New Journal of Physics
The open access journal at the forefront of physics

Deutsche Physikalische Gesellschaft DPG IOP Institute of Physics

PAPER • OPEN ACCESS

Number of hidden states needed to physically implement a given conditional distribution

Recent citations

- Biochemical Szilard engines for memory-limited inference
Rory A Brittain *et al*

To cite this article: Jeremy A Owen *et al* 2019 *New J. Phys.* **21** 013022

View the article online for updates and enhancements.

IOP ebooks™

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the collection - download the first chapter of every title for free.

This content was downloaded from IP address 192.12.12.234 on 19/08/2019 at 23:18

New Journal of Physics
The open access journal at the forefront of physics

Deutsche Physikalische Gesellschaft DPG
IOP Institute of Physics

Published in partnership with: Deutsche Physikalische Gesellschaft and the Institute of Physics

OPEN ACCESS

RECEIVED
31 August 2018

REVISED
16 November 2018

ACCEPTED FOR PUBLICATION
12 December 2018

PUBLISHED
28 January 2019

Original content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence.

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

PAPER

Number of hidden states needed to physically implement a given conditional distribution

Jeremy A Owen[1], **Artemy Kolchinsky**[2] **and David H Wolpert**[2,3,4,5]

[1] Physics of Living Systems Group, Department of Physics, Massachusetts Institute of Technology, 400 Tech Square, Cambridge, MA 02139, United States of America
[2] Santa Fe Institute, United States of America
[3] Arizona State University, United States of America
[4] http://davidwolpert.weebly.com
[5] Massachusetts Institute of Technology.

E-mail: jaowen@mit.edu

Keywords: nonequilibrium statistical physics, stochastic thermodynamics, entropy production, Markov processes, master equation, embedding problem, hidden states

Abstract

We consider the problem of how to construct a physical process over a finite state space X that applies some desired conditional distribution P to initial states to produce final states. This problem arises often in the thermodynamics of computation and nonequilibrium statistical physics more generally (e.g. when designing processes to implement some desired computation, feedback controller, or Maxwell demon). It was previously known that some conditional distributions cannot be implemented using *any* master equation that involves just the states in X. However, here we show that any conditional distribution P can in fact be implemented—if additional 'hidden' states not in X are available. Moreover, we show that it is always possible to implement P in a thermodynamically reversible manner. We then investigate a novel cost of the physical resources needed to implement a given distribution P: the minimal number of hidden states needed to do so. We calculate this cost exactly for the special case where P represents a single-valued function, and provide an upper bound for the general case, in terms of the non-negative rank of P. These results show that having access to one extra binary degree of freedom, thus doubling the total number of states, is sufficient to implement any P with a master equation in a thermodynamically reversible way, if there are no constraints on the allowed form of the master equation. (Such constraints can greatly increase the minimal needed number of hidden states.) Our results also imply that for certain P that can be implemented *without* hidden states, having hidden states permits an implementation that generates less heat.

1. Introduction

Master equation dynamics over a discrete state space play a fundamental role in nonequilibrium statistical physics and stochastic thermodynamics [1–3], and are used to model a wide variety of physical systems. Such dynamics can arise after coarse-graining the continuous phase space of a Hamiltonian system coupled to a heat bath [4–6], or as semiclassical approximations of the evolution of an open quantum system with discrete states [1, 7].

The linearity of the master equation implies a linear relationship between the distribution over the system's states at the initial time $t = 0$ and some final time (here taken to be $t = 1$ without loss of generality). This is true even if the master equation is time-inhomogeneous (i.e. non-autonomous). We can express this relation as

$$p(1) = Pp(0),$$

where $p(0)$ and $p(1)$ are column vectors whose entries sum to one, representing probability distributions at times $t = 0$ and $t = 1$ respectively. If our system has n states, the matrix P is an $n \times n$ stochastic matrix, representing the conditional distribution of the system state at the final time given its state at the initial time. The entry P_{ij} is the probability that the system is in state i at the final time given that it was in state j at the initial time.

© 2019 The Author(s). Published by IOP Publishing Ltd on behalf of the Institute of Physics and Deutsche Physikalische Gesellschaft

The problem of characterizing the set of all stochastic matrices P that can arise this way from a (time-inhomogeneous) continuous-time master equation is known as the *embedding problem* for Markov chains [8–13], which can be viewed as the classical and time-inhomogeneous analog of the *Markovianity problem* studied in quantum information [14, 15].

Despite the ubiquity of master equation dynamics in models of physical systems, it turns out that many P are *non-embeddable*, i.e. they cannot be implemented with any master equation. As we discuss below, examples of non-embeddable matrices include common operations such as bit erasure and the bit flip (a logical NOT)

$$P_{\text{erase}} = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \quad P_{\text{flip}} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}. \tag{1}$$

While neither of these operations are embeddable, there is a crucial difference between them. Bit erasure is infinitesimally close to being embeddable, meaning that it can be implemented by a master equation if one is willing to tolerate some finite, but arbitrarily small, probability of error. We call such matrices *limit-embeddable* (LE).

On the other hand, as we show below, many stochastic matrices P, such as the bit flip, are not even LE. In light of this, suppose we encounter a physical system with n 'visible states' that we know obeys a continuous-time master equation, but is observed to evolve according to a P that is not even LE between $t = 0$ and $t = 1$. In such a situation we can conclude that the master equation dynamics must actually take place over some larger state space, including some unseen *hidden states* in addition to the n visible states.

Alternatively, we can imagine that instead of observing some stochastic dynamics, we attempt to *build* a system that carries out a specified P using master equation dynamics. This is the challenge we might face, for example, if P is the update function of the logical state of a computer we wish to build. In this scenario, the minimal number of hidden states needed to implement P with master equation dynamics emerges as a fundamental 'state space cost' of the computation P. (See also [16].)

In fact, by adding hidden states one can construct master equations that meet more desiderata than just implementing a given P. Specifically, below we show that for any given P and initial distribution $p(0)$, one can construct a master equation that implements P on $p(0)$ while being arbitrarily close to thermodynamically reversible.[6]

In this paper, we establish upper bounds on the number of hidden states required to implement any stochastic matrix P using a master equation, and also establish bounds when we require that P be implemented in a thermodynamically reversible way. We do this using explicit constructions that show any stochastic matrix P can be implemented by composing some appropriate set of fundamental transformations that we call *local relaxations* (LR), defined in detail below.

Our main result is that any $n \times n$ stochastic matrix P can be implemented with no more than $r - 1$ hidden states, where r is the non-negative rank of P. Because $r \leqslant n$, this result implies a simple corollary that one additional binary degree of freedom (which doubles the number of available states) is sufficient to carry out *any* P in a thermodynamically reversible way. We also derive exact results for some particular kinds of P, including those representing single-valued functions, which show that the number of required hidden states can be much smaller than $r - 1$.

Finally, we show that if P can be implemented with some number of hidden states while incurring some nonzero EP, then it can be implemented without any EP at the cost of at most one additional hidden state. Such results imply, for instance, that for a system coupled to a heat bath, adding hidden states can allow one to implement the same P while generating less heat, as we illustrate in detail in an example below.

Note that our results assume complete freedom to use any master equation to implement a given P. However, in the real world there will often be major constraints on the form of the master equation that can be considered, e.g. due to known properties of an observed system we wish to model using a master equation, or due to limitations on what kind of system we can build. An example of the former is if we know that the system's Hamiltonian can only couple degrees of freedom in certain restricted ways. An example of the latter is if P must be implemented using a digital circuit made out of separate gates. In cases where there are such constraints on the form of the master equation, our results can be considered as upper bounds on *lower bounds* of the number of hidden states that are really required. (See section 6.3 below.)

Previous research has shown how to carry out arbitrary P thermodynamically reversibly [18–21]. Moreover, the constructions in those papers can all be formulated in terms of master equation dynamics[7]. In addition, they all exploit what in our terminology are called hidden states. However, none of those papers considered the issue

[6] $p(0)$ must be specified in addition to P since in general, the same master equation dynamics run on different initial distributions $p(0)$ will implement the same P, but with different amounts of irreversible entropy production (EP). See [17].

[7] Specifically, the processes considered in those papers are all (in our terminology) 'LE', and so could be expressed using master equations.

of the minimal number of hidden states needed to implement P, i.e. the state space cost of implementing P, which is the focus of this paper.

In a companion paper [16] we focus on the special case where the stochastic matrix P is 'single-valued', i.e. it represents a deterministic function. It turns out that for that case at least, there is a second kind of cost arising in master equations dynamics that implement P, in addition to the state space cost which is the focus of this paper. Roughly speaking, that second cost is the minimal number of times that the set of allowed state-to-state transitions changes (which in a physics context may correspond to raising or lowering infinite energy barriers between states). This can be viewed as a 'timestep cost' of implementing P. Interestingly, there is a tradeoff between the timestep cost of implementing any (single-valued) P and the state space cost of implementing that P.

This paper focuses on the general problem of bounding the state space cost of arbitrary (not necessarily single-valued) stochastic matrices and the relationship of this cost to thermodynamic reversibility. In the next section we provide relevant background. Then in section 3, we define what it means for a master equation to implement a stochastic matrix P to arbitrary precision. In section 4, we define 'LRs', which are the building blocks of all our constructions. We present our main results in section 5. We also investigate how to extend our framework beyond finite state spaces to countably infinite state spaces, for the special case where P is a single-valued map over X. Since this topic is a bit different from the main focus of the paper, it can be found in appendix F. The other appendices contain proofs that are not in the main text.

2. Background

2.1. Master equations

Consider a physical system with a finite state space X of size n. We write the probability distribution of the state at time t as $p(t)$, where $p_i(t)$ is the probability that the system is in state i at time t. We suppose that $p(t)$ evolves as a time-inhomogeneous continuous-time Markov chain (CTMC)

$$\frac{\mathrm{d}p(t)}{\mathrm{d}t} = M(t)p(t). \tag{2}$$

The elements in each column of the rate matrix $M(t)$ must sum to zero and its off-diagonal entries are positive. The entry $M_{ij}(t)$ is the transition rate from state j to state i at time t. Equation (2) is commonly referred to as 'the master equation'.

