PHYSICS OF INFORMATION IN NONEQUILIBRIUM SYSTEMS

A THESIS SUBMITTED TO THE GRADUATE DIVISION OF THE UNIVERSITY OF HAWAIʻI AT MĀNOA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

PHYSICS

MAY 2019

By

Elan Stopnitzky

Thesis Committee:

Susanne Still, Chairperson
Jason Kumar
Yuriy Mileyko
Xerxes Tata
Jeffrey Yepez

To my late grandmother,

Rosa Stopnitzky

ACKNOWLEDGMENTS

I thank my wonderful family members Benny, Patrick, Shanee, Windy, and Yaniv for the limitless love and inspiration they have given to me over the years. I thank as well my advisor Susanna Still, who has always put great faith in me and encouraged me to pursue my own research ideas, and who has contributed to this work and influenced me greatly as a scientist; my friend and collaborator Lee Altenberg, whom I have learned countless things from and who contributed significantly to this thesis; and my collaborator Thomas E. Ouldridge, who also made important contributions. Finally, I would like to thank my partner Danelle Gallo, whose kindness and support have been invaluable to me throughout this process.

ABSTRACT

Recent advances in non-equilibrium thermodynamics have begun to reveal the fundamental physical costs, benefits, and limits to the use of information. As the processing of information is a central feature of biology and human civilization, this opens the door to a physical understanding of a wide range of complex phenomena. I discuss two areas where connections between non-equilibrium physics and information theory lead to new results: inferring the distribution of biologically important molecules on the abiotic early Earth, and the conversion of correlated bits into work. I show that a non-equilibrium model for the chemistry of the early Earth, which incorporates our uncertainty about the available conditions, predicts average abundances of life's building blocks that are radically larger than their equilibrium estimates and may explain how these molecules were present in appreciable quantities on the early Earth. I also demonstrate fundamental limits on the conversion of correlated bits into work, which arise from physical constraints on the transition matrices that govern the process of conversion.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

Although developed largely independently, information theory and physics are becoming increasingly intertwined. On the one hand, this is consonant with the history in physics of incorporating previously unconnected developments in mathematics, as well as the unification over time of physical laws. On the other hand, it points to the apparent universality of information theory in its own right. Among the ways it might shed light on our world, it is the centrality of information processing in biological processes–including evolution, sensation, and cognition–that is most salient to me. There lies a wide gap between the abstract beauty and technical precision of physics, and the comparatively complex, idiosyncratic, and variegated realm of biology. The motivation behind this work has therefore been to explore links between information theory and physics, with an eye toward the potential for such links to bring a deeper understanding to biology.

My first projet, discussed in Sec. 1.1, Ch. 3, and Sec. 5.1, uses[1] information theory to make inferences about the non-equilibrium distributions of molecules that might have been present on the early Earth, and helps to explain how appreciable quantities of biologically important molecules could be found. The second project, discussed in Sec. 1.2, Ch. 4, and Sec. 5.2, explores[2] the limits of converting information to energy, and was originally conceived as a way to abstract our conception of life to include hypothetical organisms that evolve to metabolize information itself.

Neither of these projects would have been possible without the remarkable developments in non-equilibrium statistical mechanics over the last 30 years. Perhaps the most important of these is the generalization of free energy for non-equilibrium systems [103, 115], which I review in Sec. 2.3. This result is essential for bounding the average work cost of information processing tasks (see, e.g., Eq. 2.55). Another stunning achievement, although of less direct relevance to the work presented here, is a generalization of the of the Second Law due to Jarzynski [58] which takes the form of an equality: $\langle e^{-W/kT} \rangle = e^{-\Delta F/kT}$, where the average is taken over trajectories through state space, given that the system is subject to a

[1]The text of these sections is taken from the manuscript "Non-equilibrium abundances for the building blocks of life", which was coauthored with Susanne Still and is currently under review at Physical Review E.

[2]The text of these sections is taken from the paper "Physical limitations of work extraction from temporal correlations", which was coauthored with Susanne Still, Thomas E. Ouldridge, and Lee Altenberg. It is in press at Physical Review E.

fixed non-equilibrium driving protocol (for example, compressing a piston according to some predefined schedule). Application of Jensen's inequality reduces the Jarzynski expression to the traditional statement that $\langle W \rangle \geq \Delta F$. The Jarzynski equality was later shown by Crooks [30] to follow from the even more powerful statement that $\frac{p(A \to B)}{p(B \to A)} = e^{\beta(W - \Delta F)}$, where $p(A \to B)$ is the probability of observing a particular trajectory through state space due to a given driving protocol $\lambda(t)$, and $p(B \to A)$ is the probability of getting the time-reversed trajectory given that the protocol is time-reversed. That is, this relationship quantifies the probability of observing the reversal of a process as a function of the amount of work that is dissipated in the forward realization of the process, $W - \Delta F$. In this way, the Crooks theorem links precisely the emergence of an arrow of time with dissipation. It has been experimentally verified [99, 28].

These developments and others have opened the door to understanding a wide range of problems that were previously outside the purview of thermodynamics, two of which are explored in this thesis. In the next two sections, I introduce these two projects in more detail.

1.1 Non-equilibrium physics and the origin of life

Biology requires the coordination of many complex molecules to store and copy genetic information, harness energy from the environment, and maintain homeostasis. The spontaneous emergence of life thus hinges upon the abundances of such molecules in an abiotic environment. At first glance, statistical mechanics seems to pose a serious barrier: the high molecular mass and structural specificity of many biomolecules severely limit their abundances in thermodynamic equilibrium and thus make the emergence of life implausible [36, 122, 72, 27, 21]. Many biomolecules require considerable free energy to form, and this leads to an exponential suppression of their equilibrium concentrations.

The apparent severity of this problem, which appears under rather general considerations, has motivated researchers to search for special environments, either extant or belonging to the early Earth, which would be ideally suited for producing the necessary molecules in significant quantities. Due to the free energy requirement, an essential feature of these environments is that they include non-equilibrium driving of some kind [36, 72, 27, 21, 88, 7]. Some proposed sources of this driving on prebiotic Earth are radiation [7, 88], temperature and ion gradients [81, 7, 88, 70], concentration fluxes [4, 104], and electrical discharge [85]. Examples of such environments include hydrothermal vent systems [105, 80, 50, 2], and the surfaces of minerals [72]. Yet it remains an open question to what extent environmental conditions must be fine-

tuned in order to give rise to life, and there is considerable uncertainty about the chemistry of the early Earth [5, 24, 21].

In Ch. 3, we use a relatively new approach proposed by Crooks [31] which allows us to explore how the abundances of life's building blocks change away from thermodynamic equilibrium *on average*, where the average is taken over all the possible ways the system could be driven from equilibrium and depends only on a simple parametric measure of the distance from equilibrium.

Our calculation does not hinge upon specific assumptions about the conditions that might have created life, and therefore does not require significant knowledge about the early Earth. The question we answer is more general: can we quantify how much non-equilibrium conditions *typically* change the abundances of the complex molecules that life relies on? We study this dependence for two simple models describing, respectively, the concentrations of heavy amino acids, and their polymerization into peptides. The result is that away from equilibrium, the abundance of rare molecules become, on average, increasingly favorable, potentially boosted by many orders of magnitude. The specific forms of non-equilibrium driving previously considered can thus be recognized as part of a much more general phenomenon, whereby driving is expected on average to increase the probabilities of rare states as one moves further from equilibrium. By dramatically augmenting the concentrations of biologically important molecules without fine-tuning conditions, this effect makes the appearance of life on Earth a much more plausible event from the standpoint of statistical mechanics.

1.2 Conversion of bits into work

Leo Szilard proposed a simple Gedankenexperiment almost 90 years ago to resolve the paradox of Maxwell's demon, arguing that information about a system could be converted to work by an automated mechanism, in place of a sentient being [113]. Szilard's proposed information engine cyclically repeats two distinct phases: that of acquiring information and recording it into a stable memory, and that of using this information to extract work with a given mechanism. This allowed him to compute a bound on the costs associated with acquiring and recording information, necessary to prevent a violation of the Second Law [113]. Various concerns, for example that a single particle could not be consistently treated as an ideal gas and that the insertion of the barrier might cost work [59], have been alleviated by doing a careful quantum mechanical treatment of the problem that arrives at the

same bounds [127]. More recently, an experimental verification of the Szilard engine was performed using a charge excess to represent the "particle" on a pair of conducting plates [69]. Many extensions to Szilard's engine have also been explored in the literature, e.g. [127, 62, 78, 66, 118, 97].

Independently, it was shown by Landauer that there are thermodynamic consequences to logical irreversibility in computers [73], with $kT\ln(2)$ of dissipated heat for every erased or discarded bit of information. Criticisms of Landauer's results were extensively investigated by Bennett [9, 11] and convincingly shown to be unwarranted. Landauer's principle has since been experimentally verified [13]. Please see Eq. 2.55 and the accompanying discussion for a generalized form of Landauer's bound that is set by the change in Shannon entropy of a memory.

One type of information engine recently proposed by Mandal and Jarzynski exploits a data-carrying tape to extract work from a single heat bath [76], by advancing along a sequence of 0s and 1s that contains an overall bias towards either 0 or 1. In this case, a data-carrying tape refers to any sequence of two-state systems with a degenerate Hamiltonian and an energy barrier between the two states high enough to prevent thermally activated transitions from one state to the other. A sequence of Szilard boxes is a convenient model for such a tape. The device couples to one input bit at a time, and, whilst in contact with the bit, undergoes free-running dynamics that can alter the bit. This interaction increases the entropy of the tape upon output of the changed bit, and it is this entropy increase that is used to compensate for the entropy decrease of the heat bath.

The device is qualitatively different from a traditional Szilard engine, where the ability to do work arises from the correlation between the state of the gas and the state of the memory, the latter being what informs the direction to move the partition. By contrast, the ability to do work in the Mandal/Jarzynski device arises solely from the presence of having a low-entropy medium (the tape) that is out of equilibrium with its environment, and whose transformation toward an equilibrium distribution can therefore be used to do work. In this sense, the interpretation of the information in the two types of device is somewhat different. Another important difference is that while the Szilard engine can draw the full $kT\ln(2)$ per bit that is available to it, the Mandal/Jarzynski device cannot. Rather, the more work it draws the worse its efficiency, with perfect efficiency coinciding with the same limit in which the drawn work goes to zero (see Fig. 2.3). Due to the fact that the Mandal/Jarzynski device was originally presented in a rather abstract manner, and it is a predecessor to the device described in Ch. 4, we present an explicit construction for such a device and analyze

its behavior in Sec. 2.5.

Modern formulations of non-equilibrium thermodynamics naturally incorporate correlations as a potential source of work (see e.g. [39]). The connection between correlations and the ability to do work holds for quantum systems as well, with special difficulties and advantages associated with quantum correlations [55, 93, 92, 23]. Extensions of the Mandal/Jarzynski device exploit statistical information within the tape in the form of temporal correlations [19], or spatial correlations between tapes [83], rather than an overall bias in the input bits. The resulting simple dynamical models of all these proposals help develop a concrete physical understanding of the role information plays in thermodynamics. To serve this purpose, it is important that these devices are physically realizable.

Real physical systems have underlying time-continuous dynamics. Moreover, whenever the work extraction device is designed to operate without a time-dependent, externally-applied driving protocol during the periods of interaction with an individual bit, then the time-continuous dynamics must also be time-homogeneous and obey detailed balance to be physical.

We explore in Ch. 4 how this fact constrains possible designs of the class of *temporal* correlation powered devices proposed in [19], and some references therein. We find that demanding underlying time-continuous, time-homogeneous dynamics drastically limits the set of allowable transition matrices, thereby dramatically reducing the resulting efficiency (Section 4.2).

We demonstrate this effect using two methods: first, by assessing the performance of randomly generated transition matrices, and secondly via an evolutionary algorithm. In both cases we assume that the dynamics are time-homogeneous, and examine how the performance changes when we additionally enforce time-continuous dynamics. In Section 4.3 of Ch. 4, we show that the drastic efficiency limitations arising from enforcing time-continuous dynamics disappear when transition rates are modulated by external manipulation, which is allowable when the restriction to time-homogeneous dynamics is lifted.

CHAPTER 2
BACKGROUND

2.1 Information theory

Claude Shannon introduced in 1948 his "Mathematical theory of communication", now known as information theory [102]. Shannon was concerned primarily with the reliability with which information could be transmitted from a source to a receiver. This requires a suitable measure of information. Suppose that the source consists of a set of symbols labeled by integers $\{1, ..., N\}$, and that the ith symbol is emitted with probability p_i. How much information is conveyed upon receipt of symbol i? Shannon reasoned that any meaningful information measure must satisfy the following properties:

- It must be a continuous function of the probability p_i
- It must be a monotonically decreasing function of p_i . This captures the intuition that we learn more from a surprising, i.e. improbable outcome than an unsurprising one.
- The information gained by two symbols given independently and in succession is equal to the sum of the information gained by each symbol individually.

Shannon proved [102] that the only function satisfying all three of these properties is the so called surprisal $-k \log p_i$, for some constant k. The information conveyed by the source as a whole is then given by the expectation value of the surprisal

$$S = -k \sum_{i=1}^{N} p_i \log p_i. \tag{2.1}$$

This is the famed Shannon entropy or Shannon information. Two connections with the use of entropy in physics become immediately clear. For one, if the ergodicity assumption is invoked, i.e. $p_i = 1/N$ for all i, and k is set equal to the Boltzmann constant, then S reduces to the definition of entropy introduced by Boltzmann, $S = k \log N$. Second, von Neumann argued [121] that the entropy of a mixed quantum system with density operator ρ is given by $S = -Tr(\rho \log \rho)$, where the logarithm of an operator is defined in terms of the Taylor expansion of the logarithm. It is straightforward to show that the von Neumann entropy

of ρ is equal to the Shannon entropy of its eigenvalues. A major theme of this thesis is the exploration of further, often surprising and profound links between information theory and physics.

2.1.1 The source coding theorem

To further motivate the use of Shannon information in the context of information theory, I prove here one of the most important results of information theory, the source coding theorem. The source coding theorem is concerned with the minimal number of symbols needed to encode a source with negligible probability of misidentification. For a source of k input symbols, we might naively think that we require k output symbols to encode the source. However, if we allow small probabilities for errors, then depending on the probabilities p_i with which the symbols occur, it might in fact be possible to encode the source with a significantly reduced alphabet. The source coding theorem quantifies the limits to this task and in doing so, gives an operational meaning to the Shannon information.

To proceed, suppose we encode whole strings of symbols, N at a time, rather than individual symbols. This will ultimately lead to an evaluation of the bit cost per individual symbol. Our aim is to compress the set of length-N strings coming from the source by neglecting to encode some of the more improbable strings we might see from the source. Let A^N denote the set of length-N strings generated by a single-symbol alphabet A, and let δ represent the probability that there is no encoding for outcome $\mathbf{x} \in A^N$. Let T_δ denote the smallest subset (there are potentially multiple choices) such that $p(\mathbf{x} \in T_\delta) \geq 1 - \delta$, and let S denote the Shannon entropy of the source.

Source coding theorem. Given $\epsilon > 0$ and $0 < \delta < 1$, there exists a positive integer N_0 such that for $N > N_0$

$$\mid \frac{1}{N} S_\delta(X^N) - S \mid < \epsilon, \tag{2.2}$$

where X^N is the ensemble of length-N strings, and $S_\delta(X^N)$ is the entropy of the ensemble of strings belonging to T_δ. For long enough strings the approximation $S_\delta(X^N) \approx NS$ can be made arbitrarily accurate, so long as any error probability δ is tolerated. In other words, we can compress length-N strings from the source down to NS bits, or equivalently, length-N strings can be encoded using 2^{NS} symbols with negligible error, but no further.

I now prove the source coding theorem. The proof is outlined in [75], but takes several important theorems from statistics as given. I follow the proof in [75] and fill in the missing details where they are needed.

Lemma 1. Let t be a non-negative real random variable and let α be a positive real number. Then $p(t \geq \alpha) \leq \frac{\bar{t}}{\alpha}$.

Proof: $p(t \geq \alpha) = \sum_{t \geq \alpha} p(t) \leq \sum_{t \geq \alpha} p(t) \frac{t}{\alpha} \leq \frac{\bar{t}}{\alpha}$.

Lemma 2: Chebyshev's inequality. Let x be a random variable and α a positive real number. Then

$$p\big[(x - \bar{x})^2 \geq \alpha\big] \leq \sigma_x^2/\alpha. \tag{2.3}$$

Proof: Follows directly from Lemma 1 with $t = (x - \bar{x})^2$.

Lemma 3: Weak law of large numbers. Let x be the sample average of n independent random variables $h_1, h_2, ..., h_n$. Let each variable have the same mean $\bar{h}$ and variance σ_h^2. Then $\bar{x} = \bar{h}$, $\sigma_x^2 = \sigma_h^2/n$, and

$$p\big[(\bar{x} - \bar{h})^2 \geq \alpha\big] \leq \frac{\sigma_h^2}{n\alpha}. \tag{2.4}$$

Proof: We begin with the result that $\bar{x} = \bar{h}$. We first prove the slightly more general statement that $\overline{a_1 h_1 + a_2 h_2 + ... a_n h_n} = (a_1 + a_2 + ... a_n)\bar{h}$ for any real numbers $\{a_i\}$, from which the statement follows by setting $a_i = 1/n$ for all i. This result will be be useful later on. It follows directly from the linearity of the expectation value: $\overline{ax} = \sum_i p(ax_i)(ax_i) = a \sum_i p(x_i) x_i = a\bar{x}$ and $\overline{x_1 + x_2} = \sum_{s = x_1 + x_2} p(x_1 + x_2)(x_1 + x_2) = \sum_{s, x_1} p(x_1) p(x_2 = s - x_1)(x_1 + x_2) = \sum_{x_1, x_2} p(x_1) p(x_2)(x_1 + x_2) = \bar{x_1} + \bar{x_2}$.

The proof that $\sigma_x^2 = \sigma_h^2/n$ is a bit more involved. First, we prove that for independent variables h_1 and h_2 with common mean $\bar{h}$

$$Var(a_1 h_1 + a_2 h_2) = a_1^2 Var(h_1) + a_2^2 Var(h_2). \tag{2.5}$$

Let $s = a_1h_1 + a_2h_2$. Then by the same argument as before, $\sum_s p(s) = \sum_{h_1,h_2} p(h_1)p(h_2)$. With this fact, and invoking the previous result that $\overline{a_1h_1 + a_2h_2} = (a_1 + a_2)\bar{h}$, we have $Var(s) = \sum_{h_1,h_2} p(h_1)p(h_2)\big[a_1h_1 + a_2h_2 - (a_1 + a_2)\bar{h}\big]^2$. We focus for now on the argument in the sum: $\big[a_1h_1 + a_2h_2 - (a_1 + a_2)\bar{h}\big]^2 = (a_1h_1 + a_2h_2)^2 - 2(a_1h_1 + a_2h_2)\bar{h} + (a_1 + a_2)^2\bar{h}^2 = a_1^2(h_1 - \bar{h})^2 + a_2^2(h_2 - \bar{h})^2 + 2a_1a_2(h_1 - \bar{h})(h_2 - \bar{h})$. Substituting each of these back into the sum gives $Var(s) = a_1^2Var(h_1) + a_2^2Var(h_2) + 2a_1a_2 \sum_{h_1,h_2} p(h_1)p(h_2)(h_1 - \bar{h})(h_2 - \bar{h})$. The last sum is the covariance of two independent variables and can easily be evaluated to zero. We now consider the case that $a_i = 1/n$ for all i, and prove that $\sigma_x^2 = \sigma_h^2/n$ with induction. The base case follows trivially from Eq. 2.5. Define s_n to be the sum of h_1 through h_n. For the inductive hypothesis, we invoke Eq. 2.5 to get $Var(\frac{s_n+h_{n+1}}{n+1}) = Var(\frac{n}{n+1}\frac{s_n}{n} + \frac{h_{n+1}}{n+1}) = (\frac{n}{n+1})^2Var(\frac{s_n}{n}) + (\frac{1}{n+1})^2Var(h_{n+1}) = (\frac{n}{n+1})^2\frac{\sigma_h^2}{n} + (\frac{1}{n+1})^2\sigma_h^2 = \big[(\frac{n}{n+1})^2\frac{1}{n} + (\frac{1}{n+1})^2\big]\sigma_h^2 = \frac{\sigma_h}{n+1}$ as desired. This completes the proof that $\sigma_x^2 = \sigma_h^2/n$. Finally, substituting $\bar{x} = \bar{h}$ and $\sigma_x^2 = \sigma_h^2/n$ into the Chebyshev inequality Eq. 2.3, we have the weak law of large numbers Eq. 2.4. It states that by choosing n large enough, the probability of obtaining a difference between the sample mean $\bar{x}$ and the true mean $\bar{h}$ greater than some tolerance $\sqrt{\alpha}$ can be made arbitrarily small.

