1}^{k} \lambda(i)}{\lambda(k)} \) in direction \( k \).

3. Do a work extraction on the levels \( k + 1, \ldots, n, e \) by \( -\infty \).

4. Make a work extraction on all levels by \( W \).

5. Make an assimilation of the obtained state to the final state.

6. Permute the levels of the obtained state such that one gets the final state.

**Theorem 9** (Bound can be achieved (second part of main theorem)). Let \( \rho \) and \( p_f \) be Gibbs rescaled probability distributions of two states \( \rho \) and \( \sigma \), with the same
number of levels $2n + 1$.
Let $\rho, \sigma$ have at least $n + 1$ levels with eigenvalues $\lambda_e = 0$ and energy eigenvalues $E_e = \infty$.
Define $W = kT \ln(M^e(p, q))$. The work extraction algorithm on $\rho$ yields the work $W$ with probability $1 - \varepsilon$. If the work extraction is successful, the final state is given by $\sigma$ with probability 1.

Proof. After step 2, the sum of the eigenvalues $\sum_{i=k+1}^{n+1} \lambda(i) = \varepsilon$ (see lemma 6). Therefore the work extraction in step 3 succeeds with probability $\varepsilon$ and if it does not succeed it yields 0 work (else $-\infty$).
After step 3, the eigenvalues are given by $\lambda_3(i) = \lambda(i) = \frac{\lambda(i)}{1 - \varepsilon}$ for $i = 1, \ldots, k$ and $\lambda_3(i) = 0$ else (if the work extraction succeeds).
After step 4, by the definition of $W$ we have that $p_4(i) \succ p_f(i)$, the extracted work is $W$. Therefore one can make an assimilation of the obtained state to the final state and one gets the final state $\rho$ (up to permutation) with probability 1 (see lemma 8). After the permutation (if the levels have some special physical meaning) we get the final state $\rho$ with probability 1.
In total we get the final state $\rho$ with probability 1, if the work extraction succeeds and the extracted work is $W$ with probability $1 - \varepsilon$. \qed
Appendix C: Recovering the relative min-entropy

We now recover the result of eq. 2 which, as discussed in the main body, was given in [11, 12]. For completeness we repeat the equation also here:

$$W^\varepsilon = kT \ln(2) D_0^\varepsilon(\sigma||\rho_T).$$

To do this we first need the corresponding expression for the Gibbs rescaled entropies:

In the case where one starts at any state $P$ on a system with only degenerate energy levels and ends up in the thermal state $f = (2^{-n}, \ldots, 2^{-n})$, we would expect the maximal work that can be extracted by taking a risk $\varepsilon$ to be $kT \ln(2) (n - H_{\max}^\varepsilon(P))$, as seen in [9] for the case of Szilard engines. Indeed this is the result. More generally:

**Theorem 10.**

$$W^\varepsilon = kT \ln(2) (H_{\max}(q) - H_{\max}^\varepsilon(p))$$

where $p = G^T(\rho)$ is the Gibbs rescaled probability distribution corresponding to the initial state $\rho$ and $q = G^T(\sigma)$ is the one corresponding to the final state $\sigma$.

For the proof of this theorem a technical lemma on the smooth max-entropy is needed.

**Lemma 11.** Let $p$ be a monotonously falling probability function on $[0, \infty)$ and $d_\varepsilon$ be defined through

$$\int_0^{d_\varepsilon} p(x)/(1 - \varepsilon) \, dx = 1$$

Then:

$$d_\varepsilon = 2^{H_{\max}^\varepsilon(p)}$$

**Proof.** Let $d_\varepsilon$ be defined as above. We need to show two things:

i) $\exists \tilde{p}^\varepsilon$ probability function on $[0, \infty)$ with $\|\supp(\tilde{p}^\varepsilon)\| = d_\varepsilon$ and trace-distance $\delta(p, \tilde{p}^\varepsilon) < \varepsilon$.

ii) $\|\supp(\tilde{p}^\varepsilon)\| \geq d_\varepsilon \forall \tilde{p}^\varepsilon$ monotonously decreasing probability functions on $[0, \infty)$ with $\delta(p, \tilde{p}^\varepsilon) < \varepsilon$.