For any CTMC, we can relate the distributions at the initial time t and some later time t' by a linear map

$$p(t') = T_M(t,\ t')p(t), \tag{3}$$

where $T_M(t,\ t')$ is known as a *transition matrix*, and equals the time-ordered exponential of $M(t)$. Where $M(t)$ is clear from context, for notational convenience we will sometimes write $T_M(t,\ t')$ simply as $T(t,\ t')$. Note that if $M(t) = M$ is constant, then $T(t,\ t') = \mathrm{e}^{(t'-t)M}$. Finally, note that we can rescale time arbitrarily by multiplying $M(t)$ by an appropriate constant. Therefore, without loss of generality we will take $t = 0$ and $t' = 1$ from now on.

2.2. Embeddability

Only some stochastic matrices P are *embeddable*, meaning that they can be written as $P = T_M(0, 1)$ for some rate matrix $M(t)$. As an illustration of a non-embeddable matrix, consider a bit flip in a two state system, represented by $P_{\mathrm{flip}} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. Now the most general master equation for a two state system is

$$\dot{p}_1(t) = -r(t)p_1(t) + s(t)(1 - p_1(t)), \tag{4}$$

where $p_1(t)$ indicates the probability that the bit is in state 1 at time t, and $r(t)$ and $s(t)$ indicate time-dependent rates of the $1 \to 0$ and $0 \to 1$ transitions, respectively. It can be shown that there is no choice of non-negative functions $r(t)$, $s(t)$ such that the solution of (4) simultaneously satisfies $p_1(1) = 0$ when $p_1(0) = 1$ and $p_1(1) = 1$ when $p_1(0) = 0$, as required to implement P_{flip}. To see why, note that the space of distributions over two states is one-dimensional. The bit flip over a pair of states requires two distinct solutions to a differential equation, with different starting and ending conditions, to cross at some time, which is impossible (see figure 1). (Note that this same argument does not apply to bit erasure.)

Some broadly applicable necessary conditions for a stochastic matrix to be embeddable are known. For example, by Liouville's formula, we have

$$\det T_M(0,\ 1) = \exp\left(\int_0^1 \mathrm{tr}M(t)\ \mathrm{d}t\right). \tag{5}$$

Since the exponential function is strictly positive, any embeddable matrix must have strictly positive determinant [12, 22]. An immediate corollary is that a map like P_{flip} (which has determinant -1) is not embeddable. Indeed,

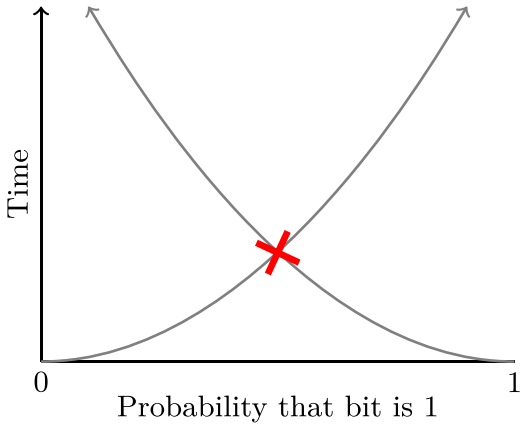

Figure 1. The difference of two solutions to (4) always decreases over time but it can never change sign.

by the continuity of the determinant, P_{flip} is not even infinitesimally close to an embeddable matrix. (In terms of our formal framework introduced below, P_{flip} is not 'LE'.)

It is also known that P must obey $\det P \leqslant \prod_i P_{ii}$ in order to be embeddable [12, 22]. Note that any (non-identity) permutation matrix has determinant that is either 1 or -1, while the product of diagonal entries is 0. Thus, no permutation matrix is embeddable, or even infinitesimally close to an embeddable matrix.

These two necessary conditions on the determinant of a matrix for it to be embeddable are not sufficient. Even if we restrict attention to the time-homogeneous case, where transition rates between states are constant, the problem of whether an arbitrary matrix is embeddable is unsolved for matrices larger than 3×3. (On the other hand, if one imposes the further constraint that the dynamics must obey detailed balance, the time-homogeneous embedding problem has been solved for all finite-size state spaces [13].)

2.3. Entropy production

The field of stochastic thermodynamics [3] has developed a consistent thermodynamic interpretation of master equation dynamics, including a way to quantify EP, the overall increase of entropy in the system and coupled reservoirs. For a physical system evolving according to a master equation like equation (2), we define the instantaneous rate of EP at time t to be [23]

$$\dot{\Sigma}(p(t),\, M(t)) = \sum_{i,j} M_{ij}(t) p_j(t) \ln \frac{M_{ij}(t) p_j(t)}{M_{ji}(t) p_i(t)}\,. \tag{6}$$

The *total EP* over the time interval [0,1], given some initial distribution $p(0)$, is

$$\Sigma(p(0),\, M(t),\, [0,\, 1]) = \int_0^1 \dot{\Sigma}(p(t),\, M(t))\, \mathrm{d}t. \tag{7}$$

The rate of EP is non-negative, and therefore so is total EP. A process that achieves 0 total EP is said to be *thermodynamically reversible*.

If the system is coupled to a single reservoir, a heat bath at temperature T, then the total EP can be written as

$$\Sigma(p(0),\, M(t),\, [0,\, 1]) = [S(p(1)) - S(p(0))] + \frac{1}{kT}\langle Q \rangle, \tag{8}$$

where $S(\,\cdot\,)$ is Shannon entropy, k is Boltzmann's constant, $\langle Q \rangle$ indicates the average heat transferred to the thermal reservoir over all trajectories of states from $t = 0$ to $t = 1$ [23].

To use a well-known example from the thermodynamics of computation [24, 25], suppose we wish to perform a bit erasure (P_{erase}), in which both initial states of a bit, $\{0, 1\}$, get mapped to a single final value, 0. Since this is a many-to-one map, it reduces the entropy of the device (e.g. a computer) in which it is performed. By the second law of thermodynamics, at least as much entropy must be produced in the environment. Often this occurs by the transfer of some amount of heat $\langle Q \rangle$ from the system into a heat bath at temperature T. In this case, assuming the initial probability over the states of the bit is uniform, the second law yields $\langle Q \rangle \geqslant kT \ln 2$, which is the well-known bound on the amount of heat produced in bit erasure that Landauer derived using semi-formal reasoning[8]. More generally, an immediate consequence of equation (8) and the non-negativity of EP is

[8] It is now understood that a logically irreversible process like bit erasure can be done in a thermodynamically *reversible* manner [26], despite how many have interpreted Landauer's reasoning.

$$\langle Q \rangle \geqslant kT[S(p(0)) - S(p(1))], \tag{9}$$

where $p(0)$ and $p(1)$ are the distributions over system states at the beginning and end of the physical process under consideration.

In the previous section, we discussed the embedding problem, which considers what P can be implemented using a continuous-time master equation. A related, thermodynamically motivated, question concerns which P can be implemented using a master equation that achieves 0 total EP, in the case where there is a single, isothermal heat bath. Generally, EP vanishes in such a situation if and only if $p(t)$ is an equilibrium distribution of $M(t)$ for all $t \in [0, 1]$. Formally, this means that $M_{ij}(t)p_j(t) = M_{ji}(t)p_i(t)$ for all i, j and $t \in [0, 1]$. In general, this condition requires that the initial distribution $p(0)$ is an equilibrium distribution and that we go to the 'quasistatic limit' where the parameters change infinitesimally slowly in relation to the relaxation time of the system [27–29]. This limit can be achieved either by implementing P increasingly slowly (in 'wall clock' time), and/or by making the transition rates (which control the relaxation time) increasingly large.

While a given CTMC $M(t)$ may achieve zero (more accurately, arbitrarily small) total EP for some initial distribution $p(0)$, in general it cannot generate zero EP for more than one such initial distribution. (See [17] for a detailed investigation of the dependence of total EP on the initial distribution.) So our thermodynamically motivated question is whether a given P has the property that for all initial distributions $p(0)$, there is some associated set of $M(t)$ that implement P with zero total EP, for that particular initial distribution $p(0)$.

3. Definitions

3.1. Limits of CTMCs

As discussed in the section 1, many common operations such as bit erasure are not embeddable (e.g. no CTMC has $T(0,1) = P_{\text{erase}}$) but are nonetheless infinitesimally close to a stochastic matrix that is embeddable. This motivates the following definition:

Definition 1. A stochastic matrix P is LE if for all $\epsilon > 0$, there is a rate matrix $M(t)$ such that $|T_M(0, 1) - P| < \epsilon$.

Our definition of limit embeddable makes no restrictions on the thermodynamic properties of the implementation of a given P. Our next task is to formalize an additional restriction that reduces the set of all LE matrices to a subset that can be implemented in a thermodynamically reversible manner. Rather than focus on whether some P can be implemented that way for one specific initial distribution, we focus on whether it can be implemented that way for *any* initial distribution. This motivates the following definition:

Definition 2. A stochastic matrix P is *quasistatically embeddable* (QE) if for all initial distributions p, all $\epsilon > 0$, and all $\delta > 0$, there is a rate matrix $M(t)$ with $|T_M(0, 1) - P| < \epsilon$ and $\Sigma(p, M(t), [0, 1]) < \delta$.

We emphasize that the rate matrix $M(t)$ that implements a given P must generally be specialized to the initial distribution p in order to make $\Sigma(p, M(t), [0, 1])$ arbitrarily small.

There are a few properties of QE matrices that are central to our analysis. The first one is the following lemma:

Lemma 1. *The set of QE matrices is closed under multiplication.*

It is impossible to implement any matrix without EP except in a limit, as illustrated by the next proposition[9]:

Proposition 2. *If $M(t)$ is the rate matrix of a CTMC with $T(0, 1) = P \neq I$, then for any initial distribution p*

$$\Sigma(p, M(t), [0, 1]) \geqslant -\frac{|Pp - p|_1^2}{2\ln(\det P)},$$

where $|\cdot|_1$ is ℓ_1 norm[10].