Returning now to the proof of the source coding theorem, we define a "typical" subset of the strings $\mathbf{x} \in A^N$, given parameters β and N, as

$$T_{N\beta} = \{\mathbf{x} \mid (\frac{-\log p(\mathbf{x})}{N} - S)^2 < \beta^2\}. \tag{2.6}$$

By applying the law of large numbers Eq. 2.4 to the average $\frac{-\log p(\mathbf{x})}{N} = \frac{-1}{N}\log(\prod_{i=1}^N x_i) = \frac{-1}{N}\sum_{i=1}^N \log(x_i)$ we get

$$p(\mathbf{x} \in T_{N\beta}) \geq 1 - \frac{\sigma^2}{N\beta^2}, \tag{2.7}$$

where x_i is the ith symbol in the string and σ^2 is the variance of the variable $-\log x$ (recall that the symbols in the string are assumed to be independent and identically distributed). Note that from the definition of the typical set Eq. 2.6, and assuming that we are measuring entropy using a logarithm of base 2, we have that the probability of any string belonging to the set is bounded by

$$2^{-N(S+\beta)} < p(\mathbf{x}) < 2^{-N(S-\beta)}. \tag{2.8}$$

Source coding theorem part 1: $\frac{1}{N}S_\delta(X^N) < S + \epsilon$. Since the total probability of belonging to $T_{N\beta}$ is at most 1 and by Eq. 2.8 each element has probability at least $2^{-N(S+\beta)}$, we have $\mid T_{N\beta} \mid 2^{-N(S+\beta)} < 1$. Therefore

$$\mid T_{N\beta} \mid < 2^{N(S+\beta)} \tag{2.9}$$

and so $\log \mid T_{N\beta} \mid < N(S+\beta)$. $\log \mid T_{N\beta} \mid$ is an upper bound for the entropy $S_{T_{N\beta}}(X^N)$ of the ensemble of strings belonging to $T_{N\beta}$, because the entropy is maximal and equal to $\log W$ when a set of W possible outcomes each have equal probability. Combining these facts gives

$$S_{T_{N\beta}}(X^N) \leq \log \mid T_{N\beta} \mid < N(S+\beta). \tag{2.10}$$

Setting $\beta = \epsilon$ and $\frac{\sigma^2}{\epsilon^2 N_0} \leq \delta$, we have from Eq. 2.7 that $p(\mathbf{x} \in T_{N\beta}) \geq 1 - \delta$. We thus recognize $T_{N\beta} = T_{N\epsilon}$ as an example of one of the "smallest sufficient subsets" T_δ that we sought. Combining this fact with Eq. 2.10 completes this half of the source coding theorem; by ignoring strings that do not belong to $T_{N\epsilon}$ we can encode all the strings in $T_{N\epsilon}$ at bit cost less than $N(S+\epsilon)$ while incurring a probability of failure (that is, of seeing a string that doesn't belong to $T_{N\epsilon}$ and for which we don't have an encoding) less than δ.

Source coding theorem part 2: $\frac{1}{N}S_\delta(X^N) > S - \epsilon$. We prove this by contradiction. Assume that for any N there exists a subset T' which is smaller than $T_{N\beta}$ and for which the probability of error remains less than δ. The probability of belonging to this alternate set is

$$p(\mathbf{x} \in T') = p(\mathbf{x} \in T' \cap T_{N\beta}) + p(\mathbf{x} \in T' \cap T^c_{N\beta}) \tag{2.11}$$

where $T^c_{N\beta}$ is the complement of $T_{N\beta}$. Since T' is assumed to be smaller than $T_{N\beta}$, we assume its cardinality has an upper bound $\mid T' \mid \leq 2^{N(S-2\beta)}$ (recall Eq. 2.9). By Eq. 2.8, $p(\mathbf{x} \in T' \cap T_{N\beta})$ is upper bounded by

$$p(\mathbf{x} \in T' \cap T_{N\beta}) < 2^{N(S-2\beta)} \cdot 2^{-N(S-\beta)} = 2^{-N\beta}. \tag{2.12}$$

The maximum value of the second term in Eq. 2.11 is given by

$$p(\mathbf{x} \in T' \cap T^c_{N\beta}) \leq p(\mathbf{x} \notin T_{N\beta}) = \frac{\sigma^2}{\beta^2 N}. \tag{2.13}$$

Combining Eqs. 2.11,2.12, and 2.13 we have

$$p(\mathbf{x} \in T') \leq 2^{-N\beta} + \frac{\sigma^2}{\beta^2 N}. \tag{2.14}$$

Thus, $p(\mathbf{x} \in T')$ is a decreasing function of N, and by choosing N_0 large enough we can ensure that for $N > N_0$, $p(\mathbf{x} \in T') < 1 - \delta$, invalidating the assumption that T' could serve as a sufficient subset T_δ. This completes the proof of the source coding theorem.

2.1.2 The Kullback-Leibler divergence and mutual information

We now introduce an important quantity from information theory known as the Kullback-Leibler divergence or relative entropy:

$$D(\rho \,||\, \phi) \equiv \sum_i \rho_i \ln \frac{\rho_i}{\phi_i} \tag{2.15}$$

where ρ and ϕ are distributions. We define $D(\rho \,||\, \phi)$ in terms of the natural logarithm rather than a logarithm of base 2 due to the fact that it will be more convenient for our uses, but this only changes $D(\rho \,||\, \phi)$ by a constant factor. The non-negativity of the Kullback-Leibler divergence is essential to its use in Sec. 2.3 and Ch. 3. We now take a minor detour to prove this fact.

Non-negativity of the Kullback-Leibler divergence: $D(p \,||\, q) \geq 0$ with equality if and only if $p = q$.

Proof: We first establish a special case of Jensen's inequality, $\langle \ln x \rangle \leq \ln \langle x \rangle$ (angled brackets denote expected value):

$$\langle \ln x \rangle - \ln\langle x \rangle = \langle \ln \frac{x}{\langle x \rangle} \rangle \tag{2.16}$$

$$\leq \langle \frac{x}{\langle x \rangle} - 1 \rangle \tag{2.17}$$

$$= \frac{\langle x \rangle}{\langle x \rangle} - 1 = 0 \tag{2.18}$$

where we used in the second line that $\ln y \leq y - 1$ for all y, with equality when when $y = 1$. This establishes the result. From here the non-negativity of the Kullback-Leibler divergence follows easily:

$$-D(p \mid\mid q) = \sum_i p_i \ln \frac{q_i}{p_i} \tag{2.19}$$

$$\leq \ln \sum_i p_i \frac{q_i}{p_i} \tag{2.20}$$

$$= \ln \sum_i q_i = 0. \tag{2.21}$$

Another important quantity in information theory is the mutual information between two variables,

$$I[X;Y] = D(P_{XY} \mid\mid P_X P_Y) \tag{2.22}$$

$$= \sum_{x,y} p(x,y) \ln \frac{p(x,y)}{p(x)p(y)} \tag{2.23}$$

$$= S(X) + S(Y) - S(X,Y) \tag{2.24}$$

where P_{XY} denotes the joint distribution of variables X and Y, and P_X,P_Y are the associated marginal distributions. The joint entropy $S(X,Y)$ is defined by $S(X,Y) = -\sum_{x,y} p(x,y) \ln p(x,y)$. The mutual information measures the reduction in uncertainty about one variable when one has access to the other, and is symmetric with respect to the two variables:

$$I[X;Y] = H[X] - H[X|Y] \tag{2.25}$$
$$= H[Y] - H[Y|X] \tag{2.26}$$

where $H[X|Y] \equiv -\sum_{x,y} p(x,y) \ln p(x|y)$. These equivalent expressions for the mutual information can easily be established through application of Bayes' Rule. By Eq. 2.22 the mutual information inherits non-negativity from the Kullback-Leibler divergence. This will prove useful in Sec. 2.3.

2.2 MaxEnt and statistical physics

E.T. Jaynes showed in his seminal work [61] that the Boltzmann distribution of equilibrium statistical mechanics could be derived using information theory. This approach naturally connects Shannon entropy $S = -\sum_i p_i \ln p_i$ with the thermodynamic entropy. Jaynes considered the distribution of states in an equilibrium system as an inference problem. That is, given that we cannot possibly hope to provide a complete microscopic description of a macroscopic system, the best we can do is to give probabilities for the various states of the system. How should these probabilities be assigned? Clearly, they must be chosen in such a way that the macroscopic properties of the system under observation are recovered. More concretely, suppose that we have measured the average energy of the system to be $\langle E \rangle$. Then, whatever distribution we assign, a necessary requirement is that

$$\langle E \rangle = \sum_i E_i p_i. \tag{2.27}$$

Yet this is just one constraint; for macroscopic systems our description would still be woefully under-constrained. Jaynes' insight was that the distribution we assign to the system should not contain any information we do not actually have access to. In other words, we want to assign probabilities that reflect maximum possible ignorance of the system (i.e. a distribution as uniform as possible), while enforcing constraints based on the knowledge we do have, which take the form of ensemble averages. Maximizing our ignorance is equivalent to maximizing the Shannon entropy of the system. This procedure suggests the following objective function to be maximized:

$$L = -\sum_i (p_i \ln p_i - \lambda_0 p_i - \lambda_1 E_i p_i) \tag{2.28}$$

where λ_0 is associated with the normalization constraint $\sum_i p_i = 1$ and λ_1 with the constraint that $\langle E \rangle = \sum_i E_i p_i$. Setting the derivative with respect to p_i of the argument to zero gives

$$-\ln p_i - 1 - \lambda_0 - \lambda E_i = 0. \tag{2.29}$$

Solving for p_i gives

$$p_i \propto e^{-\lambda E_i} \tag{2.30}$$

with the proportionality constant set by normalization. Finally, we recognize λ as corresponding with the inverse temperature of the system.

The idea that we should maximize Shannon entropy–subject to appropriate constraints–when modeling systems for which our knowledge is incomplete is now known as MaxEnt, and has been used with remarkable success to describe a wide variety of disparate systems [106, 64], including for example the distributions of species in an ecosystem [49], the distribution of spiking patterns in a neural population [116], and the distribution of antibody sequences in the immune system [86].

2.3 Non-equilibrium free energy

For transitions between equilibrium states, the Second Law says that the work required to effect the change is bounded by the change in free energy of the system: $W \geq \Delta F$. How must this expression be modified when the initial and final states are not in thermal equilibrium? It turns out that there is a state function F_{neq}, such that $W \geq \Delta F_{neq}$ for transitions between non-equilibrium states. Thus, there is a natural generalization of the free energy for non-equilibrium systems. We prove the statement $W \geq \Delta F_{neq}$ for an appropriately defined F_{neq} in this section, following the derivation given in [115].

Consider an isolated system consisting of two subsystems A and B in contact with one another. Let $X = (x, y)$ denote the state of the joint system. The total Hamiltonian for the system is given by

$$H_t(X) = H(X, \theta_t) = H^{(A)}(x, a_t) + H^{(B)}(y, b_t) + H^{(I)}(x, y) \tag{2.31}$$

where $\theta_t = (a_t, b_t)$ are time-dependent parameters that describe external manipulations applied to the system. We assume that at the beginning ($t = 0$) and end ($t = T$) of the protocol that the interaction term $H^{(I)}(x, y)$ is turned off. At each moment in time, the system has an associated equilibrium distribution

$$\rho_{eq,t}(X, \alpha) = e^{\alpha(F_t(\alpha) - H_t(X))} \tag{2.32}$$

where as usual $F_t(\alpha) \equiv -\alpha^{-1} \ln Z$ and $Z = \sum_X e^{-\alpha H_t(X)}$. α plays the role of an inverse temperature, and we have absorbed the Boltzmann constant into it for convenience. We define the entropy of any distribution ρ_t to be the Shannon entropy $S = -\sum_X \rho_t(X) \ln \rho_t(X)$. Note that in equilibrium we recover the usual relationship between entropy, energy, and free energy:

$$\begin{aligned} S(\rho_{eq,t}) &= -\sum_X \rho_t(X) \ln \rho_t(X) && (2.33) \\ &= -\sum_X \rho_t(X) \alpha \big[F_t(\alpha) - H_t(X)\big] && (2.34) \\ &= \langle E_t \rangle - \alpha F_t(\alpha) && (2.35) \end{aligned}$$

where $\langle E_t \rangle \equiv \sum_X \rho_t(X) H_t(X)$. Since the system as a whole is isolated, any changes to its total energy over the duration of the external protocol running from $t = 0$ to $t = T$ are due to work. We therefore have

$$W = \langle E_T \rangle - \langle E_0 \rangle. \tag{2.36}$$

Due to the fact that free energy is at a minimum in equilibrium, we have at the end of the protocol the inequality

$$\langle E_T^{(B)} \rangle \geq F_T^{(B)}(\alpha) + \alpha^{-1} S_T^{(B)}. \tag{2.37}$$

From the non-negativity of the mutual information we have $S_T^{(A)} + S_T^{(B)} - S_T = S_T^{(A)} + S_T^{(B)} - S_0 \geq 0$, where S_T denotes the entropy of the total system at time T, and we have used the fact that the total entropy is conserved under Hamiltonian dynamics. This means that $S_T^{(B)} \geq S_0 - S_T^{(A)}$. We use this result to extend the inequality in Eq. 2.37 to

$$\langle E_T^{(B)} \rangle \geq F_T^{(B)}(\alpha) + \alpha^{-1} S_0 - \alpha^{-1} S_T^{(A)}. \tag{2.38}$$

Before bounding the work cost of the protocol, one more observation is needed. That is,

$$\begin{aligned} D(\rho_t \,||\, \rho_{can,t}) &= \sum_i \rho_t(i) \ln \frac{\rho_t(i)}{\rho_{can,t}(i)} && (2.39) \\ &= -S_t - \sum_i \rho_t(i) \alpha \big[F_t(\alpha) - H_t(i) \big] && (2.40) \\ &= -S_t - \alpha F_t(\alpha) + \alpha \langle E_t \rangle. && (2.41) \end{aligned}$$

Altogether, we have for the work cost

$$\begin{aligned} W &= \langle E_T \rangle - \langle E_0 \rangle && (2.42) \\ &= \langle E_T^{(A)} \rangle + \langle E_T^{(B)} \rangle - \langle E_0 \rangle && (2.43) \\ &\geq \langle E_T^{(A)} \rangle + F_T^{(B)}(\alpha) - \langle E_0 \rangle + \alpha^{-1} S_0 - \alpha^{-1} S_T^{(A)} && (2.44) \end{aligned}$$

where we have used the fact that at the beginning and end of the protocol there is assumed to be no interaction Hamiltonian. We used Eq. 2.38 in line 3. Now, by applying Eq. 2.41, we replace the $\langle E_T^{(A)} \rangle - \alpha^{-1} S_T^{(A)}$ term with $\alpha^{-1} D(\rho_T^{(A)} \,||\, \rho_{can,T}^{(A)}) + F_T^{(A)}(\alpha)$. The same is done to replace $-\langle E_0 \rangle + \alpha^{-1} S_0$ with $-\alpha^{-1} D(\rho_0 \,||\, \rho_{can,0}) - F_0(\alpha)$. Substituting these expressions into Eq. 2.44 we get

$$
\begin{aligned}
W &\geq F_T^{(A)}(\alpha) + F_T^{(B)}(\alpha) - F_0(\alpha) + \alpha^{-1} D(\rho_T^{(A)} \,||\, \rho_{can,T}^{(A)}) - \alpha^{-1} D(\rho_0 \,||\, \rho_{can,0}) && (2.45)\\
&= F_T(\alpha) - F_0(\alpha) + \alpha^{-1} D(\rho_T^{(A)} \,||\, \rho_{can,T}^{(A)}) - \alpha^{-1} D(\rho_0 \,||\, \rho_{can,0}) && (2.46)\\
&= \Delta F(\alpha) + \alpha^{-1} D(\rho_T^{(A)} \,||\, \rho_{can,T}^{(A)}) - \alpha^{-1} D(\rho_0 \,||\, \rho_{can,0}) && (2.47)
\end{aligned}
$$

where we used once again that when the interaction Hamiltonian is turned off energies and entropies factor. We are interested in the special case where system B is effectively infinite and serves as a thermal reservoir at temperature T that stays in equilibrium throughout the duration of the protocol. In this case we have

$$
\begin{aligned}
D(\rho_0 \,||\, \rho_{can,0}) &= D(\rho_0^{(A)} \rho_{can,0}^{(B)} \,||\, \rho_{can,0}^{(A)} \rho_{can,0}^{(B)}) && (2.48)\\
&= \sum_{i,j} \rho_0^{(A)}(i) \rho_{can,0}^{(B)}(j) \ln \frac{\rho_0^{(A)}(i) \rho_{can,0}^{(B)}(j)}{\rho_{can,0}^{(A)}(i) \rho_{can,0}^{(B)}(j)} && (2.49)\\
&= \sum_{i,j} \rho_0^{(A)}(i) \rho_{can,0}^{(B)}(j) \ln \frac{\rho_0^{(A)}(i)}{\rho_{can,0}^{(A)}(i)} && (2.50)\\
&= D(\rho_0^{(A)} \,||\, \rho_{can,0}^{(A)}) && (2.51)
\end{aligned}
$$

and so Eq. 2.47 reduces to

$$
W \geq \Delta F^{(A)} + T\big[D(\rho_T^{(A)} \,||\, \rho_{can,T}^{(A)}) - D(\rho_0^{(A)} \,||\, \rho_{can,0}^{(A)})\big]. \tag{2.52}
$$

This equation motivates the introduction of a new state function

$$
F_{neq} \equiv F_{eq} + T D(\rho \,||\, \rho_{can}) \tag{2.53}
$$

which bounds the work cost of transitions between non-equilibrium states and naturally generalizes the equilibrium free energy. A protocol capable of saturating the bound is given

in [115], which establishes that the bound is tight. The second term $TD(\rho \mid\mid \rho_{can})$ can be understood as the amount of work potential that would be dissipated if the system were to relax to equilibrium (see the discussion at the end of the next section), where $D(\rho_{can} \mid\mid \rho_{can}) = 0$. Eq. 2.53 can also be written

$$F_{neq}(\rho) = \langle E \rangle_\rho - TS(\rho) \tag{2.54}$$

where ρ is any distribution and $S(\rho)$ is its Shannon entropy. Thus, by replacing the Boltzmann entropy with the Shannon entropy, we obtain a version of the free energy that remains valid for non-equilibrium systems and that reduces to the standard free energy when ρ is an equilibrium distribution. We make use of the non-equilibrium free energy in Ch. 3,2.5, and 4.

An important consequence of Eq. 2.52 is that it allows us to bound the cost of general information processing tasks. That is, processing information generally involves transforming one distribution ρ into another distribution ρ'. For example, erasure involves the transformation of some distribution ρ into a predefined state whose probability is thereafter 1. Logic gates also effect transformations on distributions of input bits. Assuming a degenerate Hamiltonian, the cost of any such transformation is bounded as

$$W \geq T\big[S(\rho) - S(\rho')\big]. \tag{2.55}$$

Eq. 2.55 was recently used to estimate a lower bound on the cost of synthesizing proteins in cells [65], given that it involves a random soup of amino acids getting converted to a bound polymer with a definite sequence. It was found that cells perform this task at an efficiency within a factor of about 20 of the Landauer limit. Remarkably, this is roughly 5 orders of magnitude more efficient than our best supercomputers [65].

It is worth noting that Eq. 2.52 and by extension Eq. 2.55, hold for quantum systems as well [40, 38, 57, 23]. Generalizations in the quantum regime have also been provided for the case in which one cares about single-shot bounds rather than average behavior [41, 42].

2.4 Some relevant results from stochastic thermodynamics

Stochastic thermodynamics is a powerful framework for describing systems out of equilibrium. In this framework, the distribution of states is allowed to evolve in time under the action of a transition matrix, which may itself be time-dependent. With the right definitions, the usual laws of thermodynamics can naturally be extended to this framework. This section is based on the review given in [119].

Let p_m denote the probability of occupying state m, and $W_{m,m'}$ denote the rate at which transitions are made from m' to m. The time evolution of p_m is given by a master equation

$$\dot{p}_m = \sum_{m'} W_{m,m'} p_{m'} - W_{m',m} p_m \tag{2.56}$$

$$= \sum_{m'} J_{m,m'} \tag{2.57}$$

where the flux $J_{m,m'} \equiv W_{m,m'} p_{m'} - W_{m',m} p_m$. We can simplify Eq. 2.58 by defining $W_{m,m} = -\sum_{m' \neq m} W_{m',m}$. Then Eq. 2.58 becomes

$$\dot{p}_m = \sum_{m'} W_{m,m'} p_{m'}. \tag{2.58}$$

This is the convention used in Ch.4. The new matrix $\mathbf{W}$ comprised of the rates $W_{m,m'}$ has the property that the sum of each column is equal to zero; this ensures conservation of probability. This property also guarantees the existence of a steady state distribution $\dot{p}_m^{st} = 0$, to which the system will eventually relax. Let ϵ_m denote the energy of state m. Then the energy of the system is given by

$$E = \sum_m p_m \epsilon_m \tag{2.59}$$

and its time derivative splits into two separate contributions to the change in energy:

$$\dot{E} = \sum_m \dot{p}_m \epsilon_m + p_m \dot{\epsilon}_m, \tag{2.60}$$

which are identified with the work and heat fluxes

$$\dot{Q} \equiv \sum_m \dot{p}_m \epsilon_m \tag{2.61}$$

$$\dot{W} \equiv \sum_m p_m \dot{\epsilon}_m. \tag{2.62}$$

That is, heat is associated with changes to the occupancies of the states, and work is associated with the external manipulation of energy levels. With these definitions, the First Law takes the form

$$\dot{E} = \dot{Q} + \dot{W}. \tag{2.63}$$

Physical consistency also demands that the rate matrix $\mathbf{W}$ satisfies the condition of detailed balance (we will demonstrate why shortly) at each moment:

$$W_{m,m'}(t) p^{eq}_{m'}(t) = W_{m',m}(t) p^{eq}_{m}(t) \tag{2.64}$$

where $p^{eq}_m = e^{\beta(F-\epsilon_m)}$ refers to the equilibrium, steady state distribution to which the system must eventually relax if the rate matrix is held fixed. A matrix that satisfies detailed balance is known as a reversible matrix. The condition means that in equilibrium, there cannot be any net flux from one state to another. This is closely related to the notion that in equilibrium there is no arrow of time, whose existence requires dissipation (see [30] for precise details). Eq. 2.64 also implies that changing energy levels must alter the transition rates.