Then we get that $H_{\max}^\varepsilon(p) = \log_2(\min_{\delta(p,\tilde{p}^\varepsilon)<\varepsilon}(\|\supp(\tilde{p}^\varepsilon)\|)) = \log_2(d_\varepsilon)$, as said in the lemma. The proof of i) goes as follows: Define $\tilde{p}^\varepsilon(x) = p(x) \left(\int_0^{d_\varepsilon} p(x) \, dx\right)^{-1}$ for $x \leq d_\varepsilon$ and $\tilde{p}^\varepsilon(x) = 0$ for $x > d_\varepsilon$. This $\tilde{p}^\varepsilon$ is therefore normalized to one, has support $[0, d_\varepsilon]$ and the following equation shows that it is also $\varepsilon$-near to $p$:

$$\delta(p, \tilde{p}^\varepsilon) = \frac{1}{2} \left( \int_0^\infty |p^\varepsilon(x) - p(x)| \, dx \right)$$

and

$$\delta(p, \tilde{p}^\varepsilon) = \frac{1}{2} \left( \int_0^{d_\varepsilon} |p^\varepsilon(x) - p(x)| \, dx + \int_{d_\varepsilon}^\infty p(x) \, dx \right)$$

which concludes the proof of i), ii) is proven on the next page (for typographical reasons).
For the proof of ii) assume, that: \( \exists \rho^\varepsilon \) like above, s.t. \( |\text{supp}(\rho^\varepsilon)| \leq d_\varepsilon \), then (with \( \Theta = \Theta_{0,\infty} \)):

\[
\frac{1}{2} \left( \int_0^\infty |p^\varepsilon(x) - p(x)| \, dx \right)
\]

\[
= \frac{1}{2} \int_0^{d_\varepsilon} |p^\varepsilon(x) - p(x)| \, dx + \int_0^\infty |p^\varepsilon(x) - p(x)| \, dx
\]

\[
\geq \frac{1}{2} \int_0^{d_\varepsilon} \varepsilon + \int_0^\infty \Theta(p(x) - p^\varepsilon(x))(p(x) - p^\varepsilon(x)) \, dx
\]

\[
+ \frac{1}{2} \int_0^{d_\varepsilon} \Theta(p^\varepsilon(x) - p(x))(p^\varepsilon(x) - p(x)) \, dx
\]

\[
\geq \frac{1}{2} \varepsilon + \int_0^{d_\varepsilon} \Theta(p(x) - p^\varepsilon(x))(p(x) - p^\varepsilon(x)) \, dx
\]

\[
\geq \varepsilon
\]

which is a contradiction to \( \delta(p, p^\varepsilon) \)

\[
= \frac{1}{2} \int_0^\infty |p^\varepsilon(x) - p(x)| < \varepsilon.
\]

Now we have all we need to proof the theorem above:

**Proof.** let \( p^\varepsilon \) be a probability function with the smallest possible support such that \( \delta(p, p^\varepsilon) \leq \varepsilon \) and define \( d_\varepsilon \) as

in lemma \([1]\) For \( l \leq d_\varepsilon \) the requirement for maximal work extraction reads (using the lemma)

\[
\int_0^l \frac{p(x)}{1-\varepsilon} \, dx \geq \int_0^{d_\varepsilon} \frac{p(x)}{1-\varepsilon} \, dx = \frac{l}{d_\varepsilon} \int_0^l \frac{\|\text{supp}(q)\|}{\|\text{supp}(p^\varepsilon)\|} \, dx
\]

\[
= \int_0^l q(x) \, dx
\]

The above is an equation in the case \( l = d_\varepsilon \). Which shows that the maximal \( w \) as defined in theorem \([1]\) is given by

\[w = \frac{\|\text{supp}(q)\|}{\|\text{supp}(p^\varepsilon)\|} = 2(H_{\text{max}}(q) - H_{\text{max}}(p))\]

Eq. [2] is a special case of the above theorem, recovered when the final state is a Gibbs state and has also the same energy eigenvalues as the initial.