The inequality in proposition 2 (which may not be tight) shows that only those P with determinants very close to 0 can achieve small EP for arbitrary initial distributions. Recall that by equation (5), $\det P = \exp\left(\int_0^1 \mathrm{tr}M(t)\,\mathrm{d}t\right)$. This means that a small determinant can be achieved if the transition rates

[9] While preparing this manuscript we became aware of independent result by Shiraishi *et al* [30], which implies proposition 2; see appendix B for details.

[10] For two probability distributions p, q over n states, the ℓ_1 norm is defined $|p_i - q_i| := \sum_{i=1}^n |p_i - q_i|$.

(the off-diagonal entries of $M(t)$) are very large (so that $\mathrm{tr}M(t)$ is very negative). Note that since we have scaled time out to fix the final time $t' = 1$, we are effectively measuring rates in units of $1/t'$. Therefore, the observation that the entries of $M(t)$ must be very large to achieve small *EP* is consistent with the well-known fact that quasistatic processes take an infinite amount of time. (See also our discussion of the quasistatic limit in the previous section.)

An important corollary of proposition 2 is that *any* QE matrix must be singular.

Corollary 3. *Any QE matrix, except the identity, has determinant zero.*

These points are illustrated in the following example of a canonical QE process, bit erasure, implemented with a quantum dot:

Example 1. In the model of bit erasure described in [31], a classical bit is implemented as a quantum dot, which can be either empty (state 0) or filled with an electron (state 1). The dot is brought into contact with a metallic lead at temperature T which (for an appropriate state of the dot) may transfer an electron into the dot or out of it, thereby changing the state of the bit.

At time t, the propensity of the lead to give or receive an electron is set by its chemical potential, indicated by $\mu(t)$, and the energy of an electron in the dot, indicated by $E(t)$. Specifically, let $p(t)$ indicate the two-dimensional vector of probabilities, with $p_0(t)$ and $p_1(t)$ being the probability of an empty and full dot, respectively. These probabilities evolve according to the rate matrix [31]

$$M(t) = C\begin{bmatrix} -w(t) & 1 - w(t) \\ w(t) & -(1 - w(t)) \end{bmatrix}, \tag{10}$$

where C sets the timescale of the exchange of electrons between the dot and the lead and $w(t)$ is the Fermi distribution of the lead

$$w(t) = [\exp((E(t) - \mu(t))/k_B T) + 1]^{-1}.$$

Using equations (2), (10) and conservation of probability (i.e. $p_0(t) + p_1(t) = 1$), we can write

$$\dot{p}_1(t) = C(w(t) - p_1(t)). \tag{11}$$

Suppose that $\mu(t)$ and $E(t)$ are chosen in a way that depends on $p_1(0)$, so that $w(t) = (1 - t)p_1(0) + t\delta$ for some constant δ. In this case, (11) can be explicitly solved for p_1:

$$p_1(t) = w(t) + C^{-1}(p_1(0) - \delta)(1 - \mathrm{e}^{-Ct}). \tag{12}$$

In the limit where $C \to \infty$, $p_1(1) \to w(1) = \delta$, so the transition matrix $T(0,\ 1)$ becomes

$$T(0,\ 1) = \begin{bmatrix} 1 - \delta & 1 - \delta \\ \delta & \delta \end{bmatrix}.$$

Furthermore, by equation (12), in that limit $p_1(t) \to w(t)$, and so $p_0(t) \to 1 - w(t)$. That means that for all i, j, $M_{ij}(t)p_j(t) \to M_{ji}(t)p_i(t)$. In this limit, the total EP over the course of the process

$$\Sigma = \int_0^1 \mathrm{d}t \sum_{i,j\in\{0,1\}} M_{ij}(t)p_j(t)\ln\left[\frac{M_{ij}(t)p_j(t)}{M_{ji}(t)p_i(t)}\right],$$

approaches zero (we show this rigorously, in the context of proving a more general result, in appendix D).

Since δ is arbitrary, this means that we can make $T(0,\ 1)$ arbitrarily close to bit erasure ($\delta = 0$), while having arbitrarily small total EP. This establishes that bit erasure is QE.

Note that if we cannot control C directly, we can still achieve the same effect as the limit $C \to \infty$ by running the process with some fixed C for longer and longer times (that is, by changing the endpoint from $t = 1$ to some $t > 1$). This demonstrates the equivalence of going to the limit of infinitely-long time versus infinitely-fast rates for achieving vanishing EP.

3.2. Embedding with hidden states

Many stochastic matrices P are not QE. Our main result is that despite this, any P is a (principal) submatrix of a larger matrix $\tilde{P}$ which *is* QE. We formalize this property as follows:

Definition 3. An $n \times n$ stochastic matrix P is QE *with m hidden states* if there exists some $(n + m) \times (n + m)$ matrix $\tilde{P}$ that is QE, and for all $i,\ j \in 1 \ldots n$, $\tilde{P}_{ij} = P_{ij}$ (i.e. P is a principal submatrix of $\tilde{P}$).

To understand the motivation for this definition, imagine that there is a matrix $\tilde{P}$ over state space Y that is QE, and that P is a principal submatrix of $\tilde{P}$, corresponding to the subset of states $X \subseteq Y$. If at $t = 0$ the process

that implements $\tilde{P}$ is in some state $i \in X$, then the distribution over the states in X at the end of the process at $t = 1$ will be exactly as specified by P. Furthermore, because $\sum_{j \in X} P_{ji} = 1$, $\tilde{P}_{ji} = 0$ for any $i \in X$ and $j \notin X$. This means that if the process is started on some $i \in X$, no probability can 'leak' out into the hidden states by the end of the process, although it may pass through them at intermediate times.

Mathematically, the hidden states in definition 3 are additional to the states whose evolution is controlled by P. However, there are multiple ways to map this mathematical structure into a particular physical system. In particular, hidden states can arise when Y is a set of microstates of a system and X is an associated set of macrostates; we elaborate this point in section 5.

Note that any P that is QE with m hidden states is also QE with p hidden states for all $p \geqslant m$. Note as well that by lemma 1, if an $n \times n$ stochastic matrix P can be factored into a product of $n \times n$ matrices, each of which is QE with m hidden states, then P itself is QE with m hidden states. More generally, the number of hidden states required to quasistatically embed a product of stochastic matrices is no more than the maximum number required for each of the matrices in that product. In section 5, we exploit this fact to derive upper bounds on the minimal number of hidden states needed to quasistatically embed a given matrix.

The definition of QE may seem very strict, but from the perspective of the number of hidden states needed for embedding, it is only slightly more costly than limit-embedding:

Proposition 4. *If a matrix P is LE, it is QE with at most one hidden state.*

This means the minimal number of hidden states required for limit-embedding versus quasistatic embedding are at most one apart.

Note that it might be fruitful to consider a weaker definition of QE than definition 2, in which a matrix would be considered 'QE' if it is LE while achieving vanishingly small EP for *some* specific initial distribution p, rather than requiring that it is LE that way for *any* initial distribution. With this definition, the 'state space cost' given by the number of hidden states would then be a function of P and the initial distribution p, not just of P. Our definition can be seen as a worst-case version of this weaker definition; we measure the 'state space cost' of a stochastic matrix P as the maximal number of hidden states we might need if we were given some specific initial distribution p and wanted to construct a master equation that implements P with no EP for that p.

In addition, we note that any matrix which is QE according to our stronger definition would also satisfy this weaker definition of QE. Thus, the upper bounds on the minimal number of hidden states required to quasistatically embed a given P (according to definition 2) which we derive below are also upper bounds for the number of hidden states required under the alternative weaker definition of QE. Note also that corollary 3 holds under the weaker definition, as long as the desired initial distribution p is not the fixed point of P.

4. Local relaxations

In example 1 we showed that bit erasure is QE. In fact, bit erasure is part of a much larger family of stochastic matrices which can be shown to be QE, which we call 'local relaxations'. These will serve as the 'building blocks' of the constructions we use below:

Definition 4. A stochastic matrix P is a LR if there is some permutation matrix Q (which may be the identity) such that:

1. QPQ^{-1} has a block diagonal structure,
2. Each block of QPQ^{-1} has rank one.

Note that the role of the permutation matrix Q is to just rearrange rows and columns, i.e. to relabel states.

Loosely speaking, when a LR matrix is used to evolve a system, states of the system are grouped into different 'blocks' between which no probability flows, while the states within each block relax to the same final, block-specific distribution.

Example 2. The following 4×4 stochastic matrices are all LRs (for each matrix, we assume that a, b, c, d are chosen to be non-negative and that the sum of each column equals 1):

$$A = \begin{bmatrix} a & a & a & a \\ b & b & b & b \\ c & c & c & c \\ d & d & d & d \end{bmatrix}; B = \begin{bmatrix} a & a & 0 & 0 \\ b & b & 0 & 0 \\ 0 & 0 & c & c \\ 0 & 0 & d & d \end{bmatrix}; C = \begin{bmatrix} a & 0 & a & 0 \\ 0 & c & 0 & c \\ b & 0 & b & 0 \\ 0 & d & 0 & d \end{bmatrix}.$$

The block structure in matrices A and B is immediately apparent. Matrix C is a LR since $B = QCQ^{-1}$, where Q is the permutation matrix

$$Q = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

(In other words, C can be arranged to have the block structure of B by switching rows/columns 2 and 3.)

In appendix D, we prove the following result, which establishes that we can use LR matrices as building blocks to construct QE matrices:

Proposition 5. *Any LR is QE.*

By performing one LR followed by another one involving a different partition of X into blocks, we can quasistatically implement matrices that are not themselves LRs. This means the converse of proposition 5 is false—not every QE matrix is a LR.

However, it *is* the case that any 2×2 QE matrix P is a LR. If P is the identity, it is a LR. If not, then by corollary 3, its determinant is zero, which implies it has rank one and so is a LR (with a single block).

Since they are all QE, products of LR matrices have determinant zero (again, except for the identity). On the other hand, not all singular matrices are products of LRs. To see this, note that a (non-identity) product of LRs must always send at least two initial states to the same final distribution (since there must at least one LR in the product with a block of size larger than one). So such a product must have at least two identical columns. Therefore, a matrix like

$$\begin{bmatrix} 1 & 0 & 1/2 \\ 0 & 1 & 1/2 \\ 0 & 0 & 0 \end{bmatrix}$$

is not a product of LRs, even though it is singular.