As before, we use the Shannon entropy

$$S = -k_B \sum_m p_m \ln p_m \tag{2.65}$$

where p_m is a solution to the master equation Eq. 2.58 and k_B is Boltzmann's constant. From Eq. 2.65 the rate of change of the entropy is

$$\dot{S} = -k_B \sum_m \dot{p}_m \ln p_m + \dot{p}_m \tag{2.66}$$

$$= -k_B \sum_m \dot{p}_m \ln p_m \tag{2.67}$$

where we have used that $\sum_m \dot{p}_m$ must be 0 by conservation of probability. Substituting Eq. 2.58 into Eq. 2.66 gives

$$\dot{S} = -k_B \sum_{m,m'} W_{m,m'} p_{m'} \ln p_m \tag{2.68}$$

$$= \frac{1}{2} k_B \sum_{m,m'} \left[W_{m,m'} p_{m'} - W_{m',m} p_m \right] \ln \frac{p_{m'}}{p_m} \tag{2.69}$$

where the terms with factors of $\ln p_{m'}$ and $-\ln p_m$ each sum to the same thing because swapping indices introduces a minus sign; this is the reason for the factor of $1/2$ out front. Multiplying $\frac{p_{m'}}{p_m}$ by $\frac{W_{mm'} W_{m'm}}{W_{m'm} W_{mm'}} = 1$ and substituting into the log in Eq. 2.69 gives (we will drop the factor of k_B for simplicity from now on)

$$\dot{S} = \frac{1}{2} \sum_{m,m'} \left[W_{m,m'} p_{m'} - W_{m',m} p_m \right] \ln \frac{W_{mm'} p_{m'} W_{m'm}}{W_{m'm} p_m W_{mm'}} \tag{2.70}$$

$$= \frac{1}{2} \sum_{m,m'} \left[W_{m,m'} p_{m'} - W_{m',m} p_m \right] \ln \frac{W_{mm'} p_{m'}}{W_{m'm} p_m} \tag{2.71}$$

$$+ \frac{1}{2} \sum_{m,m'} \left[W_{m,m'} p_{m'} - W_{m',m} p_m \right] \ln \frac{W_{m'm}}{W_{mm'}} \tag{2.72}$$

$$= \dot{S}_i + \dot{S}_e \tag{2.73}$$

where we have defined the entropy production $\dot{S}_i$ as

$$\dot{S}_i = \frac{1}{2} \sum_{m,m'} \left[W_{m,m'} p_{m'} - W_{m',m} p_m \right] \ln \frac{W_{mm'} p_{m'}}{W_{m'm} p_m} \tag{2.74}$$

and the entropy flow $\dot{S}_e$ as

$$\dot{S}_e = \frac{1}{2}\sum_{m,m'}\left[W_{m,m'}p_{m'} - W_{m',m}p_m\right]\ln\frac{W_{m'm}}{W_{mm'}}. \tag{2.75}$$

The entropy production $\dot{S}_i$ is the irreversible contribution to the change in entropy and satisfies $\dot{S}_i \geq 0$ because $(x-y)\ln\frac{x}{y} \geq 0$. It is positive whenever detailed balance is broken, i.e. when the system is out of equilibrium. We can recognize the entropy flow $\dot{S}_e$ as the reversible contribution coming from the flow of heat using Eq. 2.64:

$$\begin{aligned}
\ln\tfrac{W_{m'm}}{W_{mm'}} &= \ln\frac{p^{eq}_{m'}}{p^{eq}_{m}} && (2.76)\\
&= \beta\big[F-\epsilon_{m'}\big] - \beta\big[F-\epsilon_m\big] && (2.77)\\
&= \beta(\epsilon_m - \epsilon_{m'}) && (2.78)\\
&= \beta q_{m,m'} && (2.79)
\end{aligned}$$

where $q_{m,m'}$ is the heat associated with a jump from state m' to m (recall the definition of the heat coming from Eq. 2.61).

We can gain further insight by decomposing the entropy production $\dot{S}_i$ as follows:

$$\dot{S}_i = \dot{S}_a + \dot{S}_{na} \tag{2.80}$$

where the adiabatic entropy production $\dot{S}_a$ and non-adiabatic entropy production $\dot{S}_{na}$ are defined by

$$\dot{S}_a = \frac{1}{2}\sum_{m,m'}\left[W_{m,m'}p_{m'} - W_{m',m}p_m\right]\ln\frac{W_{mm'}p^{st}_{m'}}{W_{m'm}p^{st}_{m}} \tag{2.81}$$

and

$$\dot{S}_{na} = \frac{1}{2} \sum_{m,m'} \left[W_{m,m'} p_{m'} - W_{m',m} p_m \right] \ln \frac{p_{m'} p^{st}_m}{p_m p^{st}_{m'}}. \tag{2.82}$$

We see from Eq. 2.82 that the non-adiabatic entropy production vanishes when the system is in steady state (which is not necessarily equilibrium), and is therefore associated with the relaxation to steady state. On the other hand, from Eq. 2.81 we see that the adiabatic entropy production persists in steady state if the steady state is not the equilibrium distribution (recall the detailed balance condition Eq. 2.64), and vanishes only if the system is in equilibrium. $\dot{S}_{na}$ can also be written (recall the definition of the Kullback-Leibler divergence given in Eq. 2.15)

$$\dot{S}_{na} = -\dot{D}(p_m \,||\, p^{st}_m) \tag{2.83}$$

$$= -\frac{d}{dt} \sum_m p_m \ln \frac{p_m}{p^{st}_m} \tag{2.84}$$

$$= -\sum_m \dot{p}_m \ln \frac{p_m}{p^{st}_m}. \tag{2.85}$$

We now show that $\dot{D}(p_m \,||\, p^{st}_m) \leq 0$ and by extension $\dot{S}_{na} \geq 0$:

$$\dot{D}(p_m \,||\, p_m^{st}) = \sum_m \dot{p}_m \ln \frac{p_m}{p_m^{st}} \tag{2.86}$$

$$= \sum_{m,m'} W_{m,m'} p_{m'} \ln \frac{p_m}{p_m^{st}} \tag{2.87}$$

$$= \frac{1}{2} \sum_{m,m'} W_{m,m'} p_{m'} \ln \frac{p_m p_{m'}^{st}}{p_{m'} p_m^{st}} \tag{2.88}$$

$$= \frac{1}{2} \sum_{m,m';m\neq m'} W_{m,m'} p_{m'} \ln \frac{p_m p_{m'}^{st}}{p_{m'} p_m^{st}} \tag{2.89}$$

$$\leq \frac{1}{2} \sum_{m,m';m\neq m'} W_{m,m'} p_{m'} \big[\frac{p_m p_{m'}^{st}}{p_{m'} p_m^{st}} - 1\big] \tag{2.90}$$

$$= \frac{1}{2} \sum_{m,m'} W_{m,m'} p_{m'} \big[\frac{p_m p_{m'}^{st}}{p_{m'} p_m^{st}} - 1\big] \tag{2.91}$$

$$= 0 \tag{2.92}$$

where we used once again that $\ln x \leq x - 1$ with equality when $x = 1$ (we show the restricted sum in the fourth and fifth lines so that taking a log of 0 is avoided). The bound is thus saturated if and only if the distribution is in steady state. That is, $D(p_m \,||\, p_m^{st})$ decreases in value until steady state is reached, and its rate of change is the negative of the non-adiabatic entropy production. This validates our earlier claim at the end of section 2.3 that $D(p_m \,||\, p_m^{eq})$ is an additional contribution to the equilibrium free energy that corresponds with the amount of energy that would be dissipated to the environment if the system were allowed to relax to equilibrium.

2.5 An explicit design for a device that can extract work from an information tape

In this section, we provide an explicit example of a device that can use an excess of 1s or 0s in an input tape to extract work from a heat bath. The idea is that during the operation of the device the frequencies of 1s and 0s is shifted towards being more even, and so the

increase in entropy of the tape is analogous to the transfer of heat from a hot to a cold reservoir in a traditional heat engine. Note that the type of information used by such a device is different from the information utilized by a Szilard engine [114]. In the former, it is a statistical excess of 1s or 0s that serves as a source of "fuel", whereas in a Szilard engine the ability to extract work arises from correlation between bits (specifically, between the bit stored by the memory and the bit reflecting the position of the particle).

A more abstract description of such a device is given in [76], and we show that it can be implemented with a Gedenkenexperiment that extends the Szilard engine. Let each site on the input "tape" be a rectangular box with an impermeable barrier down the middle that separates the box into two equally sizes cubes. A single particle of gas resides in one or the other cube, with the other cube empty. Let the initial distribution describing which of the two sides of the box the particle resides in be denoted by p. The boxes sit in a sequence, with the particle locations independent and identically distributed. The axis along which we define left and right is perpendicular to the length of the tape. The operation of the work extraction device is as follows:

- Step 1: A step of height h is raised quasi-statically on the right side of the box, compressing its volume from l^3 to $l^2(l-h)$, where l is the side length of the cubic half-box. If we let the symbol 1 denote the right side of the box, then from the ideal gas law we have $W_{in} = p_1 kT \ln \frac{l^3}{l^2(l-h)} = p_1 kT \ln \frac{l}{l-h}$ per input site must be drawn from a work reservoir to raise the step, where p_1 is the initial probability of finding the particle on the right side of the partition[1].

- Step 2: The partition running down the middle of the box is removed, and the particle is given time to relax to its equilibrium distribution, which we denote by p^{eq}.

- Step 3: The partition is reinserted.

- Step 4: The step is quasi-statically lowered, and in the process $W_{out} = p_1^{eq} kT \ln \frac{l}{l-h}$ per site is added to the work reservoir.

[1]We write all work expressions as positive quantities, and state explicitly whether they correspond to energy added to or taken from the work reservoir.

- Step 5: The device advances to the next site on the tape and the process is repeated.

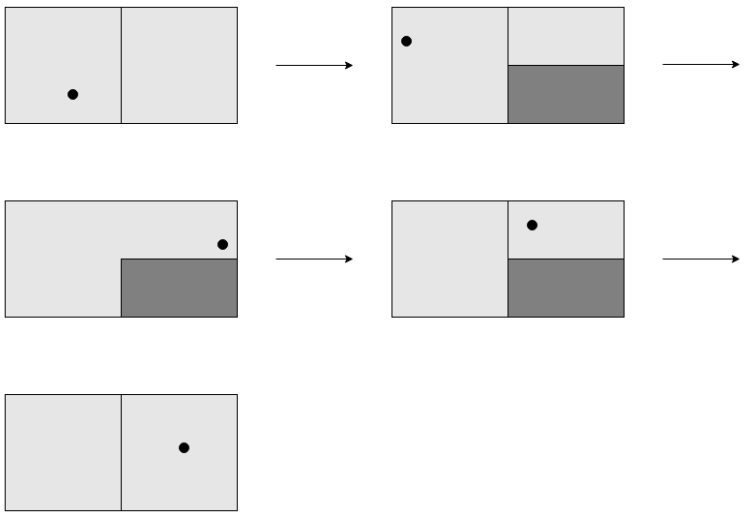


Figure 2.1: The sequence of events leading to positive work extraction. Initially, the input box has a particle on the left side of the partition, indicating a "0" bit. A step is then raised on the right side of the partition at no work cost. The partition is removed and the particle relaxes to equilibrium, filling the available volume with uniform probability. The partition is then reinserted, trapping the particle on the right hand side with probability equal to the volume fraction of the right side. Finally, the step is lowered, extracting work in the process. In contrast, if the particle begins and ends on the same side the total cost/gain in work is zero, while if it begins on the right and ends on the left there is net loss.

These steps are displayed schematically in Fig. 2.1. Altogether, the average change in energy of the work reservoir during a single cycle of operation on an input site (steps 1-4) is

$$W_{total} = W_{out} - W_{in} \quad = (p_1^{eq} - p_1)kT \ln \frac{l}{l-h} \tag{2.93}$$

$$= (p_1^{eq} - p_1)kT \ln \frac{1}{1-y} \tag{2.94}$$

$$= (p_1^{eq} - p_1)E. \tag{2.95}$$

In the second line we set l to be 1 and dimensionless, and introduced y to denote h expressed in these units. In the third line we defined $E \equiv kT \ln \frac{1}{1-y}$ to be the work that is extracted when a bit is flipped from 0 to 1, i.e. when a particle initially found on the left side of the box is trapped on the right at the end of the protocol. Note that the equilibrium distribution p^{eq} is the distribution of bits on the output tape. Eq. 2.95 then reflects the same work extraction capability as the Mandal/Jarzynski demon [76].

The equilibrium probability of finding the particle on the right side of the box, p_1^{eq}, is simply the fraction of the box's volume that is taken up by the right side while the step is raised:

$$p_1^{eq} = \tfrac{l^2(l-h)}{l^2(l-h)+l^3} \quad = \frac{l-h}{2l-h} \tag{2.96}$$

$$= \frac{1-y}{2-y} \tag{2.97}$$

$$= \frac{e^{-E/kT}}{1+e^{-E/kT}}. \tag{2.98}$$

Our choice of E thus re-expresses p^{eq} as a Boltzmann distribution for an effectively two-state system with energy 0 on the left side of the box and energy E on the right hand side of the box. Substituting Eq. 2.98 into Eq. 2.95, we can re-express the net work entirely in terms of the initial probability of residing on the right side p_1, the temperature of the environment T, and E:

$$W(E) = E\big[\frac{e^{-E/kT}}{1+e^{-E/kT}} - p_1\big]. \tag{2.99}$$

We show a plot of this function for various values of p_1 in Fig. 2.2. In Fig. 2.3 we show the maximum work that can be extracted as a function of p_1, as well as the associated change in entropy of the particle over the cycle. The maximum work must be calculated numerically, because setting the derivative of Eq. 2.99 equal to 0 results in a transcendental equation.

Additionally, we may consider the cost of resetting the tape, so that it may be fed back into

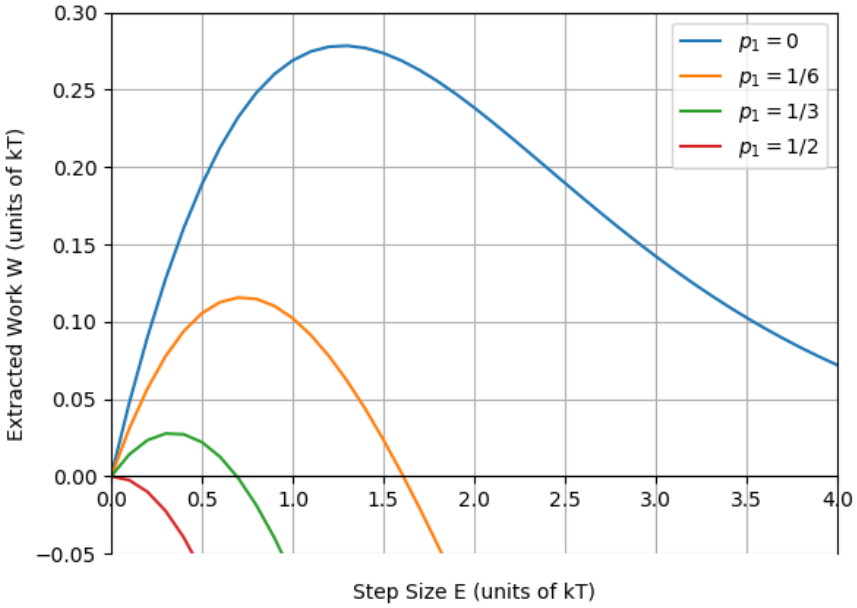


Figure 2.2: Plots of the net gained work W as a function of E (equation 2.99) for various values of p_1, starting at the top with $p_1 = 0$ and increasing in equally sized steps to $p_1 = 1/2$. Note that when $p_1 \geq 1/2$ net work extraction becomes impossible, as is demanded by the Second Law. kT has been set here to 1. Note that having a finite step size means that the distribution at the end of the process can never reach $p_0^{eq} = p_1^{eq} = 1/2$, and thus the machine described here cannot extract all the available fuel in the tape.

the device. That is, we want to spend some work to convert the distribution of bits on the output tape from p^{eq} back to p. To accomplish this, we must raise a step of the appropriate size so that p becomes the equilibrium distribution. The procedure involves the following steps:

- Step 1: A step of height y_{reset} is raised on the right side. In this process $W_{in,reset}$ per input site is drawn from the work reservoir. y_{reset} is chosen so that p is the equilibrium distribution of the associated Hamiltonian.

- Step 2: The partition is removed, and the particle is given time to relax to its equilibrium distribution, which is p by design.

- Step 3: The partition is reinserted. The step is lowered, with associated addition to

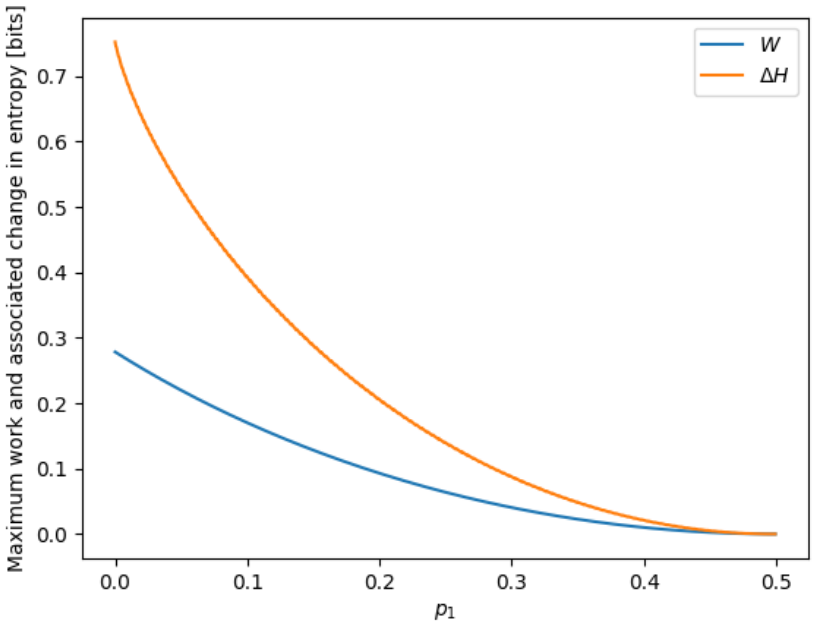


Figure 2.3: The maximum amount of work that can be extracted as a function of p_1 is shown in blue. The associated change in entropy $\Delta H = H(p^{eq}) - H(p)$ of the particle is shown in orange.

the work reservoir $W_{out,reset}$.

- Step 4: The reset box is returned to the input queue, and a newly outputted box is given to the reset device.

The total cost is given by the sum of the contributions from Step 1 and Step 3. Since y_{reset} is chosen so that p is the equilibrium distribution, we have $p_1 = \frac{1-y_{reset}}{2-y_{reset}}$ (recall Eq. 2.97). Solving for y_{reset}, we have $y_{reset} = \frac{1-2p_1}{1-p_1}$. The total average work cost per site of this step is then

$$W_{in,reset} = p_1^{eq} kT \ln \frac{1}{1 - y_{reset}} \tag{2.100}$$

$$= p_1^{eq} kT \ln \frac{1 - p_1}{p_1}. \tag{2.101}$$

Note that if $p_1 > 1 - p_1$, i.e. $p_1 > 1/2$ then this would involve lowering the step and work extraction. However, as was discussed previously, only values of p_1 less than $1/2$ lead to positive work extraction by the device, so this reset scenario can be ignored. $W_{out,reset}$ is simply $W_{out,reset} = p_1 kT \ln \frac{1}{1-y_{reset}}$. The total cost of resetting is then

$$W_{total,reset} = W_{in,reset} - W_{out,reset} \tag{2.102}$$

$$= (p_1^{eq} - p_1)kT \ln \frac{1}{1 - y_{reset}} \tag{2.103}$$

$$= (p_1^{eq} - p_1)kT \ln \frac{1 - p_1}{p_1}. \tag{2.104}$$

If we imagine cyclic operation of the device, in which bits are continually reset and fed back in for work extraction, the total energy change of the work reservoir per input site is on average given by the difference of Eqs. 2.95 and 2.104:

$$W_{cycle} = W_{total} - W_{total,reset} \tag{2.105}$$

$$= (p_1^{eq} - p_1)kT\big[\ln \frac{1}{1 - y} - \ln \frac{1 - p_1}{p_1}\big] \tag{2.106}$$

$$= kT\big[\frac{1 - y}{2 - y} - p_1\big] \ln \frac{p_1}{(1 - y)(1 - p_1)}. \tag{2.107}$$

The Second Law demands that this quantity must be negative for the regime $p_1^{eq} = \frac{1-y}{2-y} > p_1$ that leads to positive work extraction when the reset step is ignored. In this regime we have $p_1 < \frac{1-y}{2-y}$ and from this relation also that $\frac{1}{1-p_1} < 2 - y$. Substituting each of these inequalities into the argument of the log in Eq. 2.107 establishes that it is less than 1 and therefore that the log is negative.

It is worth noting that given the same sequence of input boxes, what we have described is not the optimal way to extract work. One could instead simply expand the partition to the right each time, by some fixed distance less than the full length of the box (to prevent the

pressure from diverging when the particle is on the wrong side). A simple calculation reveals that the optimal fraction of the right-side box to expand to is $1-2p_1$, and that this yields on average an extracted amount of work equal to $kT\big[\ln(2)-H(p)\big]$, where $H(p)$ is the Shannon entropy of the input distribution [98]. Fig. 2.3 reveals that this limit is not achievable for the earlier device we have described. The reason for the difference is that the work extraction method we have described relies on harnessing thermal fluctuations (transitioning from one side of the partition to the other), whereas a traditional Szilard engine does not. Thus, the mechanism of operation is fundamentally different, and it is this harnessing of fluctuations that serves as the basis for the operation of the device described in Ch. 4.

CHAPTER 3
NON-EQUILIBRIUM ABUNDANCES FOR THE BUILDING BLOCKS OF LIFE

The difficulty of obtaining appreciable quantities of biologically important molecules in thermodynamic equilibrium has long been identified as an obstacle to life's emergence, and determining the specific non-equilibrium conditions that might have given rise to life is challenging. To address these issues, we investigate[1] how the concentrations of life's building blocks change as a function of the distance from equilibrium *on average*, in two example settings: (i) the synthesis of heavy amino acids, and (ii) their polymerization into peptides. We find that relative concentrations of the heaviest amino acids can be boosted by four orders of magnitude, and concentrations of the longest peptide chains can be increased by hundreds of orders of magnitude. The average non-equilibrium distribution does not depend on the details of how the system was driven from equilibrium, indicating that environments might not have to be fine-tuned to support life.