**Corollary.** Let \( \rho \) be a diagonal state with energy eigenvalues \( E_i \) and \( \sigma^T \) be the Gibbs state with the same energy eigenvalues \( E_i \) at the bath temperature \( T \). Then the maximal extractable work at risk \( \varepsilon \) is given by:

\[
W_\varepsilon = kT \ln(2) D^{\varepsilon}_0(\rho, \sigma^T)
\]

**Proof.** Let \( p \) be the Gibbs-rescaled probability function corresponding to \( \rho \) and \( \rho_j \) the eigenvalues of \( \rho \). Let \( a \) be the flat energy probability function corresponding to \( \sigma^T \). Let \( A(\varepsilon) = \frac{\exp(\varepsilon)}{Z} \), where \( E(\varepsilon) \) are the energy eigenvalues of \( \rho \) and \( \sigma^T \) and \( Z \) is the corresponding partition function. This means by definition, that

\[
p \left( Z \int_0^x A \left( \left\lfloor \frac{y \cdot n}{A} \right\rfloor \right) \, dy \right) = \frac{P \left( \left\lfloor \frac{x n}{A} \right\rfloor \right)}{A \left( \left\lfloor \frac{x n}{A} \right\rfloor \right) Z}
\]

and likewise \( a(x) = 1/\varepsilon \) (both defined for \( x \in [0, Z] \)).

From the above theorem we get:

\[
W_\varepsilon = kT \ln(2) (H_{\text{max}}(a) - H_{\text{max}}(p))
\]

\[
= kT \ln(2) \left( \log_2(Z) - \log_2 \left( \inf_{\delta(p^\varepsilon, p) < \varepsilon} \text{supp}(p) \right) \right)
\]

\[
= -kT \ln(2) \log_2 \left( \frac{1}{Z} \right)
\]

\[
\cdot \min \left\{ x \mid \int P(\lfloor y \cdot n \rfloor) \, dy > 1 - \varepsilon \right\}
\]

\[
= kT \ln(2) R^{\varepsilon}_0(P, A)
\]
Appendix D: Triangle inequality

The logarithmic relative mixedness respects a triangle inequality:

Lemma 12 (Triangle inequality). Let \( \rho, \sigma \) be states and \( \varepsilon_{1,2} \in [0,1) \)
Let \( m_1 = M \left( \frac{G^T(\rho)}{\varepsilon_1} \bigg\| G^T(\tau) \right) \)
and \( m_2 = M \left( \frac{G^T(\tau)}{\varepsilon_2} \bigg\| G^T(\sigma) \right) \).

\[
M \left( \frac{G^T(\rho)}{\varepsilon_1 + \varepsilon_2} \bigg\| G^T(\sigma) \right) \geq m_1 m_2
\]

For all states \( \tau \).

Proof. Let \( \rho, \tau \) and \( \sigma \) be states and \( \varepsilon_{1,2} \in [0,1) \). Let \( m_1 = M \left( \frac{G^T(\rho)}{\varepsilon_1} \bigg\| G^T(\tau) \right) \) and \( m_2 = M \left( \frac{G^T(\tau)}{\varepsilon_2} \bigg\| G^T(\sigma) \right) \). Let \( p = G^T(\rho) \), \( q = G^T(\sigma) \) and \( s = G^T(\tau) \).

\[
\int_{0}^{1} q(x) \, dx \leq \int_{0}^{1} \frac{s(x)}{1 - \varepsilon_2} \, dx
\]
\[
\leq \int_{0}^{1} \frac{p(x)}{(1 - \varepsilon_2)(1 - \varepsilon_1)} \, dx
\]
\[
\leq \int_{0}^{1} \frac{p(x)}{1 - \varepsilon_2 - \varepsilon_1} \, dx
\]

Therefore there is a \( m \geq m_1 m_2 \) such that

\[
\int_{0}^{1} \frac{p(x)}{1 - \varepsilon_2 - \varepsilon_1} \, dx \geq \int_{0}^{1} q(x) \, dx.
\]

It follows:

\[
M \left( \frac{G^T(\rho)}{\varepsilon_1 + \varepsilon_2} \bigg\| G^T(\sigma) \right) \geq m \geq m_1 m_2.
\]

\( \square \)
Appendix E: Entropy increase law

Consider the interaction of the working medium system with the heat bath. Let $S$ be the Von Neumann entropy of the system, $\beta$ the inverse temperature associated with the bath, and $\langle E \rangle = \sum_i \lambda_i E_i$ the expected internal energy of the system. This section compares the standard law for entropy increase:

$$\Delta S \geq \beta \Delta \langle E \rangle,$$

(E1)

with the one we propose should replace it:

$$W^0(\rho \to \rho') \geq 0.$$

(E2)

1. Our model respects standard expression

Lemma 13. In the model for thermalisation used here Eq. (E7) is always respected.