To establish the results in the next section, we will repeatedly make use of proposition 5 and lemma 1 to prove that a given matrix is QE by writing it as a finite product of LRs.

One might conjecture that a matrix is QE only if it is a product of LRs. However, we do not establish this, and our results below do not rely on it.

5. Upper bounds on minimal number of hidden states

5.1. Single-valued maps

We refer to a stochastic matrix that represents a deterministic function (i.e. all its entries are either 0 or 1) as a *single-valued map*.

To begin, note that if a single-valued map P corresponds to an invertible function, then P is a permutation and is not LE, as discussed in section 1. If, on the other hand, a single-valued map P corresponds to a non-invertible function, then it has determinant 0. Moreover, we can use LR matrices to establish that any such P is LE, and in fact QE:

Proposition 6. *Any single-valued map with determinant zero is QE.*

Proof. Consider a single-valued map P over some finite set X with determinant zero, which represents some non-invertible function $f: X \to X$. It is known that any non-invertible function $f: X \to X$ is a composition of finitely many idempotent $X \to X$ functions[11], $f = f_N \circ \ldots f_2 \circ f_1$ [32]. In our context, this means that P can be represented as a product of N single-valued maps $P = \prod_k M^{(k)}$, each $M^{(k)}$ representing the idempotent function f_k. Note that the image of an idempotent function consists entirely of fixed points, thus each $M^{(k)}$ can be written as

[11] A function is *idempotent* if applying it twice is the same as applying it once, $f \circ f = f$.

$$M^{(k)}_{ij} = \begin{cases} 1 & \text{if } j \in f_k^{-1}(i) \text{ and } i = f_k(i) \\ 0 & \text{otherwise} \end{cases}.$$

It can be verified that any such $M^{(k)}$ is a LR, since it consists of blocks corresponding to the set of pre-images $\{f^{-1}(i)\colon i \in f(X)\}$, and each block $f^{-1}(i)$ contains identical columns: all 0s, except for a 1 in the row corresponding to i.

The result then follows by applying first proposition 5 and then lemma 1. □

While a single-valued map with nonzero determinant is not QE (by corollary 3), it can always be made QE by adding a single hidden state. Before establishing this for the general case, we illustrate the basic proof technique with an example showing that the bit flip is QE with one hidden state:

Example 3. The bit flip $P_{\text{flip}} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ is a map over a binary space X. It is not QE because it has negative determinant. However, it *is* QE with one hidden state.

This follows by constructing a special set of three LR matrices, each defined over a space Y that has three elements, such that the restriction to the first two elements of the product of those matrices is the bit flip:

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}. \tag{13}$$

The restriction of the matrix on the lhs to its first two elements (i.e. the upper left block) is the bit flip operation, by inspection. In addition, the first two matrices on the rhs are LR, by inspection. The third matrix on the rhs is LR as well, but to see that we need to relabel the elements of Y in such a way that that matrix becomes block-diagonal. Specifically, if we permute the second and the third elements of Y, we transform the third matrix as:

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 0 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

which confirms that the rightmost of the three matrices is LR.

So the rhs is a product of three LR matrices, and therefore the lhs must be QE. This confirms that the bit flip is QE with one hidden state.

As an aside, note that the matrix on the lhs is not itself LR, even though it is a product of three LR matrices. (This follows by verifying that there is no relabeling of the elements of Y that changes the matrix on the lhs of equation (13) into block-diagonal form.)

We can easily generalize this technique to establish that any transposition is QE with one hidden state. Since any permutation can be written as a product of transpositions, and since the product of QE matrices is QE, the next result follows immediately:

Proposition 7. *Any permutation is QE with one hidden state.*

Together corollary 3 and proposition 7 imply that any permutation matrix requires exactly one hidden state to be QE.

As mentioned in section 1, it is possible to extend our analysis of single-valued maps to countably infinite spaces X, for suitable extensions of our definitions. Appendix F introduces one such extension of our definitions, and then proves that with those extended definitions, any single-valued function over a countably infinite X can be implemented with at most one hidden state.

5.2. General case

We now present our main result for arbitrary stochastic matrices. To begin, recall that the *non-negative rank* of an $n \times n$ stochastic matrix P is the smallest m such that P can be written as $P = RS$, where R is an $n \times m$ stochastic matrix and S is an $m \times n$ stochastic matrix [33]. Roughly speaking, the non-negative rank of a matrix is analogous to the number of independent mixing components in a mixture distribution.

Theorem 8. *An $n \times n$ stochastic matrix P with non-negative rank r is QE with $r - 1$ hidden states.*

To understand this bound intuitively, suppose we have decomposed a given $n \times n$ stochastic matrix P into a product of two rectangular stochastic matrices of dimensions $n \times r$ and $r \times n$. These two rectangular matrices can be interpreted as representing transfers of probability between disjoint sets of states: the first matrix transfers probability from the n original states to r hidden ones, and the second map transfers probability back to the n

original states. These transfers of probability can be shown to be QE, establishing that no more than r hidden states are needed to implement P in a QE manner.

As we show in appendix E, it is possible to reduce this number of hidden states needed by one, which yields theorem 8.

Since the non-negative rank of an $n \times n$ matrix is n or less, theorem 8 implies that any $n \times n$ stochastic matrix is QE with at most $n - 1$ hidden states. This is fewer than the number of hidden states provided by adding a single independent, extra binary degree of freedom to the system, since adding such a bit doubles the size of the state space. Thus, simply by adding a hidden bit to a system, we can implement any stochastic matrix in a thermodynamically reversible manner (presuming we have freedom to use arbitrary sequences of LR matrices to implement the stochastic matrix).

Recall from proposition 2 that some stochastic matrices P cannot be directly implemented with zero EP, even if they are embeddable. Theorem 8 tells us that it is sometimes possible to reduce EP of implementing an embeddable P by adding hidden states. This is illustrated with the following example.

Example 4. The 'partial' bit flip

$$P = \begin{bmatrix} 2/3 & 1/3 \\ 1/3 & 2/3 \end{bmatrix}$$

is embeddable [9] but has nonzero determinant ($\det P = 1/3$), and so cannot be QE by corollary 3. Specifically, for any initial probability distribution $p(0)$ (except possibly the fixed point of P, $p(0) = (1/2,\ 1/2)^T$), there is some unavoidable EP. For example, by proposition 2, for the initial distribution $p(0) = (1,\ 0)^T$,

$$\Sigma \geqslant -\frac{4/9}{2 \ln 1/3} \approx 0.2.$$

However, by theorem 8, P is QE with one hidden state. Thus, by adding a hidden state, the partial bit flip can be carried out while achieving zero EP. Note that since the change in the Shannon entropy $S(Pp(0)) - S(p(0))$ is independent of how P is implemented, the reduction in EP realized when implementing P using a hidden state implies a decrease in the amount of entropy produced in the environment (e.g. as a reduction in generated heat).

As a final comment, while the non-negative rank provides a general bound on the minimal number of hidden states needed to quasistatically embed a given matrix, other properties of the matrix can sometimes be used to further reduce the bound. As a simple illustration, suppose that we know the non-negative rank of some stochastic matrix P—but also know that P is block diagonal (up to a rearrangement of rows/columns by some permutation matrix), and that the greatest non-negative rank of any of the blocks is k. Now we can implement P by implementing each block in P independently, one after the other. Implementing P this way would allow us to repeatedly 'reuse' whatever hidden states we have, for each successive implementation of a block. This means that P is QE using $k - 1$ hidden states. This is true even if the non-negative rank of P is (substantially) larger than k, in which case the direct application of theorem 8 would give a much weaker bound on the minimal required number of hidden states.

5.3. Coarse-grained states

When introducing our notion of hidden states, we indicated that our upper bounds would also apply if the stochastic matrix P described the evolution of probability over some number m of coarse-grained 'macrostates' which collectively have internal to them n 'microstates'. To see this, it suffices to notice that a single LR could send all the microstates associated to a macrostate to one of its microstates. This concentrates probability in as many microstates as there are macrostates, and the resulting 'empty' $m - n$ microstates can now be used as hidden states in the sense of our definitions and constructions.

For example, consider a system with three microstates $\{a,\ b,\ c\}$ which are grouped into two macrostates $0 = \{a\}$ and $1 = \{b,\ c\}$, representing the two states of a bit, that we wish to flip. This can be done by first applying the LR $\{a\} \mapsto a,\ \{b,\ c\} \mapsto b$, leaving c with no probability. Now, by carrying out the bit flip operation on microstates a and b using c as a hidden state (as shown in example 3), we carry out a bit flip over the macrostates 0 and 1.

Table 1. Minimum number of hidden states required to limit-embed (LE) or quasistatically embed (QE) a given stochastic matrix P. r is the non-negative rank of P.

Type of matrix P	Min. # hidden states		
Single-valued			
Invertible ($\neq I$)	1		
Non-invertible	0		
Noisy	Lower bound (LE)	(QE)	Upper bound (QE)
$\det P = 0$	0	0	$r-1$
$\det P > 0$	0	1	$r-1$
$\det P < 0$	1	1	$r-1$
$\det P > \prod_i P_{ii}$	1	1	$r-1$
Limit embeddable	0	0	1

6. Discussion

In this paper, we consider implementing stochastic matrices P using a master equation. For some P, this is not possible. However, we show that it is always possible to implement any P (to arbitrary accuracy), if we have sufficiently many hidden states, in addition to the visible states that P works on. We also show that it is always possible not only to implement P (to arbitrary accuracy), but to do so in a thermodynamically reversible manner by using hidden states. The minimal number of required hidden states for such thermodynamically reversible implementation of a given stochastic matrix P is a novel and fundamental kind of 'state space cost' of carrying out P.

We go on to derive some bounds on the minimal number of such hidden states required for any given P, either just to implement it, or to do so in a thermodynamically reversible manner. In particular, we derive bounds on this minimal number of states needed for these two kinds of implementation, stated in terms of basic properties of P.

Our results for different kinds of P are summarized in table 1, which lists upper and lower bounds for both limit-embedding (i.e. implementing to arbitrary accuracy) and quasistatic embedding (i.e. implementing to arbitrary accuracy in a thermodynamically reversible manner) for different classes of matrices. Any QE matrix is LE, so the upper bounds established for quasistatic embedding also apply to limit-embedding. All the lower bounds in this table are tight, in the sense that in each class of matrices listed, there is a matrix that satisfies the lower bound.