3.1 The non-equilibrium model

Statistical mechanics tells us that we do not need to describe the full microscopic state of a system in order to predict macroscopic characteristics, as those are understood as expectation values, or ensemble averages. Therefore, all we need to infer is the *probability*, ρ_i, of every state, $i = 1, \dots N$. This is a hard problem, as we have only a handful of constraints, namely measured average quantities, together with normalization of probability. Say we have M constraints. Then we are still lacking $N - M$ equations to determine the ρ_i. These probabilities can be assigned by choosing the probability distribution with the largest entropy, $S[\rho] \equiv -\sum_{i=1}^{N} \rho_i \ln(\rho_i)$, subject to the constraints imposed by the system's bulk properties [60]. This maximization of entropy can be interpreted as choosing a model that makes use of only the information provided by the measured quantities [60, 61, 46], ensuring that information that we do not actually have is not falsely being ascribed to the system. This powerful inference tool has been applied successfully to many other problems in a diverse range of fields from ecology to neuroscience, and is commonly known under the name of MaxEnt [106, 64]. In statistical physics, we find that under the constraint that

[1] The contents of this chapter are currently in review at Physical Review E. The manuscript was coauthored with Susanne Still

only the average energy is known, the Boltzmann distribution, describing thermodynamic equilibrium states, $\rho = \frac{1}{Z}e^{-E/kT}$, is recovered by this MaxEnt inference method [60]. Here, the temperature is denoted by T, Bolzmann's constant by k, and normalization is ensured by the partition function, Z, which is related to the equilibrium free energy, F, by $Z = e^{-F/kT}$.

On the early Earth, conditions governing the processes preceding life were *not* consistently in thermodynamic equilibrium. It is much harder to infer the distribution, θ, of a system that is away from thermodynamic equilibrium without detailed information. The distribution can no longer be inferred straight from a MaxEnt argument, and information is lacking to make up for the missing equations. Without specific knowledge about some particular process generating biomolecules on early Earth, little can be done.

Here, we propose to calculate instead the *average* non-equilibrium distribution. The idea is that there are many diverse environments on Earth and a large variety of energy sources that act as non-equilibrium drives. If all we are interested in are the expected abundances we would get somewhere on Earth, then we can average out details of the non-equilibrium driving. We do so, following [31], by giving probability distributions a weight, i.e. we will assume that there is a *distribution over distributions*, $P(\theta)$, and compute the average: $\langle\theta\rangle = \int \theta P(\theta) d\theta$.

For our purposes, we need only consider distributions on a discrete state space. We will compare the probability of finding the building blocks of life as computed from this average non-equilibrium distribution to that computed from the equilibrium distribution for two biologically relevant model systems in the following sections. Clearly, the answer will depend on the probabilities assigned to different non-equilibrium probability distributions, $P(\theta)$. Crooks suggested [31] to find $P(\theta)$ by maximizing the entropy of the distribution over distributions subject to physical constraints, in analogy to what is done in equilibrium [60]. In the absence of additional information, this maximum entropy approach ought to best describe the ensemble of non-equilibrium distributions, as it ensures that only available information is included in the description.

We elaborate on the details of Crooks' approach in the Appendix, and mention here only the resulting formula:

$$\langle\theta\rangle = \frac{1}{\mathcal{Z}(\beta,\lambda)} \int \theta e^{-\lambda D(\theta\|\rho)} d\theta. \tag{3.1}$$

The normalization constant, $\mathcal{Z}(\beta,\lambda)$ depends on the inverse temperature, $\beta = 1/kT$, where k denotes the Boltzmann constant. The factor $e^{-\lambda D(\theta\|\rho)}$ determines the weight given to each distribution θ. It is controlled by the product of the distribution-independent parameter $\lambda \geq 0$, and the relative entropy $D(\theta||\rho)$ between the non-equilibrium distribution, θ, in

question, and the corresponding equilibrium distribution ρ:

$$D(\theta||\rho) = \sum_i \theta_i \ln \left[\frac{\theta_i}{\rho_i}\right]. \tag{3.2}$$

A system away from thermodynamic equilibrium can contain free energy in excess of the corresponding equilibrium system. This additional free energy is given by $kTD(\theta||\rho)$ [103, 115, 111].

The second law of thermodynamics implies that the work input to a system is always greater than or equal to the corresponding change in free energy, and so this formalism assigns higher probabilities to distributions that require a lower minimum amount of work to create. At a fixed value of the parameter λ, a non-equilibrium distribution is thus more likely to occur, if less work is needed to produce it. Relative entropy also measures the coding cost encountered when the canonical distribution ρ is used as a model for θ [71, 29]. Relative entropy is thus both a physically and an information-theoretically meaningful measure for deviation from equilibrium. In equilibrium, it is the free energy difference between reactants and products alone that sets their relative abundances. Thus, a natural measure for the difficulty of creating a molecule is its free energy of formation. The hyperensemble extends this notion, in a sense, to the situation away from equilibrium.

We stress that using this approach does not imply that possible path-dependences of the non-equilibrium states are being neglected; they may very well retain some memory of their history. Each system in the ensemble is driven to a non-equilibrium distribution in a path-dependent way, as the arrival at a distribution, in general, depends on the trajectory generated by the drive.

In the limit $\lambda \to \infty$, all non-equilibrium distributions will have negligible probability, and the average non-equilibrium distribution converges to the equilibrium distribution: $\langle\theta\rangle = \rho$. For finite values of λ, the distribution $\langle\theta\rangle$ is in general flatter than its equilibrium counterpart, thereby augmenting the probabilities of states that would otherwise be rare [31]. This is most apparent in the limit $\lambda \to 0$. In that case, all distributions become equally likely In that sense, λ encodes the extend to which driving conditions can push the system out of equilibrium. In the most extreme out of equilibrium limit ($\lambda \to 0$), the average distribution over a finite state-space has, by symmetry, equal probabilities for every state. The overall flattening effect would persist if we were to replace the measure $D(\theta||\rho)$ with any other function (see the appendix for details). Conclusions we draw for the extreme nonequilibrium limit ($\lambda \to 0$) in Section 3.2, are therefore invariant with respect to how distance from

equilibrium is measured.

The extreme non-equilibrium limit is different from the high temperature limit of an equilibrium distribution, because the free energies of the molecules are themselves temperature dependent, and so the high temperature limit would not assign equal weight to every possible distribution of molecules. For example, polymerization of amino acids into long chains would generally be disfavored in the high temperature limit (compare Section 3.2.2).

Probabilities of rare states are only augmented on average. There are individual non-equilibrium systems that give rise to worse-than-equilibrium odds for forming the desired molecules. The non-equilibrium distributions describing those systems are included in the average. Individual distributions that exhibit large numbers of rare molecules are less probable at all finite values of λ, due to the exponential dependence on $D(\theta||\rho)$ (see Eq. (A.1). In what follows, it is the average non-equilibrium distribution $\langle\theta\rangle$, and not any particular non-equilibrium distribution θ, that we use for our analysis.

We interpret the average non-equilibrium distribution as describing the result that would be obtained if one took samples from a diverse collection of non-equilibrium environments, and averaged the concentrations of the various molecules found. The average non-equilibrium distribution provides the expected value, or best guess, for what we would find in a single sample, taken anywhere on the planet. In the context of molecules relevant for forming living structures, using the average non-equilibrium distribution to make an inference about relative abundances should be more appropriate than using the equilibrium distribution, because we know that conditions on early earth were not consistently in thermodynamic equilibrium.

The average non-equilibrium distribution does not depend on the details of any particular driving protocol, but rather on the *set* of driving protocols that generate the non-equilibrium systems in question. The set of local processes that could drive a system out of equilibrium on the Earth is extremely large and diverse, to the degree that the entropy over the set of possible distributions might be, to a good approximation, maximal. This would not necessarily be the case if, for example, the only process driving various systems on early Earth out of equilibrium was the rising and setting of the sun. That restriction would then impose additional constraints on our ensemble that would need to be taken into consideration, and we could not expect the maximization of entropy to sidestep those details. However, environments on the Earth permit a diversity of local processes. This inhomogeneity of conditions on early Earth supports the use of the maximum entropy hyperensemble, which allows us to compute averages without requiring any information beyond that captured by

the temperature T and the non-equilibrium parameter λ.

Let us now explore how the concentrations of large and complex molecules change as a function of the distance from equilibrium.

3.2 Results

3.2.1 Amino acid abundances and functional proteins

The possibility of prebiotic synthesis of amino acids was established in the landmark experiment by Miller and Urey [85]. They have since been detected in meteors [94], and produced in other experiments seeking to model the conditions of the early Earth [5, 82]. However, the abundances with which the amino acids appear in abiotic settings do not match their biotic abundances [37]. In particular, functional proteins tend to employ the various amino acids in roughly equal proportions [37, 1], whereas in abiotic sources there is an exponential suppression in the abundances of the larger amino acids, and none heavier than threonine have yet been found [51]. The apparent inability of the environment to produce heavier amino acids in sufficient quantities has been identified by several authors as a barrier to the emergence of life [1, 51, 21].

The difficulty of synthesizing the heavier amino acids in a prebiotic setting is usually ascribed to them having a larger Gibbs free energy of formation, ΔG [51]. The free energies of formation of the amino acids were calculated in [2], assuming synthesis from CO_2, NH_4^+, and H_2 in surface seawater at a temperature of $18°C$. The concentrations of amino acids relative to glycine, taken from 9 different data sets, were fit using an exponential function [51]:

$$C_{\text{rel}} = 15.8 * \exp\left[-\Delta G/31.3\right]. \tag{3.3}$$

We rescale these values so that they may be interpreted as probabilities (i.e. fraction of the total amino acid concentration occupied by amino acid x):

$$P(x) = \frac{C_{\text{rel}}(x)}{\sum_{i=1}^{N} C_{\text{rel}}(i)}\ , \tag{3.4}$$

where $C_{\text{rel}}(x)$ is the relative concentration of amino acid x, and the index $i = 1, \ldots, N$ runs over all measured amino acids. The exponential dependence of the probabilities on the free energy of formation ΔG is consistent with an equilibrium distribution [51], although we

caution that there are difficulties with this interpretation [94]. Nevertheless, we take Eq. (3.4) as our best approximation to the equilibrium distribution. We furthermore assume that this function correctly predicts the equilibrium abundances of the heavier amino acids which have not yet been found in abiotic sources, consistent with the fact that it predicts abundances too low to observe for these heavy amino acids [51].

We compare the distribution calculated from Eqs. (3.3) and (3.4) to the average non-equilibrium distribution, calculated numerically from Eq. (3.1). We assume that amino acids are the most thermodynamically costly molecules that can be formed in the system. This ought to be the case if the system is physically confined to a small volume (e.g. a mineral pore), or the reactants are very diluted. Such a restriction on the available state space is needed because in the extreme non-equilibrium limit, all states become equally probable. This means that if more costly molecules can be formed than amino acids, then the probabilities of forming any amino acids could go down relative to these more costly molecules. Yet, even without this restriction, the distribution of amino acids would become more uniform out of equilibrium. In the last section we will relax this assumption on the maximum cost of molecules, as we look at the asymptotic behavior of amino acids polymerizing into arbitrarily long chains.

Fig. 3.1 shows the probability of obtaining the rarest amino acid, tryptophan, as a function of λ. In the extreme non-equilibrium limit, the hyperensemble becomes a symmetric Dirichlet distribution, which, in a state space of dimension d, has an expectation value for each outcome of $1/d$ and a variance of $\frac{1}{d^2}\frac{(d-1)}{(d+1)}$ [6], meaning that the standard deviation is of the same order as the mean. For a state space of dimension $d = 20$, the relative concentration of tryptophan in the extreme non-equilibrium limit is then $5 * 10^{-2} \pm 4.8 * 10^{-2}$, while its equilibrium relative concentration is $\sim 6 * 10^{-6}$. On average, the relative concentrations of the rarest amino acid can be boosted by four orders of magnitude in the non-equilibrium regime. Fig. 4.1 shows a normalized histogram of samples of any amino acid, e.g. tryptophan, in the extreme non-equilibrium limit. While a significant fraction of samples end up close to their equilibrium values, the distribution as a whole gives radically more favorable odds of drawing a high concentration.

In Fig. 3.3, the average non-equilibrium distribution of amino acids is plotted as a function of the free energy of formation ΔG and compared to the equilibrium distribution, for various values of the non-equilibrium parameter λ, showing how the distribution becomes flatter as λ decreases. Importantly, the roughly uniform distribution of amino acids employed in functional proteins is exactly what the average non-equilibrium distribution gives

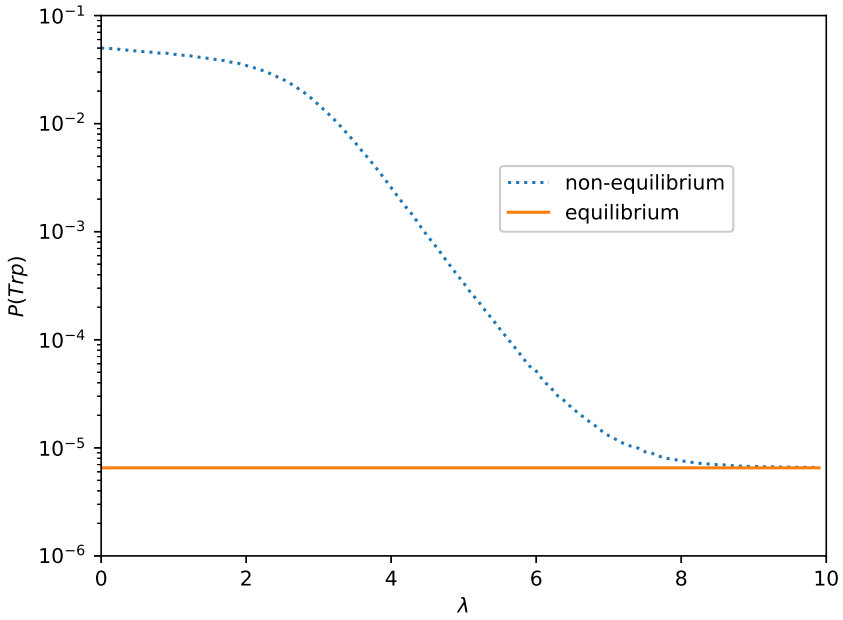


Figure 3.1: Tryptophan requires the largest free energy to form of the protein amino acids, and has not yet been found in an abiotic setting. Here we show how the relative concentration of tryptophan changes as one moves away from equilibrium, with the distance from equilibrium controlled by the parameter λ. The equilibrium probability is plotted with an orange solid line. The average non-equilibrium probability is plotted with a blue dotted line. Values are computed numerically from Eq. (3.1). We see that in the extreme non-equilibrium limit $\lambda \to 0$, the relative concentration of tryptophan can be increased up to four orders of magnitude.

in the extreme non-equilibrium regime (for values of λ close to zero). Thus, far away from equilibrium, the distribution of amino acids moves closer to its biotic distribution, thereby greatly enhancing the chances of spontaneously assembling functional proteins [122, 1].

3.2.2 Polymerization of amino acids

Amino acids may be linked with one another via the peptide bond to form long chains. These chains then fold into proteins, with a typical protein containing ~ 500 amino acids. However, the free energy, ΔG, for the peptide bond is on the order of several thousand kJ/mole [79], making the formation of long chains extremely improbable in thermodynamic equilibrium. It has been estimated that a solution containing 1 molar concentrations of each of the amino acids would require a volume 10^{50} times the size of the Earth to produce a single molecule

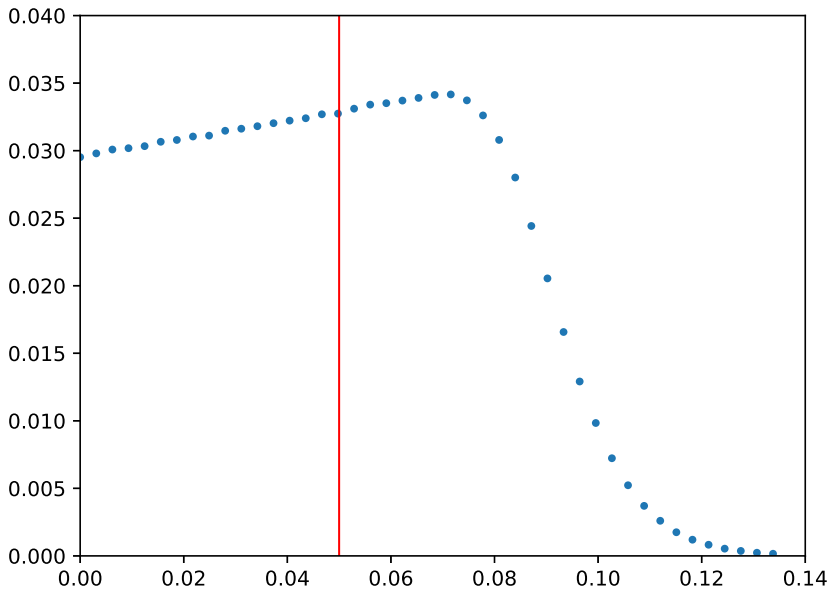


Figure 3.2: A normalized histogram of 10^7 samples of the relative concentration of any amino acid (e.g. tryptophan) from the hyperensemble in the extreme non-equilibrium limit. The red line indicates the mean value. We have confirmed numerically that there are an equal number of samples above and below the mean. On this scale, the equilibrium relative concentration of tryptophan, at $\sim 6 * 10^{-6}$, would not be distinguishable from the y-axis.

of protein in equilibrium [36].

The thermodynamics of polymerization of amino acids were explored in [79], where, for simplicity, the chains were assumed to consist entirely of glycine. It was found that dimerization of two glycine molecules requires the greatest amount of free energy per bond ($\Delta G = 3.6$ kcal/mole), being about eight times more difficult to form than subsequent additions to the chain. The relative concentration $[GG]/[G]$ is predicted to be about $1/400$ in equilibrium, and each subsequent addition of a glycine to the peptide results in a decrease by a factor of $1/50$ [79]. The probability of getting a chain of length $l \geq 2$ then follows a power-law

$$P_{\text{eq}}(l) \propto \left(\frac{1}{50}\right)^{l-2} \tag{3.5}$$

with the proportionality constant set by normalization of the probability. We examine

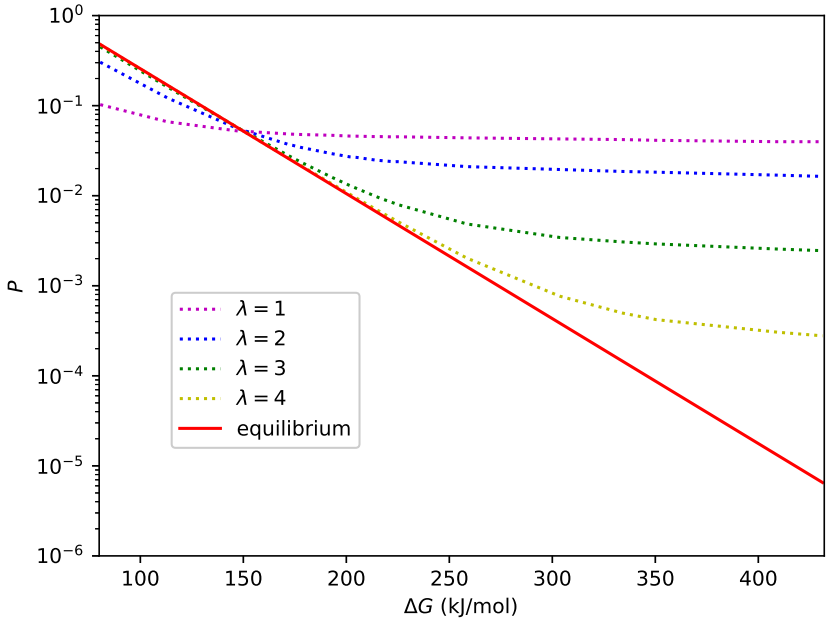


Figure 3.3: Probability distribution of amino acids, arranged on the x-axis in order of increasing Gibbs free energy, ΔG. The relative probability of formation in thermodynamic equilibrium is given by Eq. (3.4) and plotted in red. The other curves are the average non-equilibrium distribution, computed numerically from Eq. (3.1), at different distances from equilibrium (i.e. different values of λ). Note that as the distance from equilibrium increases, i.e. λ gets smaller, the distribution becomes flatter, and the probabilities of forming the rarest amino acids increase by several orders of magnitude. The flatter distribution observed out of equilibrium is consistent with the fact that roughly equal numbers of amino acids are found in functional proteins, and thus boosts the odds of forming them.

the change in this distribution for non-equilibrium systems. To proceed, we identify each macrostate of a solution containing N glycine molecules with a partition of the number N into a sum of positive integers. For example, in a solution containing $N = 3$ glycine molecules there are three possibilities: the solution could either contain three monomers (corresponding to 1+1+1), or could contain one monomer and one dimer (1+2), or one trimer. In number theory, the partition function, which we denote here by $Q(N)$, counts the number of distinct ways that a positive integer N can be decomposed into a sum of positive integers. For example, $Q(N = 3) = 3$. For tractability, we consider in this section only the extreme non-equilibrium limit $\lambda \to 0$, where all partitions of N become equally likely. First, we examine the probability of the rarest state, in which all N glycine molecules become bound into one chain of length $l = N$. The probability of observing this state is

$P(l = N) = 1/Q(N)$. For large N, we can estimate $P(l = N)$ using the Hardy-Ramanujan asymptotic expression for $Q(N)$ [3], giving us

$$P_{\text{neq}}(l = N) \approx 4N\sqrt{3} * e^{-\pi\sqrt{\frac{2N}{3}}}. \tag{3.6}$$

Far away from equilibrium, the maximum probability of the rarest state is a decreasing function of N. Yet the odds of finding all N particles bound into a single chain decrease much more rapidly in equilibrium (refer to Eq. (3.5)), meaning that as the system gets larger, the factor by which non-equilibrium driving enhances probabilities of the rarest states grows without bound. This effect radically augments the chances of forming proteins in an abiotic setting. We display the ratio $P_{\text{neq}}(l)/P_{\text{eq}}(l)$ in Fig. 3.4, computed from Eqs. (3.5) and (3.6) using an exact expression for $P_{\text{neq}}(l)$ obtained from SageMath's built-in Partitions function. With only 100 glycine molecules, the chances of finding them all bound into a single chain is found to be more than 100 orders of magnitude greater out of equilibrium than in equilibrium, and this effect will continue to become more dramatic as the number of molecules in the system increases.