**Proof.** We firstly recall the model and define certain notation.

Recall that the thermalisation model states that when two levels, 1 and 2, are coupled to the heat bath, their ratio $\lambda_1/\lambda_2$ gets closer to $\exp(-\beta(E_1 - E_2))$, and the other $\lambda$'s are untouched. In our model one may concatenate several such interactions to implement any allowed multi-level interaction with the bath. It will therefore suffice to show that Eq. (E1) holds for a single two-level interaction with the heat bath.

For notational convenience let the probability of being in level 1 or 2 be called $\lambda_{12} := \lambda_1 + \lambda_2$. This is then constant for the given two-level interaction with the bath. In the extreme case of the two levels interacting with the bath for an arbitrary amount of time we have $\lambda_1 := \lambda_1^T$ and $\lambda_2 := \lambda_2^T$ ($T$ reminds us of the temperature dependence). These values must then obey the relation

$$\lambda_1^T/\lambda_2^T = \exp(-\beta(E_1 - E_2))$$

(E3)

We also assume without loss of generality that $E_2 \leq E_1$. This implies that $\lambda_1^T \leq 0.5 \lambda_{12}$.

Now we begin to prove the statement. Firstly we simplify $\delta S$ by noting that only two levels change their probabilities. We write

$$S = -\sum_i \lambda_i \log \lambda_i$$

$$= -\lambda_1 \log \lambda_1 - (\lambda_{12} - \lambda_1) \log(\lambda_{12} - \lambda_1) - \sum_{i=3}^{i_{\max}} \lambda_i \log \lambda_i$$

$$\equiv S_{12} - \sum_{i=3}^{i_{\max}} \lambda_i \log \lambda_i.$$

We see that in any two-level interaction

$$\Delta S = \Delta S_{12}.$$

(E4)

It is helpful to re-express $S_{12}$ in terms of an actual entropy $S_{12}$, so that we can use known properties of entropies to make statements about $S_{12}$. We let $\overline{\lambda}_1 := \lambda_1/\lambda_{12}$ and $\overline{\lambda}_2 := \lambda_2/\lambda_{12}$ such that $\overline{\lambda}_1 + \overline{\lambda}_2 = 1$. We define

$$\overline{S}_{12} := -\overline{\lambda}_1 \log \overline{\lambda}_1 - \overline{\lambda}_2 \log \overline{\lambda}_2.$$

One can then see in a few lines of algebra that

$$S_{12} = \lambda_{12} \overline{S}_{12} - \lambda_{12} \log \lambda_{12}.$$

It follows that

$$\Delta S_{12} = \lambda_{12} \Delta \overline{S}_{12}.$$

(E5)

We accordingly now want to show that $\lambda_{12} \Delta \overline{S}_{12} \geq \beta \Delta \langle E \rangle$.

We can now use a well known property of the Shannon/von Neumann entropy: $\overline{S}_{12}$ is concave in $\lambda_{12} = \lambda_1/\lambda_{12}$. The function is accordingly upper bounded by any tangential line, as in Figure E1. Consider the tangential line at $\lambda_1 = \lambda_1^T$. At that point it follows from a few lines that

$$\frac{d}{d\lambda_1} S_{12|\lambda_1 = \lambda_1^T} = \frac{d}{d\overline{\lambda}_1} \overline{S}_{12} = \beta(E_1 - E_2).$$

(E6)

Note now that $\langle E \rangle$ may similarly to the entropy be written as

$$\langle E \rangle = -\sum_i \lambda_i E_i$$

$$\equiv \langle E \rangle_{12} + \langle E \rangle_{\text{rest}},$$

with the tangent upper bound having gradient $\beta(E_2 - E_1)$. It is helpful to re-express $S_{12}$ in terms of an actual entropy $S_{12}$, so that we can use known properties of entropies to make statements about $S_{12}$. We let $\overline{\lambda}_1 := \lambda_1/\lambda_{12}$ and $\overline{\lambda}_2 := \lambda_2/\lambda_{12}$ such that $\overline{\lambda}_1 + \overline{\lambda}_2 = 1$. We define

$$\overline{S}_{12} := -\overline{\lambda}_1 \log \overline{\lambda}_1 - \overline{\lambda}_2 \log \overline{\lambda}_2.$$

One can then see in a few lines of algebra that

$$S_{12} = \lambda_{12} \overline{S}_{12} - \lambda_{12} \log \lambda_{12}.$$

It follows that

$$\Delta S_{12} = \lambda_{12} \Delta \overline{S}_{12}.$$

(E5)

We accordingly now want to show that $\lambda_{12} \Delta \overline{S}_{12} \geq \beta \Delta \langle E \rangle$.