6.1. Interpretation of hidden states

Often in physics there are some states of a system or even entire physical variables that are hard to observe and control experimentally. Such 'hidden states' are often a problem to be circumvented, e.g. by coarse-graining. Whatever method we use for dealing with hidden states invariably affects the predictions we make (e.g. coarse-graining can result in entropy increasing with time). However, such methods can allow us to proceed in an analysis despite our incomplete knowledge.

At other times, the presence of hidden states can be a crucial property of a system, necessary for us to use a master equation to model the dynamics. Our results highlight the extent to which this is the case in different situations. In some cases, hidden states cannot be ignored—an engineer designing a physical system to implement a given P must ensure that there is appropriate dynamic coupling between the hidden states and the visible ones that P operates on, either implicitly or explicitly. In a different context, the number of hidden states are a measure of how many internal states are being overlooked by a scientist who notices that some naturally occurring system evolves (over discrete time) according to P, and wants to presume that there is a master equation dynamics underlying that evolution.

It is important to emphasize that the role for hidden states uncovered in our analysis is different from the role of the states of the history tape in Bennett's reversible computation construction [25, 34, 35], or the states of the extra bits in reversible Toffoli gates [36]. Like hidden states, the states used in reversible computation 'facilitate' the desired dynamics P, in a broad sense, without being in the space that P runs over. However, the role of those states in reversible computation is to allow logical reversibility of the conditional distributions giving the update rules, and therefore to lower the minimal thermodynamic work (rather than the EP) required to perform certain

computations. In contrast, the conditional distributions we construct here are made from LR matrices and are not logically reversible in general.

Indeed, the state space cost of a conditional distribution P in some ways behaves in a manner 'opposite' to the costs of P discussed in the early thermodynamics of computation literature. A logically reversible function needs a hidden state to be implementable by a CTMC, while a function that is logically irreversible does not. Therefore, as far as the number of hidden states is concerned, there are advantages to being logically irreversible rather than being logically reversible.

6.2. Biochemical oscillations

One possible application of our results is to recent stochastic thermodynamics analyses of 'Brownian clocks'[37, 38]. These are biochemical oscillations in which a component (e.g. a protein) undergoes Markovian transitions (governed by some rate matrix $M(t)$) through a cycle of discrete states[12].

Such clock-like oscillations are forbidden at equilibrium. However, they can be sustained out of equilibrium two ways: By fixed driving forces ($M(t) = M$ is constant but not detailed balanced), or by periodic driving (time-dependent variation of $M(t)$).

In the former case, oscillations invariably dephase, though coherence can be preserved over many periods if the system is driven strongly and has many states [38]. Even in the latter, more general case, it is challenging to preserve phase information. In particular, if we try to design a Brownian clock that remembers its phase exactly, we run into precisely the embedding problem discussed in this paper. For example, a clock with three states that remembers its phase (perfectly) must transform its own states according to the cyclic permutation:

$$P_{\text{tick}} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

each time it ticks. As we observe in section 1, this is impossible (since the product of the diagonal entries of P_{tick} is less than its determinant [12, 22]). In fact, we cannot even approximate these clock dynamics by allowing an arbitrarily small amount of error. For example, a 'lazy' clock $P_{\text{lazy}} = (1 - \epsilon)P_{\text{tick}} + \epsilon I$, which fails to tick with probability ϵ and slowly dephases, is not embeddable.

However, as we show above, all permutations, including P_{tick}, are QE with one hidden state, which means that if there is a hidden state in the clock (which may be occupied between ticks) it can be arbitrarily precise and remember its phase, while producing arbitrarily little entropy. Our more general results (e.g. theorem 8) say that periodic driving can be used to implement any variant of the stochastic matrix P_{tick} (e.g. modeling a three state clock that loses coherence in a specific way or rate), as long as at least two hidden states are available.

6.3. Future technical directions

Many of our results establish upper bounds on the number of hidden states required to quasistatically embed a given stochastic matrix P. We generally do not know how tight these upper bounds are, even in the worst cases. In fact, we have no example of a matrix which requires more than one hidden state for limit-embedding or quasistatic embedding. One strategy to prove lower bounds would be to show that all QE matrices can be written as products of LRs, and then improve our arguments to establish that the number of hidden states used in our constructions (or modifications thereof) is minimal. This is a natural target for future work.

We are also keen to explore the potentially fruitful connections between the classical questions we have tackled in this paper and questions arising in the study of open quantum systems. The state of such systems can be described by a density matrix, and the transformation of state between two times is given by a completely positive trace-preserving map, or 'quantum channel', which is the quantum analog of a stochastic matrix. The analog of the classical master equation in this context is the Lindblad master equation [39], and the analog of the embedding problem is known as the Markovianity problem [14]. Prior work related to our concerns in this paper include [40], which studies the relationship between time-homogeneous and inhomogeneous embedding in the quantum context, as well the results in [41] on mapping certain non-Lindblad master equations to systems obeying a Lindblad master equation coupled to a single extra 'ancillary' qubit.

Another possible direction involves pursuing continuous space generalizations of our results. Although master equations over a finite number of states are well-suited to represent many systems at some scales, continuous state spaces unavoidably appear in microscopic, classical descriptions. One way to handle this would be to introduce a fine discretization of space (see e.g. [42]) to turn say, diffusion in a compact region into a finite-state master equation. Although our results presented here would apply to such an approximation, they would not be addressing the natural questions for such a context, because they ignore the continuous structure of state

[12] Note that the rate equation of a monomolecular chemical reaction network is also of the form of equation (2), so the same mathematics can arise with a different interpretation.

space. For example, we have assumed throughout that the entries of $M(t)$ can be controlled freely and entirely independently of each other. In a discretized model of a continuous system, this would be like supposing that force fields or diffusivities could vary arbitrarily over the (very small) discretization scale. As an alternative, one could consider the embedding problem for Fokker–Planck equations, characterizing the set of conditional probability density functions that can be implemented with Fokker–Planck equations, and seeing whether adding additional dimensions can enlarge this set.

Of course, there is also an 'intermediate' regime, in which the state space is infinite, but still countable. We present a very preliminary investigation of that regime in appendix F. Some possibly fruitful future work would involve extending the investigations there, e.g. to apply to noisy P as well as single-valued P. It might also be fruitful to consider alternatives to the ways that the analysis in appendix F extends our basic concepts of 'implementing P' from the finite state space case to the infinite state space case.

Even in naturally discrete systems, our ability to independently control transition rates between states could be constrained, which could give rise to state space costs greater than the very modest upper bound (one hidden bit) we establish here for any matrix P, and deserve further exploration.

Our proofs involve implementing a given conditional distribution P via a sequence of discrete steps—LRs. The number of such discrete steps required to implement a given P presents an interesting way of measuring the difficulty of physically implementing a stochastic matrix. Analyzing the number of steps required to implement a given P, and how this number depends on the number of hidden states and the details of P, is an important direction for future work. As a initial result in this direction, we note that our construction in the proof of theorem 8 requires $4rn - 10$ LRs to be performed in sequence. Such investigations are closely linked to the state space size versus timestep tradeoff for single-valued stochastic matrices, which we analyze in a companion paper [16]. (See also work on a related question for the special case $n = 3$ in [43].)

Acknowledgments

We would like to thank Max Tegmark for useful discussions, and the Santa Fe Institute for helping to support this research. This paper was made possible through the support of Grant No. FQXi-RFP-1622 from the FQXi foundation, and Grant No. CHE-1648973 from the US National Science Foundation.

Appendix A. Proof of lemma 1

Lemma 1. *The set of QE matrices is closed under multiplication.*

Proof. If $T(t,\ t')$ is the transition matrix associated with rate matrix $M(t)$ and $S(t,\ t')$ is the transition matrix associated with $N(t)$, then $S(1/2,\ 1)T(0,\ 1/2) = U(0,\ 1)$, where $U(t,\ t')$ is associated with rate matrix:

$$L(t) = \begin{cases} M(t) & \text{if } t \in [0,\ \frac{1}{2}] \\ N(t) & \text{if } t \in (\frac{1}{2},\ 1] \end{cases}.$$

To complete the proof note that the EP over the whole process is the sum of EP over both subintervals. □

Appendix B. Proof of proposition 2

Proposition 2. *If $M(t)$ is the rate matrix of a CTMC with $T(0,\ 1) = P \neq I$, then for any initial distribution p*

$$\Sigma(p,\ M(t),\ [0,\ 1]) \geqslant -\frac{|Pp - p|_1^2}{2\ln(\det P)},$$

where $|\cdot|_1$ is ℓ_1 norm[13].

Proof. Let $p(0)$ indicate the initial state distribution, and $p(1) = Pp(0)$ indicate the final state distribution under P.

Consider a CTMC with rate matrix $M(t)$ that implements P over $t \in [0,\ 1]$. The probability flux from state j to i at time t is

$$J_{ij}(t) \coloneqq p_j(t)M_{ij}(t) - p_i(t)M_{ji}(t).$$

[13] For two probability distributions p, q over n states, the ℓ_1 norm is defined $|p - q|_1 \coloneqq \sum_{i=1}^{n}|p_i - q_i|$.