Of interest is also the number of chains of each possible length l, which we denote by m_l. When every partition is equally likely, the average number of chains of length l is given by [25, 56]

$$\langle m_l \rangle = \frac{1}{Q(N)} \sum_{n=1}^{\text{floor(N/l)}} Q(N - nl). \tag{3.7}$$

This distribution was previously studied in the context of a fragmentation process, e.g. where a nucleus is broken apart and each partition is equally likely [56, 87, 84, 74, 25, 17]. We calculate the expected number of chains of length l, $\langle m_l \rangle$, numerically for a system of size $N = 100$ and compare to that computed from the equilibrium distribution, Eq. (3.5). The results are displayed in Fig. 3.5. The numbers of chains of all lengths are increased dramatically, whereas in equilibrium most molecules would remain unbound to one another.

When N is large and the chains are not too long relative to N, Eq. (3.7) is well approximated by [25]

$$\langle m_l \rangle \approx \frac{1}{\exp\left[\sqrt{\frac{\pi^2}{6N}}l\right] - 1} \tag{3.8}$$

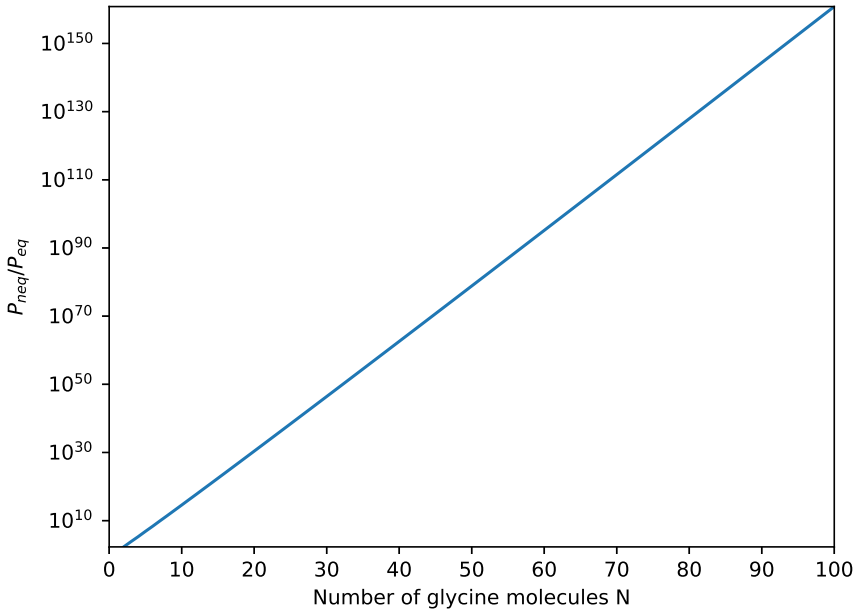


Figure 3.4: Glycine molecules can be linked together via a peptide bond to form chains. Due to the large amount of free energy required per bond, the concentrations of longer chains drop precipitously in thermodynamic equilibrium (Eq. (3.5)). Here we consider a system of N glycine molecules, and compare the probability of finding all of them bound into a single long chain, in thermodynamic equilibrium (P_{eq}) to that far away from thermodynamic equilibrium (P_{neq}), in the extreme non-equilibrium limit (given approximately by Eq. (3.6) but using exact values here). We plot the ratio P_{neq}/P_{eq} as a function of N, and see an exponential increase. This effect helps to explain how amino acids can be spontaneously linked together to form proteins in an abiotic setting.

which again will drop off much more slowly than the equilibrium distribution. Overall, this means that in the extreme non-equilibrium limit, the abundances of long peptide chains, and therefore proteins, can be increased by hundreds of orders of magnitude. This shows that obtaining appreciable quantities of proteins on the early Earth, which is all but excluded in equilibrium statistical mechanics, is a viable possibility considering the average odds out of equilibrium.

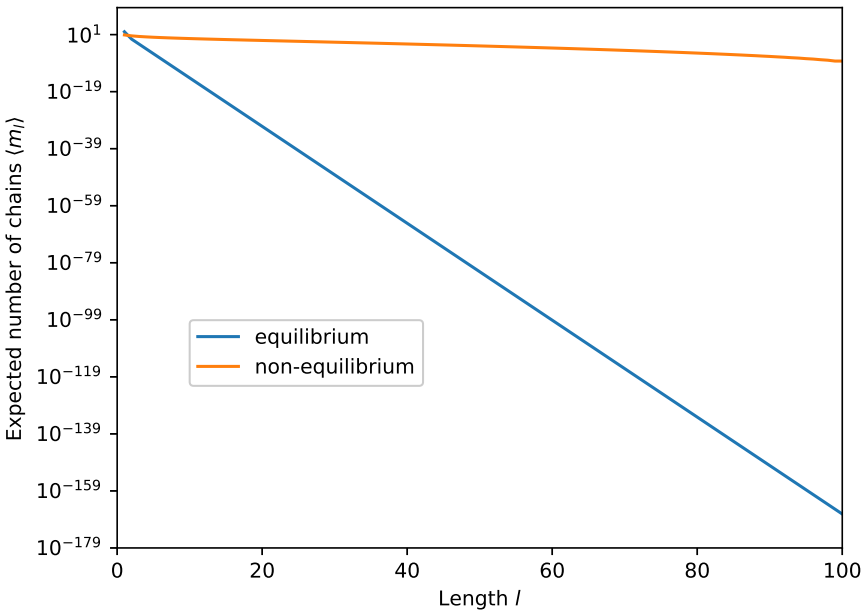


Figure 3.5: The expected number of chains of length l in the extreme non-equilibrium limit is given by Eq. (3.7) and plotted in orange for a system of size $N = 100$. We compare it to the values computed from the equilibrium distribution, given by Eq. (3.5) (plotted in blue). In equilibrium, long chains are suppressed exponentially. This is not the case far away of equilibrium, where concentrations of the longest chains are increased by hundreds of orders of magnitude.

CHAPTER 4
PHYSICAL LIMITATIONS OF WORK EXTRACTION FROM TEMPORAL CORRELATIONS

Recently proposed information-exploiting systems extract work from a single heat bath using temporal correlations on an input tape. We study how enforcing time-continuous dynamics, which is necessary to ensure the device is physically realizable, constrains possible designs and drastically diminishes efficiency. We show that these problems can be circumvented by means of applying an external, time-varying protocol, which turns the device from a "passive", free-running machine into an "actively" driven one[1]. We begin by introducing the model.

4.1 Model of a temporal correlation powered work extraction device.

This model largely follows [19]. Imagine a work extraction system with two internal states, $s \in \{A, B\}$, which can be coupled to and decoupled from a work reservoir (such as a weight), an input tape with bits $b^{in} \in \{0, 1\}$, an output tape with $b^{out} \in \{0, 1\}$, and a heat bath. The joint state-input value of the coupled system is then $(s, b) \in \{A, B\} \times \{0, 1\} = \{A0, A1, B0, B1\}$, where b denotes a coupled bit. Each of these four joint states possesses a potential energy, E_i, $i \in \{A0, A1, B0, B1\}$. The dynamics are described by a time-dependent vector containing the probabilities that the system is in one of the joint states at a given time, $\boldsymbol{p}_{sb} \equiv [p_{A0}, p_{A1}, p_{B0}, p_{B1}]^\top$ ($^\top$ denotes the transpose).

The engine alternates between an "interaction step", during which the internal state interacts with a bit, and a "switching step", during which the bit is changed. Any changes in energy during an interaction step are due to the exchange of heat with the heat bath, and changes in energy during a switching step are due to exchange of work with the work reservoir. We therefore talk about "heat" steps and "work" steps, and will use these words to label transformations, as a reminder.

An interaction step is represented by the transformation $\boldsymbol{p}_{sb} \xrightarrow{\text{heat}} \boldsymbol{M}\boldsymbol{p}_{sb}$, where the joint state evolves under the action of a reversible, column stochastic matrix, $\boldsymbol{M}$, for a duration

[1]The contents of this chapter are in press with Physical Review E. The paper is coauthored with Susanne Still, Thomas E. Ouldridge, and Lee Altenberg.

of time, τ. During interaction steps, the system undergoes a free-running relaxation towards equilibrium. Thus, $\boldsymbol{M}$ must be reversible, and the stationary distribution, $\boldsymbol{\rho}^{(\boldsymbol{M})} = \boldsymbol{M}\boldsymbol{\rho}^{(\boldsymbol{M})}$, satisfies detailed balance

$$M_{ij}\rho_j^{(\boldsymbol{M})} = M_{ji}\rho_i^{(\boldsymbol{M})} \tag{4.1}$$

for all i, j in $\{A0, A1, B0, B1\}$. The coupled device and bit relax towards thermodynamic equilibrium, described by the Boltzmann distribution $\rho_i^{(\boldsymbol{M})} = e^{-E_i/kT}/Z$. For simplicity, we choose units such that kT=1 and set the energy scale so that Z=1. We can then write the energy of each joint state as $E_i = -\ln \rho_i^{(\boldsymbol{M})}$, with $i \in \{A0, A1, B0, B1\}$.

In a switching step, the internal state is held fixed, and a new bit is coupled to the device. Whichever bit comprised the joint state prior to the switching step is printed to the output tape. That is, switching from the machine's n-th cycle to cycle $n+1$ changes the state of the bit that is interacting with the machine from $b = b_n^{out}$ to $b = b_{n+1}^{in}$. Depending on whether the input is 0 or 1, switching corresponds to the following transformation of the joint state probability vector:

$$\boldsymbol{p}_{sb} = [p_{A0}, p_{A1}, p_{B0}, p_{B1}]^\top \xrightarrow{\text{input 0}} \bar{\boldsymbol{p}}_{sb} = [p_{A0}{+}p_{A1}, 0, p_{B0}{+}p_{B1}, 0]^\top \equiv \boldsymbol{F}_0\boldsymbol{p}_{sb}, \tag{4.2}$$

$$\boldsymbol{p}_{sb} = [p_{A0}, p_{A1}, p_{B0}, p_{B1}]^\top \xrightarrow{\text{input 1}} \bar{\boldsymbol{p}}_{sb} = [0, p_{A0}{+}p_{A1}, 0, p_{B0}{+}p_{B1}]^\top \equiv \boldsymbol{F}_1\boldsymbol{p}_{sb}, \tag{4.3}$$

where the matrices $\boldsymbol{F}_0$ and $\boldsymbol{F}_1$ represent the switching:

$$\boldsymbol{F}_0 = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \qquad \text{and} \qquad \boldsymbol{F}_1 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}.$$

To extract work, the machine must on average raise the energy of the joint (s, b)-state during interaction steps (i.e. absorb heat), and lower the energy during switching steps (i.e. deposit energy into the work reservoir).

In the following, we limit our analysis for simplicity to an input tape consisting of alternating 1s and 0s. This is an interesting special case, because the per-symbol entropy of the input tape is maximal, as $\mathsf{Prob}(b^{in}=0) = \mathsf{Prob}(b^{in}=1) = \frac{1}{2}$, and hence cannot be leveraged for work extraction. Any net gain is thus due to exploiting temporal correlations.

A single complete cycle of operation is defined by the product of transition matrices $\boldsymbol{C} = \boldsymbol{M}\boldsymbol{F}_0\boldsymbol{M}\boldsymbol{F}_1$ (reflecting the alternating switching and interaction steps), taking the probability distribution of the four states from $\boldsymbol{p}_{sb}^{(n)}$ to $\boldsymbol{p}_{sb}^{(n+2)} = \boldsymbol{C}\boldsymbol{p}_{sb}^{(n)}$. We require that repeated application of the matrix $\boldsymbol{C}$ to any starting distribution $\boldsymbol{p}_{sb}^{(0)}$ converges to a steady state distribution $\boldsymbol{\pi}_0(s,b)$, defined by $\boldsymbol{C}\boldsymbol{\pi}_0(s,b) = \boldsymbol{\pi}_0(s,b)$. Thus $\boldsymbol{C}$ must be a primitive matrix (irreducible and aperiodic), which is assured if $M_{ij} > 0$ for all $i, j \in \{A0, A1, B0, B1\}$. We can define a steady state distribution if we census the system at each step of the cycle ($\boldsymbol{\pi}_1$ to $\boldsymbol{\pi}_3$ in the equation array below). Starting with feeding in a 1, a cycle is then characterized by the following changes (we use b' to denote a bit that is about to be transferred to the output tape):

$$\boldsymbol{\pi}_0(s_{n-1}, b'_{n-1}) \xrightarrow{\text{work}} \boldsymbol{\pi}_1(s_{n-1}, b_n) = \boldsymbol{F}_1\boldsymbol{\pi}_0(s_{n-1}, b'_{n-1}) \tag{4.4}$$

$$\boldsymbol{\pi}_1(s_{n-1}, b_n) \xrightarrow{\text{heat}} \boldsymbol{\pi}_2(s_n, b'_n) = \boldsymbol{M}\boldsymbol{\pi}_1(s_{n-1}, b_n) \tag{4.5}$$

$$\boldsymbol{\pi}_2(s_n, b'_n) \xrightarrow{\text{work}} \boldsymbol{\pi}_3(s_n, b_{n+1}) = \boldsymbol{F}_0\boldsymbol{\pi}_2(s_n, b'_n) \tag{4.6}$$

$$\boldsymbol{\pi}_3(s_n, b_{n+1}) \xrightarrow{\text{heat}} \boldsymbol{\pi}_0(s_{n+1}, b'_{n+1}) = \boldsymbol{M}\boldsymbol{\pi}_3(s_n, b_{n+1}) \tag{4.7}$$

The average work supplied to the work reservoir per input symbol is given by the sum of the average energy changes in the two switching steps: $\langle W \rangle = -\frac{1}{2}\big[\langle E \rangle_{\boldsymbol{\pi}_0(s_{n-1}, b'_{n-1})} - \langle E \rangle_{\boldsymbol{\pi}_1(s_{n-1}, b_n)} + \langle E \rangle_{\boldsymbol{\pi}_2(s_n, b'_n)} - \langle E \rangle_{\boldsymbol{\pi}_3(s_n, b_{n+1})}\big]$. The factor of 1/2 is due to two bits being encountered per cycle [2].

4.2 Work extraction by time-continuous, free-running devices

We now depart from the approach of [19]. To be physically realizable in the absence of externally applied driving during the interaction period, the matrix $\boldsymbol{M}$ should correspond with a continuous-time equilibration process for some time τ. In other words, we require there be a generator, $\boldsymbol{G}$, such that $\boldsymbol{M} = e^{\tau \boldsymbol{G}}$, where $\boldsymbol{G}$ is a reversible rate matrix with non-negative off-diagonal elements, in which every column sums to 0. If $\boldsymbol{M}$ can be constructed in this way, then $\boldsymbol{M}$ is said to be "reversibly embeddable" [63]. The following results from [63] are crucial: 1) If $\boldsymbol{M}$ is reversible, then $\boldsymbol{M}$ is diagonalizable and the eigenvalues of $\boldsymbol{M}$

[2] Note that this is work *extracted* from the joint (s, b) system. By convention, work done on the system is positive, as is heat flowing into the system.

are real. 2) If $\boldsymbol{M}$ is also embeddable, then the eigenvalues of $\boldsymbol{M}$ are all positive, and the generator, $\boldsymbol{G}$, of $\boldsymbol{M}$ is unique.

Due to these properties, processes governed by reversible and embeddable transition matrices generally extract much less work than those governed by reversible but not embeddable ones, as we will see shortly. In the rest of this paper, we will only be considering matrices that are reversible. For brevity, we henceforth use the terms *embeddable* and *non-embeddable* to refer to the two different classes of matrices. In Fig. 4.1, we display histograms showing the number of randomly generated matrices that achieve various values of positive work, for the two categories. The procedure used to make these matrices is detailed in Appendix B. The best randomly found embeddable matrices extract roughly a factor 20 less work than the best randomly found non-embeddable ones. In comparison, the construction given in Figure 6 of [19], which is non-embeddable, extracts $\frac{kT}{e} \approx 0.368\ kT$ of work per input symbol. We see from the histogram that finding a machine with a comparable efficiency by chance is rather unlikely.

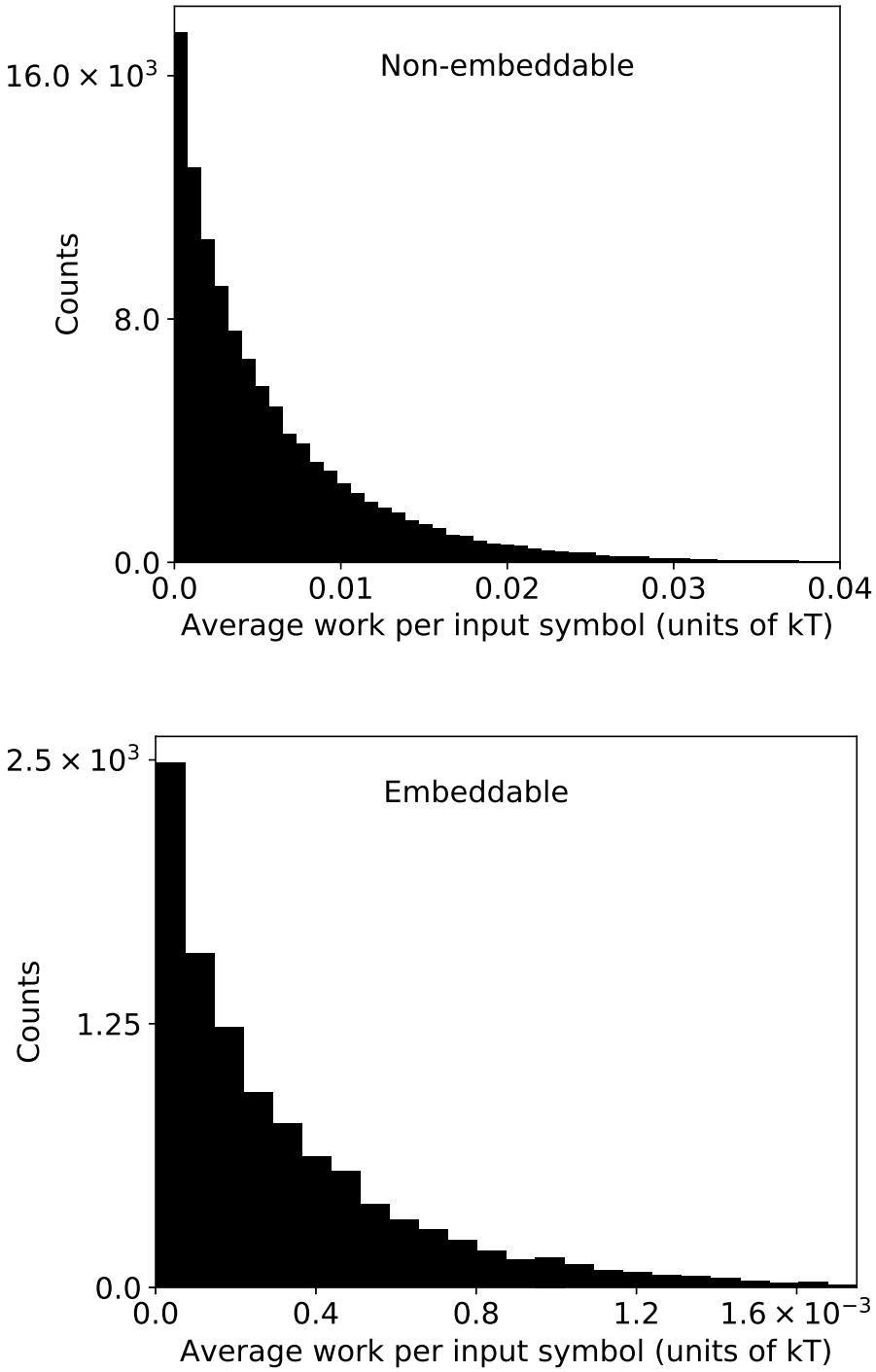


Figure 4.1: Histogram of work, W_{out}, extracted by randomly generated reversible transition matrices. Only positive work values are shown. With 10^6 randomly generated matrices of each type, 11% of non-embeddable matrices and 0.3% of embeddable ones achieved positive work production.

To improve upon random sampling, we constructed an evolutionary algorithm to explore the search space. The algorithm applied mutations to individual machine dynamics and fixed those mutations whenever they led to improved performance. Due to the high dimension of the space of transition matrices, this algorithm performed better than a grid search. When ignoring the embeddability constraint, the evolutionary algorithm readily returned the best design of [19], and never found one better. However, enforcing embeddability lowered the efficiency drastically. The best embeddable design which the evolutionary algorithm found is shown in Fig. 4.2. It achieved only $\approx 0.0146\ kT$ per input symbol. That is less than 4% of the yield of the non-embeddable best design of [19].