We can now use a well known property of the Shannon/von Neumann entropy: $\overline{S}_{12}$ is concave in $\lambda_{12} = \lambda_1/\lambda_{12}$. The function is accordingly upper bounded by any tangential line, as in Figure E1. Consider the tangential line at $\lambda_1 = \lambda_1^T$. At that point it follows from a few lines that

$$\frac{d}{d\lambda_1} S_{12|\lambda_1 = \lambda_1^T} = \frac{d}{d\overline{\lambda}_1} \overline{S}_{12} = \beta(E_1 - E_2).$$

(E6)

Note now that $\langle E \rangle$ may similarly to the entropy be written as

$$\langle E \rangle = -\sum_i \lambda_i E_i$$

$$\equiv \langle E \rangle_{12} + \langle E \rangle_{\text{rest}},$$

with the tangent upper bound having gradient $\beta(E_2 - E_1)$.
such that $\Delta(E) = \Delta(E)_{12} = (\Delta\lambda_1)(E_1 - E_2)$, with $\Delta\lambda_1 = \lambda'_1 - \lambda_1$ the change in $\lambda_1$. So $\langle E \rangle(\lambda_1)$ is a line with gradient given by

$$\frac{\Delta\langle E \rangle}{\Delta\lambda_1} = E_1 - E_2.$$  

Similarly

$$\frac{\Delta\langle E \rangle}{\Delta\lambda_1} = \frac{1}{\lambda_{12}}(E_1 - E_2).$$

Comparing this with the gradient of the tangential line to $S_{\lambda 12}$ in Eq. E6, we see that $\frac{1}{\lambda_{12}} \beta(E)_{12}$ has the same gradient as the tangential line. We therefore only need to show that the change in the tangential line is upper bounded by the change in the entropy curve, as it is equivalent to showing that $\Delta S_{\lambda 12} \geq \frac{1}{\lambda_{12}} \beta(E)_{12}$. This must hold for all possible initial and final values of $\lambda_1$ and all possible values of $\lambda_1^T$ (recall that we assumed without loss of generality that $\lambda_1^T \geq 0$). These can be grouped into three cases.

1. $\lambda_1^T \leq \lambda_1^T$. Here the tangential bound above implies that $\Delta S_{\lambda 12} \geq \frac{1}{\lambda_{12}} \beta(E)_{12} \geq 0$.
2. $\lambda_1^T \leq \lambda_1 \leq 0.5$. Here the tangential bound implies that $0 \geq \Delta S_{\lambda 12} \geq \frac{1}{\lambda_{12}} \beta(E)_{12}$.
3. $\lambda_1 \geq 0.5$, also after the interaction. Here the tangential bound implies that $\Delta S_{\lambda 12} \geq 0 \geq \frac{1}{\lambda_{12}} \beta(E)_{12}$.

This implies the lemma. \[\square\]

2. Evolutions respecting standard expression may violate Kelvin’s second law

Recall that our condition on thermalising evolutions was stronger than Eq. E1. There are, as mentioned in the main body, examples of evolutions that respect Eq. E1 but violate our condition: Eq E2. In this subsection we consider whether these evolutions may violate Kelvin’s second law: No process is possible in which the sole result is the absorption of heat from a reservoir and its complete conversion into work.

We use standard results concerning majorisation, as well as our main theorem. We will consider degenerate energy levels for simplicity so that Eq. E1 reduces to $\Delta S \geq 0$. We assume the evolution corresponds to a stochastic matrix.

**Lemma 14.** Any stochastic matrix $A$ which for some state violates Eq E2 but respects the entropy condition $\Delta S \geq 0$ will for some input state, namely the uniform distribution, violate $\Delta S \geq 0$.

**Proof.** (i) Eq E2 is respected iff the matrix is bistochastic. Thus $A$ is NOT bistochastic.