Write the instantaneous EP, from equation (6) and [23], as

$$\dot{\Sigma}(p(t)) = \sum_{i,j:J_{ij}(t)>0} J_{ij}(t)\ln\frac{p_j(t)M_{ij}(t)}{p_i(t)M_{ji}(t)}, \tag{B1}$$

where we have used the symmetry property

$$J_{ij}(t)\ln\frac{p_j(t)M_{ij}(t)}{p_i(t)M_{ji}(t)} = J_{ji}(t)\ln\frac{p_i(t)M_{ji}(t)}{p_j(t)M_{ij}(t)}.$$

Define $\alpha(t) \coloneqq -\mathrm{tr}(M(t))$ to be the negative of the trace of the rate matrix at time t. We then bound the rhs of equation (B1) using $\ln x \geqslant 1 - \frac{1}{x}$,

$$\begin{aligned}
\dot{\Sigma} &\geqslant \sum_{i,j:J_{ij}(t)>0} J_{ij}(t)\frac{p_j(t)M_{ij}(t) - p_i(t)M_{ji}(t)}{p_j(t)M_{ij}(t)} \\
&= \sum_{i,j:J_{ij}(t)>0} \frac{(p_j(t)M_{ij}(t) - p_i(t)M_{ji}(t))^2}{p_j(t)M_{ij}(t)} \\
&= \frac{1}{2}\sum_{i,j}\frac{J_{ij}(t)^2}{p_j(t)M_{ij}(t)} \\
&= \frac{1}{2}\sum_{i,j}\frac{\alpha(t)^2}{p_j(t)M_{ij}(t)}\left(\frac{J_{ij}(t)}{\alpha(t)}\right)^2 \\
&\geqslant \frac{1}{2}\sum_{i,j}\alpha(t)\left(\frac{J_{ij}(t)}{\alpha(t)}\right)^2,
\end{aligned} \tag{B2}$$

where in the last inequality we used that $p_j(t) \leqslant 1$ and $\frac{M_{ij}(t)}{\alpha(t)} \leqslant 1$.

Before proceeding, we note that

$$\begin{aligned}
|p(1) - p(0)|_1 &= \sum_i |p_i(1) - p_i(0)| \\
&= \sum_i \left| \int_0^1 \sum_j J_{ij}(t)\,\mathrm{d}t \right| \leqslant \sum_{i,j}\int_0^1 |J_{ij}(t)|\,\mathrm{d}t.
\end{aligned} \tag{B3}$$

For convenience, we also define

$$K \coloneqq \int_0^1 \alpha(t)\,\mathrm{d}t = -\int_0^1 \mathrm{tr}\,M(t)\,\mathrm{d}t = \ln(\det P), \tag{B4}$$

where the last equality is equation (5).

We now use equations (B3), (B2) to bound the integrated EP, $\Sigma \coloneqq \int_0^1 \dot{\Sigma}\,\mathrm{d}t$.

$$\begin{aligned}
\Sigma &\geqslant \frac{1}{2}\int_0^1 \sum_{i,j}\alpha(t)\left(\frac{J_{ij}(t)}{\alpha(t)}\right)^2 \mathrm{d}t \\
&= \frac{K}{2}\int_0^1 \sum_{i,j}\frac{\alpha(t)}{K}\left(\frac{J_{ij}(t)}{\alpha(t)}\right)^2 \mathrm{d}t \\
&\geqslant \frac{K}{2}\left(\int_0^1 \sum_{i,j}\frac{\alpha(t)}{K}\left|\frac{J_{ij}(t)}{\alpha(t)}\right| \mathrm{d}t\right)^2 \\
&= \frac{K}{2}\left(\frac{1}{K}\int_0^1 \sum_{i,j}|J_{ij}|\,\mathrm{d}t\right)^2 \\
&\geqslant \frac{1}{2K}\,|p(1) - p(0)|_1^2,
\end{aligned}$$

where the third line follows by Jensen's inequality. The proposition follows immediately by rearranging. □

Corollary 3, that any non-identity QE matrix P is singular, follows directly from proposition 2. Choose p to be any non-fixed-point of P. Then, in order for EP to be made arbitrarily small, as required by the definition of QE, it must be that $\det P = 0$.

Note also that a related result has recently been derived in [30]. Equation (14) in that paper, with minor rearranging, states

$$\Sigma(p(0),\, M(t),\, [0,\, \tau]) \geqslant \frac{|p(\tau) - p(0)|_1^2}{2\tau \langle A \rangle_\tau}, \tag{B5}$$

where $\langle A \rangle_\tau$ is the so-called 'dynamical activity'

$$\begin{aligned}\langle A \rangle_\tau &:= \frac{1}{\tau} \int_0^\tau \sum_{i \neq j} M_{ij}(t) p_j(t) \mathrm{d}t \\ &= -\frac{1}{\tau} \int_0^\tau \sum_i M_{ii}(t) p_i(t) \mathrm{d}t.\end{aligned}$$

Since for all i, $p_i(t) \leqslant 1$ and $M_{ii}(t) \leqslant 0$, we can lower bound the dynamical activity as

$$\langle A \rangle_\tau \leqslant -\frac{1}{\tau} \int_0^\tau \sum_i M_{ii}(t) \mathrm{d}t = \frac{1}{\tau} \ln(\det P), \tag{B6}$$

where we have used $P = T_M(0,\, \tau)$ and equation (5). It can be seen that proposition 2 is implied by combining equations (B5), (B6) and taking $p = p(0)$, $Pp = p(\tau)$.

Appendix C. Proof of proposition 4

Proposition 4. *If a matrix P is LE, it is QE with at most one hidden state.*

Proof. Let P be a LE matrix. By definition, there is an embeddable matrix P' within any desired distance ϵ_1 of P.

Now consider a stochastic matrix which keeps every state fixed except state i, which it sends to j with probability α, and leaves alone with probability $1 - \alpha$. In control theory, such matrices are called *Poisson matrices*. It is known that any embeddable matrix P' can be approximated arbitrarily closely by a finite product of such matrices [44]. Let P'' indicate a product of Poisson matrices which approximates P' within a distance ϵ_2.

A Poisson matrix is QE with one hidden state by lemma 9. Since P'' is a product of Poisson matrices, P'' is QE with one hidden states. Thus, P is arbitrarily close (less than $\epsilon_1 + \epsilon_2$) to P'', a matrix which is QE with one hidden state.

Now choose such a P'' within $\epsilon/2$ of P. If we wish to implement P within ϵ while producing total entropy less than δ, with one hidden state, we can do so with a rate matrix $M(t)$ that implements P'' to within $\epsilon/2$ while producing less than δ entropy, with one hidden state. There always is one since P'' is QE with one hidden state. This establishes the result. □

Appendix D. Proof of proposition 5, that any LR matrix is QE

Proposition 5. *Any LR is QE.*

Proof. We first show this for LRs that have a single block, that is, those of the form

$$P = \pi \mathbf{1}^T,$$

where π is an n-dimensional column-vector with positive entries that sum to 1, and $\mathbf{1}$ is an n-dimensional column-vector of 1s. We construct a sequence of CTMCs that approximates P arbitrarily well while achieving arbitrarily small total EP for some initial distribution q.

Consider the rate matrix whose off-diagonal entries are given by $M_{ij}(t) = \alpha w_i(t)$, where $w_i(t) = (1 - t)q_i + t\pi_i$. The associated master equation decouples

$$\dot{p}_i(t) = \alpha(w_i(t) - p_i(t)), \tag{D1}$$

and can be solved explicitly (using the initial condition $p(0) = q$),

$$p_i(t) = w_i(t) + \alpha^{-1}(q_i - \pi_i)(1 - \mathrm{e}^{-\alpha t}). \tag{D2}$$

It is clear that by making α sufficiently large, this CTMC will transform any initial distribution at $t = 0$ arbitrarily close to final distribution π at $t = 1$. That is

$$\lim_{\alpha \to \infty} p_i(1) = \lim_{\alpha \to \infty} w_i(1) = \pi_i(t).$$

Thus, it will approximate P arbitrarily well. We must now show that by making α large we can also make the total EP as small we like.

Since the rates $M_{ij}(t)$ satisfy detailed balance at all times, $M_{ij}(t)w_j(t) = M_{ji}(t)w_i(t)$, the total EP can be written as

$$\Sigma = -\sum_i \int_0^1 \dot{p}_i(t) \ln \frac{p_i(t)}{w_i(t)}\, \mathrm{d}t.$$

We split each integral into two parts

$$\int_0^1 \dot{p}_i(t) \ln p_i(t)\, \mathrm{d}t - \int_0^1 \dot{p}_i(t) \ln w_i(t)\, \mathrm{d}t.$$

The first integral can be evaluated as

$$\begin{aligned}\int_0^1 \dot{p}_i(t) \ln p_i(t) \mathrm{d}t &= \int_0^1 \frac{d}{\mathrm{d}t}[p_i(t) \ln p_i(t) - p_i(t)]\, \mathrm{d}t \\ &= p_i(1) \ln p_i(1) - q_i \ln q_i + q_i - p_i(1),\end{aligned}$$

which, as $\alpha \to \infty$, converges to

$$\pi_i \ln \pi_i - q_i \ln q_i + q_i - \pi_i. \tag{D3}$$

The second integral can be written as

$$\int_0^1 \dot{p}_i(t) \ln w_i(t)\, \mathrm{d}t = (\pi_i - q_i) \int_0^1 (1 - \mathrm{e}^{-\alpha t}) \ln w_i(t)\, \mathrm{d}t,$$

where we've plugged equation (D2) into equation (D1). As $\alpha \to \infty$, the integrand converges pointwise and monotonically to $\ln w_i(t)$, which is bounded for all $t \in (0,\ 1)$. By the dominated convergence theorem, as $\alpha \to \infty$, the second integral converges to

$$(\pi_i - q_i) \int_0^1 \ln w_i(t)\, \mathrm{d}t \tag{D4}$$

$$= (\pi_i - q_i) \int_0^1 \ln((1 - t) q_i + t\pi_i)\, \mathrm{d}t$$

$$= \pi_i \ln \pi_i - q_i \ln q_i + q_i - \pi_i. \tag{D5}$$

Comparing equations (D3) and (D5), we see that in the limit of $\alpha \to 0$, the two integrals cancel, and so $\Sigma \to 0$.

The general case of a LR with *multiple* blocks follows immediately, because a block diagonal matrix is QE if the blocks are—a block diagonal matrix A with blocks B_i can be factored as a product $A = (B_1 \oplus I \oplus \cdots)(I \oplus B_2 \oplus \cdots)\cdots$. □

Appendix E. Proof of main result in section 5

To establish our main result we first establish some preliminary results.