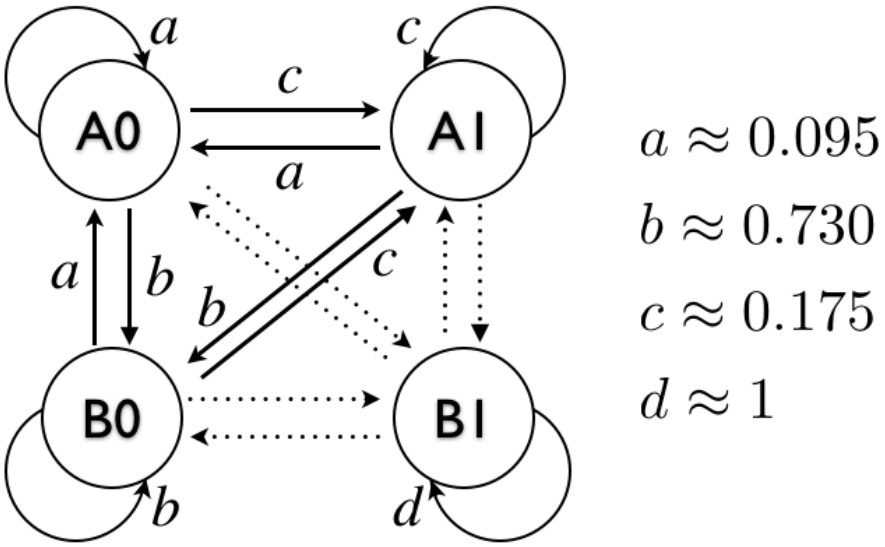


Figure 4.2: Design of the best embeddable transition matrix found by the evolutionary algorithm, with $W_{out} \approx 0.0146\ kT$. Dotted arrows denote transition probabilities close to zero. The transition matrix for this graph has a rank-1 submatrix, a second eigenvalue very close to 1, and the two smallest eigenvalues very close to zero.

We now discuss why the performance of embeddable designs is so poor. Optimal performance requires that the internal state of the device, s_n, contain predictive information [3], $I(s_n; b_{n+1})$, about the next incoming bit, b_{n+1}. To see why, note that the non-equilibrium free energy of system coupled to bit, $F[\pi] = \langle E \rangle_\pi - kTH[\pi]$, cannot increase spontaneously during an interaction step. Thereby, the heat absorbed in one interaction step is upper bounded by the entropy change, which can be written as $\langle Q_n \rangle \leq kT\big[I(s_n; b_{n+1}) - I(s_{n+1}; b'_{n+1}) + H(s_{n+1}) - H(s_n) + H(b'_{n+1}) - H(b_{n+1})\big]$. Adding two of these heat contributions to account

[3] We detail in Appendix A how to calculate information.

for the full cycle of Eqs. (4.4-4.7), we get a cancelation, because $H(s_{n+1}) - H(s_{n-1}) = 0$, due to the fact that we are in the same stationary distribution $\boldsymbol{\pi}_0$ at the beginning and end of the cycle.

Taking the average then sets an upper bound on the extractable work per input symbol, W_{out}. The bound depends on how the average predictive information about the input, $I_{\text{pred}} = (I(s_{n-1}; b_n) + I(s_n; b_{n+1}))/2$ compares to the average memory about the output, $I_{\text{mem}} = (I(s_n; b'_n) + I(s_{n+1}; b'_{n+1}))/2$, and on how the average output entropy, $H_B^{\text{out}} = (H(b'_n) + H(b'_{n+1}))/2$ compares to the average input entropy, $H_B^{\text{in}} = (H(b_n) + H(b_{n+1}))/2$:

$$W_{\text{out}} \leq kT\big[I_{\text{pred}} - I_{\text{mem}} + \Delta H_B\big] \ , \tag{4.8}$$

where $\Delta H_B \equiv H_B^{\text{out}} - H_B^{\text{in}}$. We display W_{out} as a function of I_{pred} in Fig. 4.3, for each of the two classes of transition matrices. It is clear from the plot that predictive information between device and next incoming bit is severely limited for embeddable systems, and that there is a consequent reduction in extractable work.

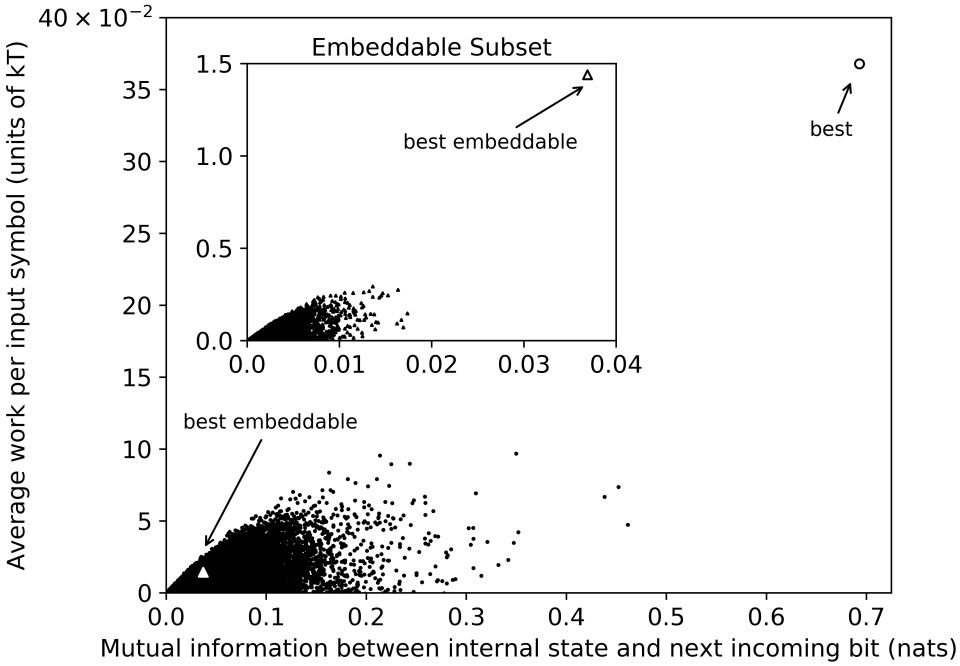


Figure 4.3: Average work per input symbol, W_{out}, vs. average predictive information per symbol. The triangle denotes the best embeddable design found by the evolutionary algorithm. The circle denotes the design of [19]. The inset shows only the subset of embeddable designs. See Appendix B for further details.

Perfect prediction of the next incoming bit requires synchronization with the input. For a period-2 input, the internal state must change in each interaction step from A to B or vice versa, necessitating a bipartite graph structure, whose associated transition matrix has negative eigenvalues [44], and therefore is non-embeddable. More generally, tracking a periodic signal with period n requires n eigenvalues of magnitude 1 equally spaced around the unit circle in the complex plane [44]. For $n > 2$ some of these eigenvalues must therefore be complex, which is impossible for any reversible matrix, embeddable or otherwise.

Synchronization is hampered by the tendency of embeddable systems to undergo "self-transitions", in which the system starts and ends in the *same* state during an interaction interval. These self-transitions are also undesirable because they are associated with no net exchange of energy with the bath, thus wasting the input. Self-transitions arise from the diagonal entries in $\boldsymbol{M}$, which can be set to zero for non-embeddable $\boldsymbol{M}$ [19], but not if $\boldsymbol{M}$ is embeddable. To see why, note that $\boldsymbol{M}$ being a stochastic matrix implies that it has an eigenvalue of 1, and $\boldsymbol{M}$ being embeddable implies that all other eigenvalues are positive. Thus, the trace of $\boldsymbol{M}$ must be greater than 1 and self-transitions cannot be neglected. The average fraction of such self-transitions in interaction steps can be written as

$$\frac{1}{2}\boldsymbol{d}(\boldsymbol{M})^{\top}[\boldsymbol{\pi}_1(s_{n-1}, b_n) + \boldsymbol{\pi}_3(s_n, b_{n+1})], \tag{4.9}$$

where $\boldsymbol{d}(\boldsymbol{M})$ is a vector of the diagonal elements of $\boldsymbol{M}$. We have found numerically that self-transitions occur on average at a minimum of 1/4 of interaction steps. We provide in Appendix C an example of a matrix satisfying this property. Note that this feature implies that there do not exist any embeddable matrices "close" to the non-embeddable design of [19]. The relationship between self-transition rate and average work extracted for randomly chosen embeddable and non-embeddable designs is shown in Fig. 4.4.

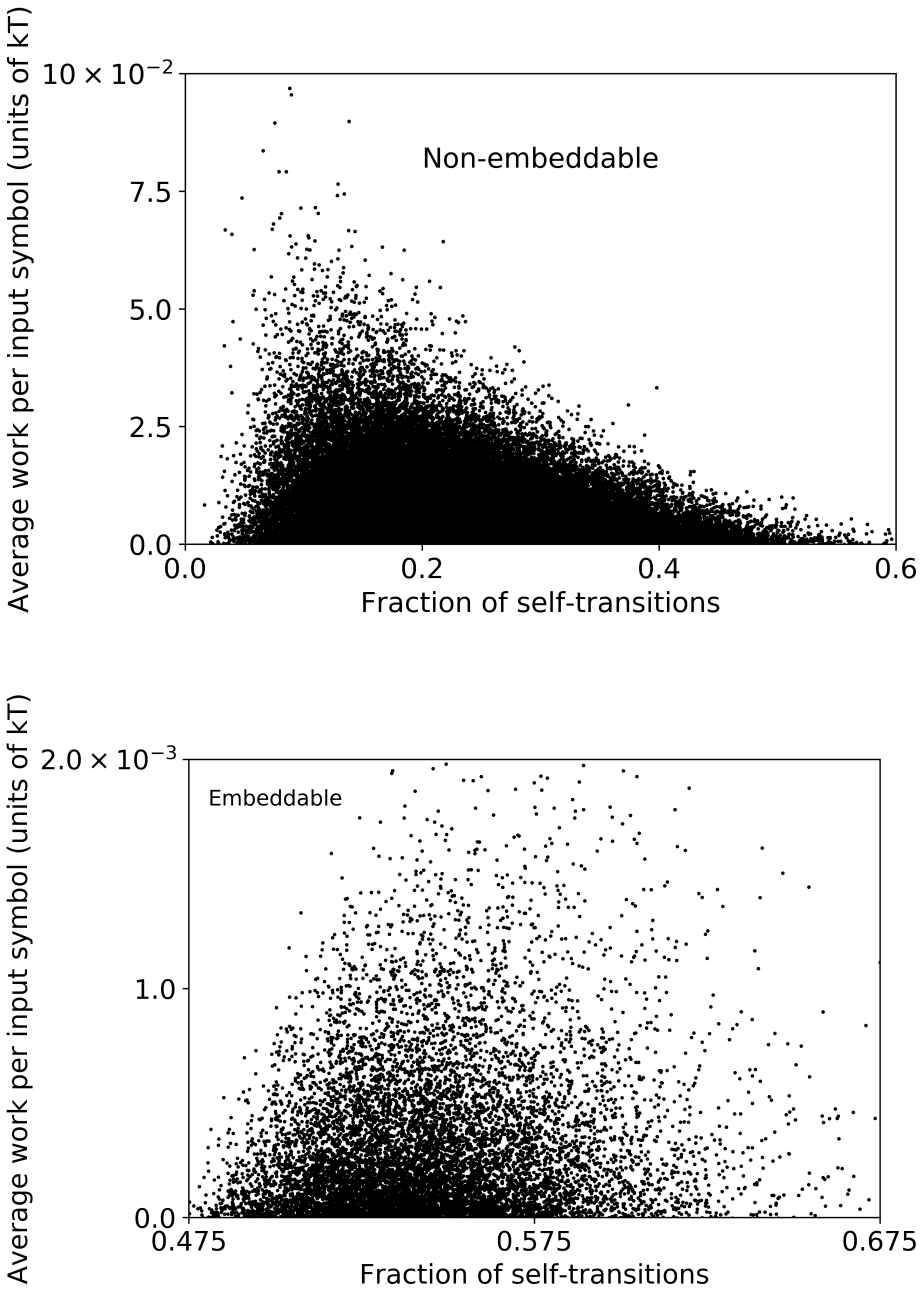
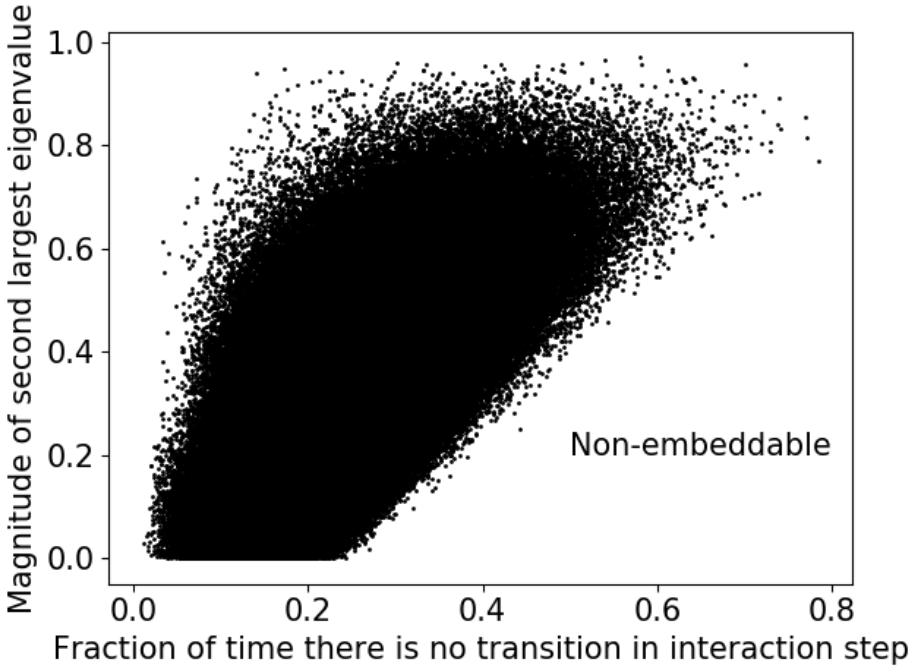


Figure 4.4: The relationship between W_{out} and the fraction of self-transitions

We display only those designs that lead to positive work extraction. There are fewer points in the bottom panel because a smaller fraction of embeddable matrices lead to positive average work. Since the non-embeddable matrices can be made to have a trace of zero, their rate of self-transitions can also be made 0, while for the embeddable matrices a minimum

trace of 1 forces the rate of self-transitions to be at least 1/4, leading to inefficiency. Moreover, approaching the minimal trace of 1 and the corresponding minimal self-transition rate of 1/4 is not a viable strategy for maximizing efficiency, because it would mean that all the smaller eigenvalues would have to approach zero. This requirement is in tension with the fact that the modulus of the second largest eigenvalue, $\lambda_2 < 1$, bounds the distance from equilibrium at the end of an interaction step through $\boldsymbol{M}\boldsymbol{p} = \boldsymbol{\rho}^{(\boldsymbol{M})} + \lambda_2 c_2 \boldsymbol{v}_2 + \lambda_3 c_3 \boldsymbol{v}_3 + \lambda_4 c_4 \boldsymbol{v}_4$, where $\boldsymbol{p} = \boldsymbol{\rho}^{(\boldsymbol{M})} + c_2 \boldsymbol{v}_2 + c_3 \boldsymbol{v}_3 + c_4 \boldsymbol{v}_4$ is an arbitrary starting distribution expanded in the basis of $\boldsymbol{M}$'s eigenvectors, $\boldsymbol{\rho}^{(\boldsymbol{M})}$,$\boldsymbol{v}_2$,$\boldsymbol{v}_3$,$\boldsymbol{v}_4$. Thus, taking the limit as the trace approaches 1 would cause instantaneous relaxation to equilibrium in every interaction step, which prevents work from being extracted.



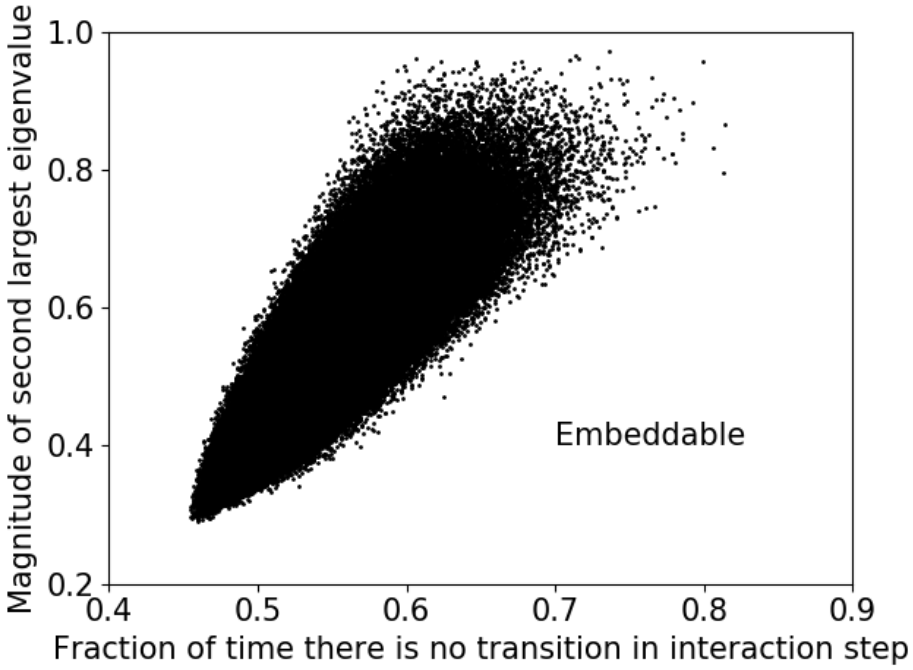


Figure 4.5: The relationship between the magnitude of the second largest eigenvalue, which bounds the distance from equilibrium after an interaction step, and the rate of self transitions. Only matrices leading to positive work extraction are shown.

Intuitively, the device cannot extract work if it relaxes fully to equilibrium in each interaction step because in that case it cannot retain memory. More formally, let E_{eq} denote the energy of the equilibrium distribution. With complete relaxation to equilibrium in each interaction step, the sum of the energy changes over the two switching steps is then $\frac{1}{Z}[E_{A1}e^{-E_{A0}} + E_{A0}e^{-E_{A1}} + E_{B0}e^{-E_{B1}} + E_{B1}e^{-E_{B0}}] - E_{eq} \geq 0$. This quantity is non-negative because the equilibrium distribution pairs the highest energies with the smallest Boltzmann factors, so any reordering of the factors cannot decrease the average energy.

Altogether, there exists a tradeoff between inefficiency coming from staying in the same state too often, and inefficiency coming from relaxing too close to equilibrium. We display the relationship between the self-transition rate and the magnitude of the second largest eigenvalue in Fig. 4.5. This tradeoff would be less severe if more internal states were included, because with a transition matrix of higher dimension, the trace could be kept relatively small even with a large second eigenvalue. However, adding additional internal states would not necessarily guarantee a substantial improvement of the overall performance, because the prohibition on bipartite graphs prevents synchronization with the input, an issue that persists.

Much of the inefficiency of embeddable designs arises from an inability to reliably track the input by switching the device state at each step. But embeddable designs also suffer from a second drawback that would limit work extraction even if the input were a pure string of 1s. To extract work, it is vital that the energy tends to increase during the interaction window. The work extracted is equal to the average number of transitions during these windows, multiplied by the increase in energy per transition. But if we increase the energy of the high energy states to which we hope the system will transition, then we decrease the net number of transitions, since we decrease their occupancy in equilibrium – and embeddable designs can only relax towards equilibrium during the interaction window. Thus there is an unavoidable trade-off for embeddable designs between facilitating many transitions that each contribute only a little to the work extracted, and allowing only a few that contribute a large amount. The overall power is maximised at intermediate values (see, for example, [76]).

4.3 Time-Inhomogeneous Protocols

We have shown that devices that are free-running during the interaction period, which are restricted to reversibly embeddable, time-homogeneous dynamics, can extract only a small

fraction of the work available from an alternating input of 0s and 1s. One might ask whether devices connected to a time-dependent, externally-applied protocol during the interaction period, resulting in time-inhomogeneous dynamics that need not satisfy detailed balance, could perform better. In this setting, for example, it can be ensured that the device's state *must* change during a cycle, allowing a better match to the input's periodicity. Previously considered devices [19] have transition matrices that would require such external control, and possibly additional auxiliary states [125]. However, the thermodynamic analyses performed to date have neglected this fact.

Here we show[4] that it is relatively straightforward to construct a device of this type that extracts, in the quasistatic limit, all the work stored in an input tape of alternating 1s and 0s. External manipulations correspond to changing the energy levels of the system over time [33], and allow the input and/or extraction of work during the interaction period of duration τ. For our purposes, it is sufficient to consider only devices in which the energies of the joint states (s, b) vary during τ, but are restricted to all being *identical* at the beginning and end of each window. In this case, the work of switching to the next input bit on the tape is zero, and only the window τ need be considered to compute the extracted work.

To demonstrate an optimal device, let us compose it of two standard operations: erasure and relaxation. Let a given pair of states within a larger state space each have an occupation probability of $p/2$. Erasure shifts all of this probability to just one state, leaving it occupied with probability p, and the other with a probability 0. Famously, erasure can in principle be performed at a work cost of $pk_{\mathrm{B}}T\ln 2$ [73, 125]. Relaxation is the inverse of erasure, and therefore work of $pk_{\mathrm{B}}T\ln 2$ can be extracted. In both cases, these optimal limits on the work are reached by *thermodynamically* reversible processes, in which manipulations must be applied quasistatically and the reversal of the protocol would restore the initial probability distribution. We also consider switching, or the transfer of probability from one state to another that has zero initial probability. Switching can be decomposed as a relaxation followed by an erase, and therefore has no total work requirement if performed in a thermodynamically reversible manner.

[4]The protocol described in this section was created by Thomas E. Ouldridge.

Let us consider the following transition matrix for the interaction step

$$\boldsymbol{T} = \begin{bmatrix} 0 & 1/2 & 0 & 1/2 \\ 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \\ 1/2 & 0 & 1/2 & 0 \end{bmatrix}, \tag{4.10}$$

where $\boldsymbol{T}\boldsymbol{p}_{sb}$ gives the evolution of $\boldsymbol{p}_{sb}$ during a single interaction window. This transition matrix ensures that the machine transitions to the A state if the input bit was in state 1, and to the B state if the input bit was in state 0. This oscillation is the central switching motif that allows the device to track the input. $\boldsymbol{T}$ is produced by composition of the following sequential operations: a switch from $(A, 0)$ to $(B, 0)$; a switch from $(B, 1)$ to $(A, 1)$; a relaxation from $(B, 0)$ to $(B, 0)$ or $(B, 1)$; and finally a relaxation from $(A, 1)$ to $(A, 0)$ or $(A, 1)$. The states $(A, 1)$ and $(B, 0)$ are effectively used as ancillary states [125] to facilitate the necessary reversing of the machine's state at each step, prior to relaxation.