Lemma 9. *Any* 2×2 *stochastic matrix is QE with one hidden state.*

Proof. For any real number $p \in [0,\ 1]$, define $\bar{p} \coloneqq 1 - p$. Without loss of generality[14], write P in matrix notation as

$$P = \begin{bmatrix} \bar{p} & q \\ p & \bar{q} \end{bmatrix}.$$

We consider two cases. If $p \leqslant \bar{q}$, then if $x = p/\bar{q}$ and $y = q$, we have

$$\begin{bmatrix} \bar{p} & q & \bar{p} \\ p & \bar{q} & p \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} y & y & 0 \\ \bar{y} & \bar{y} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x & 0 & x \\ 0 & 1 & 0 \\ \bar{x} & 0 & \bar{x} \end{bmatrix}.$$

If instead $p > \bar{q}$, then if $x = \bar{p}$, $y = \bar{q}/p$, we have

$$\begin{bmatrix} \bar{p} & q & q \\ p & \bar{q} & \bar{q} \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x & x & 0 \\ \bar{x} & \bar{x} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & y & y \\ 0 & \bar{y} & \bar{y} \end{bmatrix}.$$

In both cases, the factors on the right-hand side are LRs, so the result is established. □

[14] The other kind of transfer can be rewritten in this form by permuting its rows and columns.

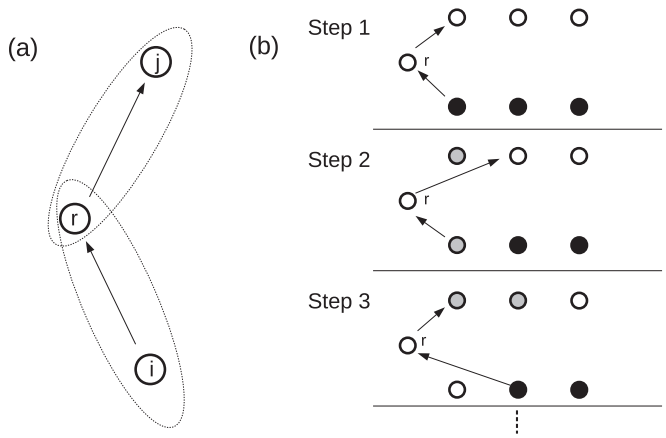

Figure 2. (a) A composition of LRs can be used to effect a transfer of probability α from i to j, using a relay state r. (b) A sequence of these operations can be composed to effect any transfer T.

E.1. Transfers of probability using a relay state

In what follows, it will be useful to consider stochastic matrices of the form:

$$\begin{bmatrix} I & P \\ 0 & D \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} D & 0 \\ P & I \end{bmatrix},$$

where P is an $m \times p$ matrix with positive entries with column sums less than or equal to one, D is a $p \times p$ diagonal matrix that makes the whole block matrix stochastic, and I is the $m \times m$ identity matrix.

We call such matrices *transfers*, because they can be viewed as representing transfers of probability between two disjoint sets of states of sizes p and m. One can show that:

Proposition 10. *Any transfer T is QE with one hidden state.*

We will make repeated use of the map, $\mathcal{R}^{i\to j}(\alpha)$, acting on a set with two elements, which keeps state j fixed, but sends state i to j with probability α and leaves it as i with probability $1 - \alpha$. Such a matrix, which is called a *Poisson matrix* in control theory, is QE with one hidden state by lemma 9. See figure 2 for an illustration.

Proof of proposition 10. Without loss of generality[14], let

$$T = \begin{bmatrix} D & 0 \\ P & I \end{bmatrix}.$$

If P has only one nonzero row, whose entries are p_i, $i = 1,\ldots, n$, then

$$T = \prod_{i=1}^{n} \mathcal{R}^{i\to j}(p_i),$$

where the indices i and j are now labeling rows (e.g. states) in the larger matrix T. The first n rows that i ranges over become the rows of D, and j is the index of the single nonzero row of P, now as a submatrix of T. So by construction $j > n \geqslant i$, and the maps $\mathcal{R}^{i\to j}(p_i)$ for different i do not interfere with one another (they commute and fix each others images).

Now suppose P has two nonzero rows. Write P as the sum of two matrices $P = P_1 + P_2$, each zero except for *one* of the rows of P. Let D_1 be the diagonal matrix which makes

$$\begin{bmatrix} D_1 & 0 \\ P_1 & I \end{bmatrix}$$

stochastic, and write

$$\begin{bmatrix} D & 0 \\ P & I \end{bmatrix} = \begin{bmatrix} D_2 D_1 & 0 \\ P_2 + P_1 & I \end{bmatrix} = \begin{bmatrix} D_2 & 0 \\ P_2 D_1^{-1} & I \end{bmatrix} \begin{bmatrix} D_1 & 0 \\ P_1 & I \end{bmatrix},$$

where D_2 is chosen to make the matrix is appears in stochastic (this is always possible since the column sums of P are all less than 1). Since the product of stochastic matrices is stochastic, the matrix appearing on the left-hand side is T, establishing the proposition for P with two nonzero rows. The result for general P follows by induction. □

E.2. Minimum number of hidden states is less than non-negative rank

Lemma 11. *An $n \times n$ stochastic matrix P with non-negative rank 2 is QE with one hidden state.*

Proof. Write the non-negative rank decomposition $P = RS$ where R and S are stochastic matrices of dimensions $n \times 2$ and $2 \times n$, respectively. We further decompose R and S into sub-matrices: $R = \begin{bmatrix} R_m \\ R_2 \end{bmatrix}$ and $S = [S_m \; S_2]$, where R_2 and S_2 are 2×2. Now note that

$$P = \begin{bmatrix} R_m S_m & R_m S_2 \\ R_2 S_m & R_2 S_2 \end{bmatrix}$$

and

$$\begin{bmatrix} R_m S_m & R_m S_2 \\ R_2 S_m & R_2 S_2 \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & R_2 D^{-1} \end{bmatrix} \begin{bmatrix} I & R_m \\ 0 & D \end{bmatrix} \begin{bmatrix} 0 & 0 \\ S_m & I \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & S_2 \end{bmatrix},$$

where D is the diagonal matrix that makes the second factor stochastic, and I is an identity matrix of dimensions suitable to where it appears. To show that P can be implemented with one hidden state, it suffices to show that each of the factors on the right hand side can be.

The middle two factors represent transfers, in the sense described above, so they can implemented with one hidden state. The first and last factors represent operations performed on just two states, so by lemma 9 can also be implemented with one hidden state. Note that $R_2 D^{-1}$ is stochastic, because the column sums of R_2 are exactly the diagonal entries of D (recall D was chosen to make the second factor stochastic). □

We are now ready to prove theorem 8.

Theorem 8. *An $n \times n$ stochastic matrix P with non-negative rank r is QE with $r - 1$ hidden states.*

Proof. Suppose $r > 2$ (if not, the result follows from lemma 11). As before, write $\pi = RS$ where R and S are stochastic of dimensions $n \times r$ and $r \times n$, respectively. But this time, decompose R and S differently: $R = [R_m \; P]$ and $S = \begin{bmatrix} S_m \\ Q \end{bmatrix}$, where P is $n \times 2$ and Q is $2 \times n$. Note that $\pi = R_m S_m + PQ$.

$$\begin{bmatrix} PQ + R_m S_m & R_m \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} I & R_m \\ 0 & 0 \end{bmatrix} \begin{bmatrix} PQD^{-1} & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} D & 0 \\ S_m & I \end{bmatrix}$$

where D is the diagonal matrix that makes the second factor stochastic, and I is an identity matrix of dimensions suitable to where it appears.

To prove the theorem, it suffices show that each of the $(n + k - 2) \times (n + k - 2)$ matrices on the right hand side care QE with one hidden state.

The first and last factors represent transfers from $r - 2$ hidden states to the original n states (and vice versa), and can be implemented using one hidden state, as described earlier. The middle factor, which represents an operation performed only on the original states, has non-negative rank 2 (note that P and QD^{-1} are stochastic), so by lemma 11 can also be implemented with one hidden state. This establishes the result. □

Appendix F. Countably infinite state spaces

We now consider the case where the state space of our system, X, is countably infinite, and restrict attention to the implementation of single-valued maps $f\colon X \to X$ over such spaces.

In our results above, we establish quasistatic embeddability of various matrices by writing them as products of LRs. In the infinite case, we can similarly ask what is possible by composing LRs, which in the context of single-valued maps are exactly the idempotent functions.

One natural way to extend our analysis to the case of countably infinite X is to consider a sequence of idempotent functions over some $Y \supset X$ that, when restricted to X, gives the desired f. To allow the sequence of functions to implement $f(x)$ for *all* $x \in X$, in general we must consider sequences that are infinite. However, the infinite product of LRs need not be QE (or even be well-defined). To circumvent this issue we impose a 'practical' interpretation of what it means to implement f, by requiring that any particular input $x \in X$ is mapped to $f(x)$ after finitely many idempotents.

Adopting this interpretation, in this appendix we establish the following result:

Proposition 12. *Let X be a countable set, and take $Y \coloneqq X \cup \{z\}$. For any function $f\colon X \to X$ there is a sequence $\{g_i\}$ of idempotent functions $g_i\colon Y \to Y$ such that for all $x \in X \subset Y$, there is an m (which can depend on x) such that for all $r \geqslant m$, $(g_r \circ g_{r-1} \circ \cdots \circ g_1)(x) = f(x)$.*

Proposition 12 is a kind of infinite analog of propositions 6 and 7, which showed that any function over a finite X can be implemented with at most one hidden state.

To prove proposition 12 we begin with the following lemma:

Lemma 13. *Let Y be $\mathbb{Z}$ with one point added. There is a sequence $\{g_i\}$ of idempotent functions $g_i\colon Y \to Y$ such that for all $n \in \mathbb{Z} \subset Y$, there is an m (which can depend on n) such that for all $r \geqslant m$, $(g_r \circ g_{r-1} \circ \cdots \circ g_1)(n) = n + 1$.*

Proof. First, let $[\alpha]$ indicate the cyclic permutation $(-1, 0, 1)$ and, for any integer $i \in \mathbb{Z}$, let $[i]$ indicate the cyclic permutation $(-i, i+1, -i-1)$.