Regardless of the initial condition, a single application of $\boldsymbol{F}_1\boldsymbol{T}\boldsymbol{F}_0\boldsymbol{T}$ will bring the system to

$$\hat{\boldsymbol{p}}_{sb} = \left[p_{A0}, p_{A1}, p_{B0}, p_{B1}\right]^\top = \left[0, 0, 0, 1\right]^\top, \tag{4.11}$$

in which the state of the device and tape are perfectly coordinated. Since $\hat{\boldsymbol{p}}_{sb}$ is an eigenvector of $\boldsymbol{F}_1\boldsymbol{T}\boldsymbol{F}_0\boldsymbol{T}$ with eigenvalue 1, subsequent applications of $\boldsymbol{F}_1\boldsymbol{T}\boldsymbol{F}_0\boldsymbol{T}$ will also return $\hat{\boldsymbol{p}}_{sb}$. Given the initial condition of $\hat{\boldsymbol{p}}_{sb}$, the switch and relaxation procedures underlying T can be implemented in a thermodynamically reversible manner, yielding $k_B T \ln 2$ of work per T operation (or $2k_B T \ln 2$ per full cycle) due to the relaxation steps. Thus the device extracts all of the available work after an initial alignment cycle.

CHAPTER 5
DISCUSSION

In the following two sections, we discuss implications and future directions for the two projects that constitute this thesis. Together, these results add to an extensive and growing literature linking information theory with thermodynamics. The approach we have taken to understanding the origin of life is, to our knowledge, a significant departure from all previous work on the subject. We hope that this approach may serve as a fruitful new paradigm with which to think about the problem. Our work on the physical limits of information engines is more of a refinement and partial correction of earlier work, and is in this sense less radical, ambitious, and speculative than the other project. The analysis we have performed will undoubtedly advance the scientific understanding of information engines by revealing limits to their functionality, and may help to realize practical implementations of information engines in the future.

5.1 Non-equilibrium abundances for the building blocks of life

Using two examples for which equilibrium thermodynamics seems to prohibit the spontaneous emergence of biologically important molecules, we demonstrated in Ch. 3 that, under very modest assumptions, the concentrations of these molecules might be significantly larger (by many orders of magnitude) when odds are calculated from an average non-equilibrium distribution instead of the equilibrium distribution. It is well known that non-equilibrium conditions of some kind are necessary for life. The degree by which the abundances are improved depends, of course, on how far from equilibrium the system has been driven. Since this is not known, we can not determine the parameter λ in our model, and hence can not provide a definitive number for the concentrations of life's building blocks. But what this study reveals is that, on average, non-equilibrium systems exhibit significantly more favorable conditions, provided that the distance from equilibrium is large enough. Importantly, this approach does not rely on specific knowledge about the conditions on the early Earth.

Another model-independent approach to assessing the odds of life's formation was presented in [101]. The chance of life's emergence on other worlds was calculated from estimating parameters in a Drake-type equation. One of the parameters appearing in this equation is the abiogenesis probability, which estimates the chances of life forming per unit time within

a set of building blocks. An implication of our conclusions is that this parameter ought to be increased on planets where conditions are far from equilibrium, as for example on planets with rich weather phenomena, tectonic activity, or tidal interactions [7]. The necessity for chemical disequilibrium on a planetary scale for the emergence of life has been identified by several authors [88, 7, 96]. The average non-equilibrium distribution (Eq. (3.1)) provides a concrete way of quantifying this effect as a function of how far conditions are from equilibrium.

Explaining the presence of heavy amino acids and peptides is, of course, far from a complete account of life's origins. But we wish to emphasize that the average non-equilibrium distribution's increased odds for attaining otherwise rare states should be independent of the details of any particular reaction. Thus, the same effect is likely to play an important role in other situations where equilibrium thermodynamics appear to create barriers to the emergence of life, e.g. the polymerization of nucleotides in RNA and DNA [4]. It is also possible that the effect might be compounded. This could happen, for example, if a more favorable distribution of amino acids, resulting from a non-equilibrium process is input to another non-equilibrium system that assembles the amino acids into peptides.

Moreover, the biological relevance of this effect need not be limited to the origin of life. Indeed, it is possible that early metabolic processes drove intracellular molecular distributions even further from equilibrium, creating a feedback process whereby the state-space of useful molecules could be more effectively sampled. A similar effect can be observed in kinetic proofreading, where energy is expended to drive reactions out of equilibrium and reduce the rate at which disadvantageous molecules are formed [52].

Altogether, the approach we presented here raises the possibility that the formation of life does not require a particular environment that has been fine-tuned for life. Rather, it may be sufficient to have a set of environments that have been driven far enough away from thermodynamic equilibrium. Not only is non-equilibrium driving a prerequisite for life, but non-equilibrium driving may thus, in this very general way, be a catalyst for life's emergence.

One topic for future investigation is the realizability of various values of λ. That is, what timescales and work inputs are needed to achieve a particular value? The validity of the approach as a whole must also be tested. Specifically, Crooks' model predicts that averaging over a set of non-equilibrium systems that have been driven in different ways should yield a distribution approximated by Eq. 3.1 for some value of λ. This approximation should become better as the set of non-equilibrium systems grows larger, and if it doesn't, that would be an indication that the constraints must be modified or new constraints must be

introduced. Finally, it would be interesting to measure how many systems must be included in the average before the approximation yields an accurate prediction. Taken together, the answers to these questions would help to establish how strongly it should be believed that the mechanism we have discussed is able to explain the origin of life.

5.2 Physical limitations of work extraction from temporal correlations

In Ch. 4 we learned that building physically realistic devices that exploit temporal correlations with a well-defined period to extract work from a heat bath at high efficiency can be challenging. Specifically, devices with time-continuous dynamics cannot extract much work from an alternating input of 0s and 1s, if they operate in a free-running fashion during interaction with the input bit. External manipulation by a time dependent protocol alleviates this issue.

For highly efficient work-extraction systems to emerge (for example due to an evolutionary process), they would then have to develop the ability to operate in an actively driven fashion, rather than passively, in order to reach maximum efficiency. An interesting implication that arises from this is the need for a higher order control structure for active driving. Hierarchical organizations are ubiquitous in biology, and it would be interesting to modify our evolutionary algorithm to explore the emergence of hierarchical structures for greater work extraction.

For such systems, there will be a trade-off between the speed at which they operate and the amount of energy they can extract. In our simple example, we can extract all the work we put in, plus extract net gain from the heat bath, because external manipulations are applied reversibly. But, if we were to put constraints on the execution time, then we should see a trade-off between power and efficiency, similar to effects discussed for example in [95], and references therein. Throughout this work we have considered a particular type of strong correlation in the input string, which is a perfect alternating sequence of 0s and 1s. More complex inputs with weaker correlations may be more challenging to exploit. This question is the topic of recently submitted work [22].

Another question raised by this work is how performance might change if more internal states are allowed. For example, as we argued in Ch. 4, this may help to alleviate the issue of self-transitions. Finally, we do not yet know if the design presented in Fig. 4.2, which was found by our evolutionary algorithm, is truly optimal. Furthermore, we do not have a

good conceptual understanding of why it is even a strong design. As such, we cannot yet predict how the optimal design should change as the structure of the input is altered or more internal states are permitted. Altogether, this leaves a rich set of problems to be explored in the future.

APPENDIX A
APPENDIX

A.1 Crooks' model

Crooks' approach [31] finds the $P(\theta)$ that maximizes the entropy $S[P(\theta)] = -\int P(\theta)\ln P(\theta)d\theta$, subject to:

1. Normalization of probability: $\int P(\theta)d\theta = 1$.

2. $\langle \bar{E}[\theta] \rangle = \int P(\theta)\bar{E}(\theta)d\theta$, a constraint on the average energy. Here, $\bar{E}[\theta] = \sum_i E_i\theta_i$ denotes the average energy, averaged over an individual non-equilibrium distribution, θ.

3. $\langle S \rangle = \int P(\theta)S[\theta]d\theta$, a constraint on the average entropy. The entropy of a non-equilibrium distribution is given by $S[\theta] = -\sum_i \theta_i \ln(\theta_i)$. The Lagrange multiplier , λ, used to enforce this constraint then parameterizes the deviation from the equilibrium distribution. While this constraint is necessary to distinguish equilibrium systems from non-equilibrium ones, it also implicitly introduces the quantitative measure of deviation from equilibrium.

Solving the above constrained optimization problem results in the distribution [31]

$$P(\theta) = \frac{1}{\mathcal{Z}(\beta,\lambda)}\exp\left[-\lambda D(\theta \parallel \rho)\right], \tag{A.1}$$

where $\mathcal{Z}(\beta,\lambda)$ is a normalization constant, and $D(\theta||\rho)$ is the relative entropy between the non-equilibrium distribution, θ, and the corresponding equilibrium distribution ρ. The average non-equilibrium distribution is then found by integrating:

$$\langle\theta\rangle = \int \theta P(\theta)d\theta = \frac{1}{\mathcal{Z}(\beta,\lambda)}\int \theta e^{-\lambda D(\theta\|\rho)}d\theta. \tag{A.2}$$

The flattening effect on the average distribution, which is observed as $\lambda \to 0$, is invariant with respect to the choice of distance measure used in constraint number 3. If this constraint was replaced by a generic constraint on the average distance from equilibrium, $\int P(\theta)d(\theta,\rho)d\theta$,

for any distance measure $d(\theta, \rho)$, then $P(\theta) \propto e^{-\lambda d(\theta,\rho)}$. This would change the exact form of $\langle\theta\rangle$, but the limit $\lambda \to 0$ would nonetheless give a flat average distribution.

Numerical calculations were performed in SageMath. To calculate $\langle\theta\rangle$ in Fig. 3.1 and Fig. 3.3, we generated $20,000$ random distributions, calculated the relative entropy of each one (Eq. 3.2) using the corresponding equilibrium distribution, then weighted them using Eq. A.1 and took the average using Eq. A.2. We also added a sample of the equilibrium distribution to the set of random distributions, in order to correct for the possibility that no samples would be generated close enough to the equilibrium distribution to obtain appreciable weight, when λ was high. For Fig. 4.1, each possible non-equilibrium distribution was generated from a list of 20 random numbers. Each entry in the list was sampled uniformly from the interval $[0, 1]$. The list was then normalized. We generated 10^7 such distributions, and examined the distribution of a single element in the list, which corresponds with the relative concentration of an amino acid. Due to symmetry, the distribution is the same for each amino acid in the limit $\lambda \to 0$. For Figs. 3.4 and 3.5 we were only interested in the extreme non-equilibrium limit $\lambda \to 0$ where all states become equally likely, in which case the probability of each state is just the inverse of the number of states, and the number of states is given by the partition function. The partition function was calculated exactly, using Sage's built in Partitions function.

A.2 Predictive information in steady state

In steady state, the distribution of internal states prior to receiving a 1 is $\boldsymbol{\pi}_{0,s} = [\pi_{0,A}, \pi_{0,B}] = [\pi_{0,A0} + \pi_{0,A1}, \pi_{0,B0} + \pi_{0,B1}]$ (see Eq. 4.4). The distribution prior to receiving a 0 is $\boldsymbol{\pi}_{2,s} = [\pi_{2,A}, \pi_{2,B}] = [\pi_{2,A0} + \pi_{2,A1}, \pi_{2,B0} + \pi_{2,B1}]$ (see Eq. 4.6). The overall probability of being in state A at the end of an interaction step is then $\pi(A) = \frac{1}{2}[\pi_{0,A0} + \pi_{0,A1} + \pi_{2,A0} + \pi_{2,A1}]$, and $\pi(B) = 1 - \pi(A)$. With these expressions, and noting that the overall probability of receiving each bit is $1/2$, the mutual information between the internal state at the end of an interaction interval and the next incoming bit simplifies to: $I_{pred} = \frac{1}{2}[\pi_{0,A} \ln \frac{\pi_{0,A}}{\pi(A)} + \pi_{0,B} \ln \frac{\pi_{0,B}}{\pi(B)} + \pi_{2,A} \ln \frac{\pi_{2,A}}{\pi(A)} + \pi_{2,B} \ln \frac{\pi_{2,B}}{\pi(B)}]$.

A.3 Making reversible random matrices

The following procedure for making reversible matrices at random is taken from [16]. Let $\{K_{ij} | \, j \leq i \leq 4\}$ be a set of 10 real, random variables created by sampling uniformly from

an interval $(0, N_{max}]$. This set forms the lower triangle of a symmetric 4 by 4 matrix with $K_{ij} = K_{ji}$. Define $\pi_j \equiv \sum_i K_{ij}$. Then the matrix $\boldsymbol{M}$ given by $M_{ij} = \frac{K_{ij}}{\pi_j}$ is a reversible stochastic matrix, with stationary distribution $\boldsymbol{\pi}$. We can also make generators $\boldsymbol{G}$ via $\boldsymbol{G} = \boldsymbol{M} - \boldsymbol{I}$, where $\boldsymbol{I}$ is the identity matrix. We generated 10^6 $\boldsymbol{M}$ by this procedure with $N_{max} = 100$, as well as the corresponding generators given by $\boldsymbol{G} = \boldsymbol{M} - \boldsymbol{I}$. Note that this procedure gives transition rates in the range $(0, 1]$. From this set of generators, we made 3×10^6 embeddable transition matrices, with 10^6 each for $\tau = 1$,$\tau = 0.01$,and $\tau = 100$. The interaction intervals with τ other than 1 led to poor performance, because the other values led to very high self-transition rates ($\tau = .01$) and full equilibration during the interaction interval ($\tau = 100$).

A.4 Lower bound on self-transitions

The overall chance of making a self-transition under the action of $\boldsymbol{M}$ on distribution $\boldsymbol{p}$ is $\boldsymbol{d}(\boldsymbol{M})^\top \boldsymbol{p}$, where $\boldsymbol{d}(\boldsymbol{M})$ is the vector of the diagonal entries of $\boldsymbol{M}$. We can construct a matrix achieving the smallest numerically found self-transition rate by assuming that this is achieved when the trace is minimal. For embeddable stochastic matrices, the lower bound on the trace is 1 (see argument in Section 4.2). In this case, $\boldsymbol{M}$ has a single eigenvalue of 1 and all other eigenvalues 0. Thus, $\boldsymbol{M}$ is rank-1 with 4 repeats of the same column $\mathbf{m} = [m_0, m_1, m_2, m_3]^\top$. The matrix $\boldsymbol{C} = \boldsymbol{M}\boldsymbol{F}_0\boldsymbol{M}\boldsymbol{F}_1$ is then equal to $\boldsymbol{M}$, and the steady state of the complete cycle is nothing more than the repeated column, i.e., $\boldsymbol{\pi}_0(s_n, b'_{n-1}) = \mathbf{m}$. Eq. (4.9) says that the average number of self-transitions over the cycle is $L = \frac{1}{2}\big[(m_0+m_1)^2 + (m_2+m_3)^2\big]$. Minimizing this number is a simple optimization problem with the constraint $m_0 + m_1 + m_2 + m_3 = 1$. The solution is $(m_0 + m_1) = (m_2 + m_3) = \frac{1}{2}$. Substituting these in gives $L = \frac{1}{4}$. This condition can be satisfied for approximately embeddable $\boldsymbol{M}$, which can be constructed, for example, by perturbing the matrix with all entries equal to 1/4 so that the smaller eigenvalues are just slightly positive and not exactly zero.

A.5 Constraints on the eigenvalues of reversible and embeddable matrices

The following results may be found in [63].

1. If M is reversible then M is diagonalizable and the eigenvalues of M are real.

Proof: Let $\mathbf{R} = \mathbf{diag}(\sqrt{\pi_1}, \sqrt{\pi_2}, ..., \sqrt{\pi_n})$, where $\boldsymbol{\pi}$ is the steady state vector of $\boldsymbol{M}$. We first show that $\mathbf{S} = \mathbf{R^{-1}}\boldsymbol{M}\mathbf{R}$ is symmetric:

$$S_{ij} = R^{-1}_{ik} M_{kl} R_{lj} \tag{A.3}$$

$$= R^{-1}_{ii} M_{ij} R_{jj} \tag{A.4}$$

$$= \frac{1}{\sqrt{\pi_i}} M_{ij} \sqrt{\pi_j} \tag{A.5}$$

$$= \sqrt{\frac{M_{ji}}{\pi_j M_{ij}}} M_{ij} \sqrt{\pi_j} \tag{A.6}$$

$$= \sqrt{M_{ji} M_{ij}} \tag{A.7}$$

where we use the convention that repeated indices are summed over. In the second line we used the fact that $\mathbf{R}$ and $\mathbf{R}^{-1}$ are diagonal. In the fourth line we used the relation $M_{ji}\pi_i = M_{ij}\pi_j$ that defines a reversible matrix to replace π_i with $\frac{M_{ij}\pi_j}{M_{ji}}$. The last line reveals that $\mathbf{S}$ is symmetric. Since $\mathbf{S}$ is symmetric, and symmetric matrices are a subset of Hermitian matrices, we have the well known result from quantum mechanics that S is diagonalizable and the eigenvalues of $\mathbf{S}$ are real. Then $\boldsymbol{M} = \mathbf{R}^{-1}\mathbf{S}\mathbf{R}$ is similar to $\mathbf{S}$ and is also diagonalizable, with real eigenvalues.

1. If M is reversibly embeddable, then M is reversible and the eigenvalues of M are all positive.

Proof: $\boldsymbol{M}$ being reversibly embeddable means there exists a reversible generator $\mathbf{Q}$ such that $\boldsymbol{M} = e^{\mathbf{Q}}$. Defining $\mathbf{R}$ and $\mathbf{S}$ in the analogous way for $\mathbf{Q}$, we can run through the same argument to show that $\mathbf{S}$ is symmetric. A linear combination of symmetric matrices is obviously symmetric. Any power of a symmetric matrix is also symmetric, which can easily be shown by induction: $A^2_{ij} = A_{ik}A_{kj} = A_{ki}A_{jk} = A_{jk}A_{ki} = \mathbf{A}^2_{ij}$, and for the inductive

hypothesis we have $A_{ij}^{n+1} = A_{ik}^n A_{kj} = A_{ki}^n A_{jk} = A_{jk} A_{ki}^n = A_{ji}^{n+1}$. Combining these two facts means that $e^{\mathbf{S}}$ is symmetric if $\mathbf{S}$ is. $e^{\mathbf{S}} = e^{\mathbf{R^{-1}QR}} = \mathbf{R}^{-1} e^{\mathbf{Q}} \mathbf{R} = \mathbf{R^{-1}MR}$. Symmetry of $\mathbf{R^{-1}MR}$ then implies $\boldsymbol{M}$ is reversible:

$$(R^{-1}MR)_{ij} = R_{ik}^{-1} M_{kl} R_{lj} \tag{A.8}$$

$$= R_{ii}^{-1} M_{ij} R_{jj} \tag{A.9}$$

$$= \frac{1}{\sqrt{\pi_i}} M_{ij} \sqrt{\pi_j} \tag{A.10}$$

$$= \frac{1}{\sqrt{\pi_j}} M_{ji} \sqrt{\pi_i} \tag{A.11}$$

where $\boldsymbol{\pi}$ now refers to the steady state distribution of $\mathbf{Q}$ which was used in the construction of $\mathbf{R}$. Combination of the last two lines implies $M_{ji}\pi_i = M_{ij}\pi_j$, and therefore $\boldsymbol{M}$ is reversible with the same reversible distribution as $\mathbf{Q}$. As $\mathbf{Q}$ is assumed to be reversible, it has real eigenvalues and is diagonalizable. Writing $\mathbf{Q}$ in the basis in which it's diagonal, we have

$$\boldsymbol{M} = e^{\mathbf{Q}} \tag{A.12}$$

$$= \sum_{n=0}^{\infty} \frac{\mathbf{diag}(q_1, q_2, ..., q_m)^n}{n!} \tag{A.13}$$

$$= \sum_{n=0}^{\infty} \mathbf{diag}(\frac{q_1^n}{n!}, \frac{q_2^n}{n!}, ..., \frac{q_m^n}{n!}) \tag{A.14}$$

$$= \mathbf{diag}(e^{q_1}, e^{q_2}, ..., e^{q_m}) \tag{A.15}$$

where $\{q_1, q_2, ..., q_m\}$ are the eigenvalues of $\mathbf{Q}$. We see from Eq. A.15 that the eigenvalues of $\boldsymbol{M}$ are the exponentials of the eigenvalues of $\mathbf{Q}$. As the eigenvalues of $\mathbf{Q}$ are real, this means that the eigenvalues of $\boldsymbol{M}$ are positive.

BIBLIOGRAPHY

[1] Christoph Adami. Information-theoretic considerations concerning the origin of life. *Orig. Life Evol. Biosph.*, 45(3):309–317, 2015.

[2] JP Amend and EL Shock. Energetics of amino acid synthesis in hydrothermal ecosystems. *Science*, 281(5383):1659–1662, 1998.

[3] George E Andrews. *The Theory of Partitions*. Number 2. Cambridge University Press, Cambdridge, 1998.

[4] David Andrieux and Pierre Gaspard. Nonequilibrium generation of information in copolymerization processes. *Proc. Natl. Acad. Sci.*, 105(28):9516–9521, 2008.

[5] Jeffrey L. Bada. New insights into prebiotic chemistry from stanley miller's spark discharge experiments. *Chem. Soc. Rev.*, 42:2186–2196, 2013.

[6] Narayanaswamy Balakrishnan and Valery B Nevzorov. *A primer on statistical distributions*. John Wiley & Sons, Hoboken, 2004.