For any particular $j \in \mathbb{Z}$, consider the following composition of permutations,

$$M^{(j)} \coloneqq [j] \ldots [2][1][\alpha].$$

We show by induction that $M^{(j)}$ sends each state $i \in \{-j-1, \ldots, j\}$ to $i + 1$ (while also sending $j+1$ to $-j-1$ as a 'side effect'). First, it can be verified by inspection that it holds true for $j = 0$ (when $M^{(j)} = [\alpha]$). Second, assume it holds true for $M^{(j)}$ and then observe that

$$M^{(j+1)} = [j+1]M^{(j)} = (-j-1, j+2, -j-2)M^{(j)}.$$

So, applying $[j+1]$ after $M^{(j)}$ results in:

(a) $-j-2$ being sent to $-j-1$;

(b) the 'side effect' of $M^{(j)}$ being fixed, in that, $M^{(j)}$ mapped $j + 1 \mapsto -j-1$, but $[j+1]$ maps $-j-1 \mapsto j + 2$, so the combined result is $j + 1 \mapsto j + 2$.

Thus, $M^{(j+1)}$ sends each state $i \in \{-(j+1)-1, \ldots, (j+1)\}$ to $i + 1$ (while now sending $(j+1)+1$ to $-(j+1)-1$ as a 'side effect').

Note that a cyclic permutation like $[-j]$ and $[\alpha]$ affect only a finite number of points in $\mathbb{Z}$, and thus can be written as a product of a finite number of transpositions, which can in turn be written as a product of three idempotents that use one hidden state (e.g. as in example 3). This hidden state is provided by Y, which lets us construct a sequence of idempotents $\{g_i\}$ over Y with the property we desire. □

We can now use a 'dovetailing' algorithm to construct a sequence of local idempotents that implements any specified function, thereby establishing proposition 12:

Proof of proposition 12. Suppose first that f is a bijection. Partition X according to the orbits of f—that is, two points x and y are in the same part if there is some n such that $y = f^n(x)$ or $x = f^n(y)$.

There will be countably many orbits A_j. Each of the A_j is either finite and f restricted to it is a permutation, or else A_j is countably infinite and there is a numbering of its elements (using positive and negative integers) such that f restricted to that orbit is $n \to n + 1$. In either case, using proposition 6 or lemma 13, we can form for each orbit a sequence $\{(g_j)_i\}$ that implements the restriction of f to that orbit. Extend the maps in each sequence to all of Y by having them fix the elements in all other orbits.

Now form a new sequence by interleaving these (countably many) sequences in a way that preserves the order of elements in each individual sequence. For example

$$\{(g_1)_1, (g_1)_2, (g_2)_1, (g_1)_3, (g_2)_2, (g_3)_1, \ldots\}.$$

This sequence satisfies the requirements of the theorem. Any $x \in X$ is in some orbit A_j, and so there is some element of the associated sequence $(g_j)_m$ after the application of which x is mapped to $f(x)$ and remains so. We use here the fact that the interleaving to form the larger sequence preserves order, and that idempotents that are members of sequences corresponding to the other orbits fix A_j.

If f is not a bijection, consider the idempotent map a that, for all $x \in X$, sends all elements in the inverse image $f^{-1}(x)$ to a distinguished element $w(x) \in f^{-1}(x)$. Write $Z = \operatorname{img} f \cup \operatorname{img} a$. We can write $f = h \circ a$, where $h\colon Z \to Z$ is a bijection. So using the construction above, we can make a sequence of idempotents for h with the desired property. Adding a to the beginning of the sequence forms a sequence that implements f. For each x there exists an $w(x)$, and under the implementation of h, $w(x) \mapsto f(w(x))$ once some number m idempotents are applied. But note that x was first mapped to $w(x)$, under the initial application of a. Thus, once

m idempotents of h have been applied, the combined effect on x is $x \mapsto w(x) \mapsto f(w(x))$, and $f(w(x)) = f(x)$, by the definition of $w(x)$, so this completes the argument. □

ORCID iDs

Jeremy A Owen https://orcid.org/0000-0002-9180-3794
Artemy Kolchinsky https://orcid.org/0000-0002-3518-9208

References

[1] Van den Broeck C *et al* 2013 Stochastic thermodynamics: a brief introduction *Phys. Complex Colloids* **184** 155–93
[2] Van den Broeck C and Esposito M 2015 Ensemble and trajectory thermodynamics: a brief introduction *Physica* A **418** 6–16
[3] Seifert U 2012 Stochastic thermodynamics, fluctuation theorems and molecular machines *Rep. Prog. Phys.* **75** 126001
[4] Van Kampen N G 1981 *Stochastic Processes in Chemistry and Physics* vol 1 (Amsterdam: North-Holland) pp 120–7
[5] Givon D, Kupferman R and Stuart A 2004 Extracting macroscopic dynamics: model problems and algorithms *Nonlinearity* **17** R55
[6] Gaspard P 2006 Hamiltonian dynamics, nanosystems, and nonequilibrium statistical mechanics *Physica* A **369** 201–46
[7] Esposito M, Kawai R, Lindenberg K and Van den Broeck C 2010 Finite-time thermodynamics for a single-level quantum dot *Europhys. Lett.* **89** 20003
[8] Elfving G 1937 Zur theorie der Markoffschen ketten *Acta Societas Scientiarium Fennicae Nova Series* A **2** 1–7
[9] Kingman J F C 1962 The imbedding problem for finite Markov chains *Z. Wahrscheinlichkeitstheor. Verwandte Geb.* **1** 14–24
[10] Frydman H and Singer B 1979 Total positivity and the embedding problem for Markov chains *Math. Proc. Camb. Phil. Soc.* **86** 339–44
[11] Fuglede B 1988 On the imbedding problem for stochastic and doubly stochastic matrices *Probab. Theory Relat. Fields* **80** 241–60
[12] Lencastre P, Raischel F, Rogers T and Lind P G 2016 From empirical data to time-inhomogeneous continuous Markov processes *Phys. Rev.* E **93** 032135
[13] Chen J 2016 A solution to the reversible embedding problem for finite Markov chains *Stat. Probab. Lett.* **116** 122–30
[14] Cubitt T S, Eisert J and Wolf M M 2012 The complexity of relating quantum channels to master equations *Commun. Math. Phys.* **310** 383–418
[15] Bausch J and Cubitt T 2016 The complexity of divisibility *Linear Algebr. Appl.* **504** 64–107
[16] Wolpert D H, Kolchinsky A and Owen J A 2017 The minimal hidden computer needed to implement a visible computation arXiv:1708.08494
[17] Kolchinsky A and Wolpert D H 2017 Dependence of dissipation on the initial distribution over states *J. Stat. Mech.* 083202
[18] Turgut S 2009 Relations between entropies produced in nondeterministic thermodynamic processes *Phys. Rev.* E **79** 041102
[19] Maroney O J E 2009 Generalizing Landauer's principle *Phys. Rev.* E **79** 031105
[20] Wolpert D H 2015 Extending Landauer's bound from bit erasure to arbitrary computation arXiv:1508.05319
[21] Wolpert D H 2016 The free energy requirements of biological organisms; implications for evolution *Entropy* **18** 138
[22] Goodman G S 1970 An intrinsic time for non-stationary finite Markov chains *Probab. Theory Relat. Fields* **16** 165–80
[23] Esposito M and Van den Broeck C 2010 Three faces of the second law: I. Master equation formulation *Phys. Rev.* E **82** 011143
[24] Landauer R 1961 Irreversibility and heat generation in the computing process *IBM J. Res. Dev.* **5** 183–91
[25] Bennett C H 1982 The thermodynamics of computation—a review *Int. J. Theor. Phys.* **21** 905–40
[26] Sagawa T 2014 Thermodynamic and logical reversibilities revisited *J. Stat. Mech.* P03025
[27] Salamon P and Berry R S 1983 Thermodynamic length and dissipated availability *Phys. Rev. Lett.* **51** 1127
[28] Andresen B, Berry R S, Ondrechen M J and Salamon P 1984 Thermodynamics for processes in finite time *Acc. Chem. Res.* **17** 266–71
[29] Zulkowski P R and DeWeese M R 2014 Optimal finite-time erasure of a classical bit *Phys. Rev.* E **89** 052140
[30] Shiraishi N, Funo K and Saito K 2018 Speed limit for classical stochastic processes *Phys. Rev. Lett.* **121** 070601
[31] Diana G, Bagci G B and Esposito M 2013 Finite-time erasing of information stored in fermionic bits *Phys. Rev.* E **87** 012111
[32] Howie J M 1966 The subsemigroup generated by the idempotents of a full transformation semigroup *J. London Math. Soc.y* **1** 707–16
[33] Cohen J E and Rothblum U G 1993 Nonnegative ranks, decompositions, and factorizations of nonnegative matrices *Linear Algebr. Appl.* **190** 149–68
[34] Bennett C H 1973 Logical reversibility of computation *IBM J. Res. Dev.* **17** 525–32
[35] Levine R Y and Sherman A T 1990 A note on Bennett's time-space tradeoff for reversible computation *SIAM J. Comput.* **19** 673–7
[36] Fredkin E and Toffoli T 1982 Conservative logic *Int. J. Theor. Phys.* **21** 219–53
[37] Barato A C and Seifert U 2016 Cost and precision of Brownian clocks *Phys. Rev.* X **6** 041053
[38] Barato A C and Seifert U 2017 Coherence of biochemical oscillations is bounded by driving force and network topology *Phys. Rev.* E **95** 062409
[39] Lindblad G 1976 On the generators of quantum dynamical semigroups *Comm. Math. Phys.* **48** 119–30
[40] Schnell A, Eckardt A and Denisov S 2018 Is there a floquet lindbladian? arXiv:1809.11121
[41] Hush M R, Lesanovsky I and Garrahan J P 2015 Generic map from non-lindblad to lindblad master equations *Phys. Rev.* A **91** 032113
[42] Gingrich T R, Rotskoff G M and Horowitz J M 2017 Inferring dissipation from current fluctuations *J. Phys. A: Math. Theor.* **50** 184004
[43] Johansen S and Ramsey F L 1979 A bang-bang representation for 3 × 3 embeddable stochastic matrices *Probab. Theory Relat. Fields* **47** 107–18
[44] Johansen S 1973 A central limit theorem for finite semigroups and its application to the imbedding problem for finite state Markov chains *Z. Wahrscheinlichkeitstheor. Verwandte Geb.* **26** 171–90