[7] LM Barge, E Branscomb, JR Brucato, SSS Cardoso, JHE Cartwright, SO Danielache, D Galante, TP Kee, Y Miguel, S Mojzsis, et al. Thermodynamics, disequilibrium, evolution: Far-from-equilibrium geological and chemical considerations for origin-of-life research. *Orig. Life Evol. Biosph.*, pages 1–18, 2016.

[8] Friedrich G Barth and Axel Schmid. *Ecology of sensing*. Springer Science & Business Media, 2013.

[9] Charles H Bennett. Notes on landauer's principle, reversible computation, and maxwell's demon. *Studies In History and Philosophy of Science Part B: Studies In History and Philosophy of Modern Physics*, 34(3):501–510, 2003.

[10] Charles H Bennett. Notes on landauer's principle, reversible computation, and maxwell's demon. *Studies In History and Philosophy of Science Part B: Studies In History and Philosophy of Modern Physics*, 34(3):501–510, 2003.

[11] Charles H Bennett and Rolf Landauer. The fundamental physical limits of computation. *Scientific American*, 253(1):48–57, 1985.

[12] Howard C Berg and Edward M Purcell. Physics of chemoreception. *Biophysical journal*, 20(2):193–219, 1977.

[13] Antoine Bérut, Artak Arakelyan, Artyom Petrosyan, Sergio Ciliberto, Raoul Dillenschneider, and Eric Lutz. Experimental verification of landauer's principle linking information and thermodynamics. *Nature*, 483(7388):187, 2012.

[14] William Bialek. *Biophysics: searching for principles*. Princeton University Press, 2012.

[15] Mary L Boas. *Mathematical methods in the physical sciences*. Wiley, 2006.

[16] Charles Bordenave, Pietro Caputo, and Djalil Chafai. Spectrum of large random reversible markov chains: Heavy-tailed weights on the complete graph. *Ann. Probab.*, 39(4):1544–1590, 07 2011.

[17] AS Botvina, AD Jackson, and IN Mishustin. Partitioning composite finite systems. *Phys. Rev. E*, 62(1):R64, 2000.

[18] Alexander B Boyd, Dibyendu Mandal, and James P Crutchfield. Identifying functional thermodynamics in autonomous maxwellian ratchets. *New J. Phys.*, 18(2):023049, 2016.

[19] Alexander B. Boyd, Dibyendu Mandal, and James P. Crutchfield. Correlation-powered information engines and the thermodynamics of self-correction. *Phys. Rev. E*, 95:012152, Jan 2017.

[20] Alexander B Boyd, Dibyendu Mandal, and James P Crutchfield. Leveraging environmental correlations: The thermodynamics of requisite variety. *J. Stat. Phys.*, 167(6):1555–1585, 2017.

[21] André Brack. From interstellar amino acids to prebiotic catalytic peptides: a review. *Chem. Biodivers.*, 4(4):665–679, 2007.

[22] Rory A Brittain, Nick S Jones, and Thomas E Ouldridge. Biochemical szilard engines for memory-limited inference. *arXiv preprint arXiv:1812.08401*, 2018.

[23] David Edward Bruschi, Martí Perarnau-Llobet, Nicolai Friis, Karen V. Hovhannisyan, and Marcus Huber. Thermodynamics of creating correlations: Limitations and optimal protocols. *Phys. Rev. E*, 91:032118, Mar 2015.

[24] Sherwood Chang. Prebiotic synthesis in planetary environments. In *The Chemistry of Life's Origins*, pages 259–299. Springer, Berlin, 1993.

[25] KC Chase and AZ Mekjian. Nuclear fragmentation and its parallels. *Phys. Rev. C*, 49(4):2164, 1994.

[26] Christopher Chyba and Carl Sagan. Endogenous production, exogenous delivery and impact-shock synthesis of organic molecules: an inventory for the orig. life. 1992.

[27] HJ Cleaves, AD Aubrey, and JL Bada. An evaluation of the critical parameters for abiotic peptide synthesis in submarine hydrothermal systems. *Orig. Life Evol. Biosph.*, 39(2):109–126, 2009.

[28] Delphine Collin, Felix Ritort, Christopher Jarzynski, Steven B Smith, Ignacio Tinoco Jr, and Carlos Bustamante. Verification of the crooks fluctuation theorem and recovery of rna folding free energies. *Nature*, 437(7056):231, 2005.

[29] Thomas M Cover and Joy A Thomas. *Elements of Information Theory*. John Wiley & Sons, Hoboken, 2012.

[30] Gavin E Crooks. Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Physical Review E*, 60(3):2721, 1999.

[31] Gavin E. Crooks. Beyond boltzmann-gibbs statistics: Maximum entropy hyperensembles out of equilibrium. *Phys. Rev. E*, 75:041119, Apr 2007.

[32] Oscar C O Dahlsten, Renato Renner, Elisabeth Rieper, and Vlatko Vedral. Inadequacy of von neumann entropy for characterizing extractable work. *New Journal of Physics*, 13(5):053015, 2011.

[33] Christian De Duve. *A Guided Tour of the Living Cell*, volume 2. WH Freeman Trade, 1984.

[34] Lidia Del Rio, Johan Åberg, Renato Renner, Oscar Dahlsten, and Vlatko Vedral. The thermodynamic meaning of negative entropy. *Nature*, 474(7349):61, 2011.

[35] John P DeLong, Jordan G Okie, Melanie E Moses, Richard M Sibly, and James H Brown. Shifts in metabolic scaling, production, and efficiency across major evolutionary transitions of life. *Proceedings of the National Academy of Sciences*, 107(29):12941–12945, 2010.

[36] M Dixon and EC Webb. *Enzymes*. Academic Press, Cambridge, 1964.

[37] Evan D. Dorn, Kenneth H. Nealson, and Christoph Adami. Monomer abundance distribution patterns as a universal biosignature: Examples from terrestrial and digital life. *J. Mol. Evol.*, 72(3):283–295, 2011.

[38] M. Esposito and C. Van den Broeck. Second law and landauer principle far from equilibrium. *EPL (Europhysics Letters)*, 95(4):40004, 2011.

[39] M. Esposito and C. Van den Broeck. Second law and landauer principle far from equilibrium. *EPL (Europhys. Lett.)*, 95(4):40004, 2011.

[40] Massimiliano Esposito, Katja Lindenberg, and Christian Van den Broeck. Entropy production as correlation between system and reservoir. *New Journal of Physics*, 12(1):013013, 2010.

[41] Philippe Faist, Frédéric Dupuis, Jonathan Oppenheim, and Renato Renner. The minimal work cost of information processing. *Nature communications*, 6:7669, 2015.

[42] Philippe Faist and Renato Renner. Fundamental work cost of quantum processes. *Physical Review X*, 8(2):021011, 2018.

[43] Paul J. Flory. Thermodynamics of heterogeneous polymers and their solutions. *The J. Chem. Phys.*, 12(11):425–438, 1944.

[44] Robert G Gallager. *Discrete stochastic processes*, volume 321. Springer Science & Business Media, Berlin, 2012.

[45] Andrew JP Garner, Jayne Thompson, Vlatko Vedral, and Mile Gu. Thermodynamics of complexity and pattern manipulation. *Phys. Rev. E*, 95(4):042140, 2017.

[46] J Willard Gibbs. *Elementary Principles in Statistical Mechanics*. Courier Corporation, Mineola, 2014.

[47] Sandra C Greer. Physical chemistry of equilibrium polymerization. *The J. Phys. Chem. B*, 102(28):5413–5422, 1998.

[48] Arne L Grimsmo and Susanne Still. Quantum predictive filtering. *Physical Review A*, 94(1):012338, 2016.

[49] John Harte. *Maximum entropy and ecology: a theory of abundance, distribution, and energetics*. OUP Oxford, 2011.

[50] Barry Herschy, Alexandra Whicher, Eloi Camprubi, Cameron Watson, Lewis Dartnell, John Ward, Julian R. G. Evans, and Nick Lane. An origin-of-life reactor to simulate alkaline hydrothermal vents. *J. Mol. Evol.*, 79(5):213–227, 2014.

[51] Paul G Higgs and Ralph E Pudritz. A thermodynamic basis for prebiotic amino acid synthesis and the nature of the first genetic code. *Astrobiology*, 9(5):483–490, 2009.

[52] J. J. Hopfield. Kinetic proofreading: A new mechanism for reducing errors in biosynthetic processes requiring high specificity. *Proc. Natl. Acad. Sci.*, 71(10):4135–4139, 1974.

[53] Jordan M. Horowitz, Takahiro Sagawa, and Juan M. R. Parrondo. Imitating chemical motors with optimal information motors. *Phys. Rev. Lett.*, 111:010602, Jul 2013.

[54] Howard C Howland, Stacey Merola, and Jennifer R Basarab. The allometry and scaling of the size of vertebrate eyes. *Vision research*, 44(17):2043–2065, 2004.

[55] Marcus Huber, MartÃ Perarnau-Llobet, Karen V Hovhannisyan, Paul Skrzypczyk, Claude KlÃ¶ckl, Nicolas Brunner, and Antonio AcÃn. Thermodynamic cost of creating correlations. *New Journal of Physics*, 17(6):065008, 2015.

[56] Joseph R Iafrate, Steven J Miller, and Frederick W Strauch. Equipartitions and a distribution for numbers: A statistical model for benford's law. *Phys. Rev. E*, 91(6):062138, 2015.

[57] Junichi Ishikawa, Kazuma Takara, Hiroshi-H. Hasegawa, and Dean Driebe. Application of the generalized work relation for an n-level quantum system. *Entropy*, 16(6):3471D3481, Jun 2014.

[58] Christopher Jarzynski. Nonequilibrium equality for free energy differences. *Physical Review Letters*, 78(14):2690, 1997.

[59] JM Jauch and JG Baron. Entropy, information and szilard's paradox. *Maxwell's Demon: Entropy, information, computing*, pages 160–172, 1990.

[60] E. T. Jaynes. Information theory and statistical mechanics. *Phys. Rev.*, 106:620–630, May 1957.

[61] Edwin T Jaynes. *Probability Theory: The Logic of Science*. Cambridge University Press, Cambridge, 2003.

[62] Hee Joon Jeon and Sang Wook Kim. Optimal work of the quantum szilard engine under isothermal processes with inevitable irreversibility. *New J. Phys.*, 18(4):043002, 2016.

[63] Chen Jia. A solution to the reversible embedding problem for finite markov chains. *Statistics and Probability Letters*, 116(Supplement C):122 – 130, 2016.

[64] Jagat Narain Kapur. *Maximum-Entropy Models in Science and Engineering*. John Wiley & Sons, Hoboken, 1989.

[65] Christopher P Kempes, David Wolpert, Zachary Cohen, and Juan Pérez-Mercader. The thermodynamic efficiency of computations made in cells across the range of life. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 375(2109):20160343, 2017.

[66] Sang Wook Kim, Takahiro Sagawa, Simone De Liberato, and Masahito Ueda. Quantum szilard engine. *Phys. Rev. Lett.*, 106(7):070401, 2011.

[67] M Kleiber. Body size and metabolism. *ENE*, 1(9), 1932.

[68] Kenneth J Kokjer. The information capacity of the human fingertip. *IEEE transactions on systems, man, and cybernetics*, 17(1):100–102, 1987.

[69] Jonne V Koski, Ville F Maisi, Jukka P Pekola, and Dmitri V Averin. Experimental realization of a szilard engine with a single electron. *Proceedings of the National Academy of Sciences*, 111(38):13786–13789, 2014.

[70] Moritz Kreysing, Lorenz Keil, Simon Lanzmich, and Dieter Braun. Heat flux across an open pore enables the continuous replication and selection of oligonucleotides towards increasing length. *Nat. Chem.*, 7(3):203—208, March 2015.

[71] S Kullback. *Statistics and Information Theory*. J. Wiley and Sons, Hoboken, 1959.

[72] Jean-François Lambert. Adsorption and polymerization of amino acids on mineral surfaces: a review. *Orig. Life Evol. Biosph.*, 38(3):211–242, 2008.

[73] Rolf Landauer. Irreversibility and heat generation in the computing process. *IBM J. Res. Dev.*, 5(3):183–191, 1961.

[74] SJ Lee and AZ Mekjian. Canonical studies of the cluster distribution, dynamical evolution, and critical temperature in nuclear multifragmentation processes. *Phys. Rev. C*, 45(3):1284, 1992.

[75] David JC MacKay and David JC Mac Kay. *Information theory, inference and learning algorithms*. Cambridge university press, 2003.

[76] Dibyendu Mandal and Christopher Jarzynski. Work and information processing in a solvable model of maxwells demon. *Proc. Nat. Acad. Sci. U.S.A*, 109(29):11641–11645, 2012.

[77] Dibyendu Mandal, H. T. Quan, and Christopher Jarzynski. Maxwell's refrigerator: An exactly solvable model. *Phys. Rev. Lett.*, 111:030602, Jul 2013.

[78] Rahul Marathe and JMR Parrondo. Cooling classical particles with a microcanonical szilard engine. *Phys. Rev. Lett.*, 104(24):245704, 2010.

[79] R. Bruce Martin. Free energies and equilibria of peptide bond hydrolysis and formation. *Biopolymers*, 45(5):351–353, 1998.

[80] William Martin, John Baross, Deborah Kelley, and Michael J Russell. Hydrothermal vents and the origin of life. *Nat. Rev. Microbiol.*, 6(11):805–814, 2008.

[81] Christof B Mast, Severin Schink, Ulrich Gerland, and Dieter Braun. Escalation of polymerization in a thermal gradient. *Proc. Natl. Acad. Sci.*, 110(20):8030–8035, 2013.

[82] Thomas M. McCollom. Miller-urey and beyond: What have we learned about prebiotic organic synthesis reactions in the past 60 years? *Annu. Rev. Earth Planet. Sci.*, 41(1):207–229, 2013.

[83] Thomas McGrath, Nick S. Jones, Pieter Rein ten Wolde, and Thomas E. Ouldridge. Biochemical machines for the interconversion of mutual information and work. *Phys. Rev. Lett.*, 118:028101, Jan 2017.

[84] AZ Mekjian. Model of a fragmentation process and its power-law behavior. *Phys. Rev. Lett.*, 64(18):2125, 1990.

[85] Stanley L. Miller and Harold C. Urey. Organic compound synthesis on the primitive earth. *Science*, 130(3370):245–251, 1959.

[86] Thierry Mora, Aleksandra M Walczak, William Bialek, and Curtis G Callan. Maximum entropy models for antibody diversity. *Proceedings of the National Academy of Sciences*, 107(12):5405–5410, 2010.

[87] Luciano G Moretto and Gordon J Wozniak. The role of the compound nucleus in complex fragment emission at low and intermediate energies. *Prog. Part. Nucl. Phys.*, 21:401–457, 1988.

[88] Harold Morowitz and Eric Smith. Energy flow and the organization of life. *Complexity*, 13(1):51–59, 2007.

[89] Michael A Nielsen and Isaac Chuang. Quantum computation and quantum information, 2002.

[90] Jeremy E Niven, John C Anderson, and Simon B Laughlin. Fly photoreceptors demonstrate energy-information trade-offs in neural coding. *PLoS biology*, 5(4):e116, 2007.

[91] Jeremy E Niven and Simon B Laughlin. Energy limitation as a selective pressure on the evolution of sensory systems. *Journal of Experimental Biology*, 211(11):1792–1804, 2008.

[92] J. Oppenheim, M. Horodecki, P. Horodecki, and R. Horodecki. Thermodynamical Approach to Quantifying Quantum Correlations. *Physical Review Letters*, 89(18):180402, October 2002.

[93] Martí Perarnau-Llobet, Karen V. Hovhannisyan, Marcus Huber, Paul Skrzypczyk, Nicolas Brunner, and Antonio Acín. Extractable work from correlations. *Phys. Rev. X*, 5:041011, Oct 2015.

[94] Sandra Pizzarello, Yongsong Huang, and Megan Fuller. The carbon isotopic distribution of murchison amino acids. *Geochim. Cosmochim. Acta*, 68(23):4963–4969, 2004.

[95] Karel Proesmans, Bart Cleuren, and Christian Van den Broeck. Power-efficiency-dissipation relations in linear thermodynamics. *Phys. Rev. Lett.*, 116(22):220601, 2016.

[96] Michael J Russell, Laura M Barge, Rohit Bhartia, Dylan Bocanegra, Paul J Bracher, Elbert Branscomb, Richard Kidd, Shawn McGlynn, David H Meier, Wolfgang

Nitschke, et al. The drive to life on wet and icy worlds. *Astrobiology*, 14(4):308–343, 2014.

[97] T. Sagawa and M. Ueda. Second law of thermodynamics with discrete quantum feedback control. *AIP Conference Proceedings*, 1110(1):21–24, 2009.

[98] Takahiro Sagawa. Thermodynamics of information processing in small systems. *Progress of Theoretical Physics*, 127(1):1–56, 2012.

[99] O-P Saira, Y Yoon, T Tanttu, Mikko Möttönen, DV Averin, and Jukka P Pekola. Test of the jarzynski and crooks fluctuation relations in an electronic system. *Physical review letters*, 109(18):180601, 2012.

[100] Sina Salek, Daniela Cadamuro, Philipp Kammerlander, and Karoline Wiesner. Quantum rate-distortion coding of relevant information. *IEEE Transactions on Information Theory*, 2018.

[101] Caleb Scharf and Leroy Cronin. Quantifying the orig. life on a planetary scale. *Proc. Natl. Acad. Sci.*, 113(29):8127–8132, 2016.

[102] Claude Elwood Shannon. A mathematical theory of communication. *Bell system technical journal*, 27(3):379–423, 1948.

[103] Robert Shaw. *The Dripping Faucet as a Model Chaotic System*. Aerial Press, Santa Cruz, 1984.

[104] Robert S. Shaw, Norman Packard, Matthias Schroter, and Harry L. Swinney. Geometry-induced asymmetric diffusion. *Proc. Natl. Acad. Sci.*, 104(23):9580–9584, 2007.

[105] Everett L. Shock and Mitchell D. Schulte. Organic synthesis during fluid mixing in hydrothermal systems. *J. Geophys. Res.: Planets*, 103(E12):28513–28527, 1998.

[106] John Skilling. *Maximum Entropy and Bayesian Methods*, volume 36. Springer Science & Business Media, Berlin, 2013.

[107] Paul Skrzypczyk, Anthony J Short, and Sandu Popescu. Work extraction and thermodynamics for individual quantum systems. *Nature communications*, 5:4185, 2014.

[108] Allan W Snyder, Simon B Laughlin, and Doekele G Stavenga. Information capacity of eyes. *Vision res*, 17(10):1163–1175, 1977.

[109] Susanne Still. Thermodynamic cost and benefit of data representations. *arXiv preprint arXiv:1705.00612*, 2017.

[110] Susanne Still. Thermodynamic cost and benefit of data representations. *arXiv preprint arXiv:1705.00612*, 2017.

[111] Susanne Still, David A Sivak, Anthony J Bell, and Gavin E Crooks. Thermodynamics of prediction. *Physical review letters*, 109(12):120604, 2012.

[112] Philipp Strasberg, Gernot Schaller, Tobias Brandes, and Massimiliano Esposito. Thermodynamics of a physical model implementing a maxwell demon. *Phys. Rev. Lett.*, 110:040601, Jan 2013.

[113] Leo Szilard. Über die entropieverminderung in einem thermodynamischen system bei eingriffen intelligenter wesen. *Zeitschrift für Physik*, 53(11-12):840–856, 1929.

[114] Leo Szilard. On the decrease of entropy in a thermodynamic system by the intervention of intelligent beings. *Behav. Sci.*, 9(4):301–310, 1964.

[115] K Takara, H-H Hasegawa, and DJ Driebe. Generalization of the second law for a transition between nonequilibrium states. *Phys. Lett. A*, 375(2):88–92, 2010.

[116] Gašper Tkačik, Olivier Marre, Thierry Mora, Dario Amodei, Michael J Berry II, and William Bialek. The simplest maximum entropy model for collective behavior in a neural network. *Journal of Statistical Mechanics: Theory and Experiment*, 2013(03):P03011, 2013.

[117] Arthur V. Tobolsky. Equilibrium distribution in sizes for linear polymer molecules. *The J. Chem. Phys.*, 12(10):402–404, 1944.

[118] Suriyanarayanan Vaikuntanathan and Christopher Jarzynski. Modeling maxwell's demon with a microcanonical szilard engine. *Phys. Rev. E*, 83(6):061120, 2011.

[119] Christian Van den Broeck et al. Stochastic thermodynamics: A brief introduction. *Phys. Complex Colloids*, 184:155–193, 2013.

[120] Giuseppe Vitagliano, Claude Klöckl, Marcus Huber, and Nicolai Friis. Trade-off between work and correlations in quantum thermodynamics. *arXiv preprint arXiv:1803.06884*, 2018.

[121] John Von Neumann. *Mathematical Foundations of Quantum Mechanics: New Edition*. Princeton university press, 2018.

[122] S. Walker, P. Davies, and G. Ellis. *From Matter to Life: Information and Causality*. Cambridge University Press, Cambridge, 2017.

[123] Geoffrey B West and James H Brown. The origin of allometric scaling laws in biology from genomes to ecosystems: towards a quantitative unifying theory of biological structure and organization. *Journal of experimental biology*, 208(9):1575–1592, 2005.

[124] Wikipedia contributors. Orders of magnitude (mass) — Wikipedia, the free encyclopedia, 2018. [Online; accessed 1-November-2018].

[125] D. H. Wolpert, A. Kolchinsky, and J. A. Owen. The minimal hidden computer needed to implement a visible computation. *arXiv1708:08494*, 2017.

[126] David H. Wolpert. The free energy requirements of biological organisms; implications for evolution. *Entropy*, 18(4):138, 2016.

[127] Wojciech Hubert Zurek. Maxwell's demon, szilard's engine and quantum measurements. In *Frontiers of nonequilibrium statistical physics*, pages 151–161. Springer, Berlin, 1986.