Imperial College London

Master of Applied Mathematics Thesis

Imperial College of Science, Technology and Medicine

Department of Mathematics

Path Probabilities and Entropy Production in non-Markovian Systems

Author:
Farid Kaveh

Supervisor:
Dr. Gunnar Pruessner

Second Marker:
Dr. John Doe

July 17, 2022

Abstract

Your abstract goes here

Acknowledgements

“ Thanks mum!

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Entropy and The Second Law: Classical Treatment

The original formulation of the second law of thermodynamics is due to the British physicist and mathematician William Thomson (later Lord Kelvin) who , in 1851, stated the principle as follows.

"It is impossible, by means of inanimate material agency, to derive mechanical effect from any portion of matter by cooling it below the temperature of the coldest of the surrounding objects."[15]

Put more plainly, Thomson's principle asserts that *"No process is possible whose sole result is the complete conversion of heat into work."*[7, see §13.1] In 1854, the German Physicist Rudolf Clausius independently produced the following statement of the second law.[1]

> *"Heat can never pass from a colder to a warmer body without some other change, connected therewith, occuring at the same time."*[11,14]

These statements of the second law are in fact equivalent, and upon their application to a general non-reversible cycle process one can derive the Clausius inequality

$$\oint \frac{\delta Q}{T} \leq 0, \tag{1.1}$$

for any thermodynamic cycle, with equality holding if and only if the cycle is reversible [7, see §13.4 & §13.7].[2] Here δQ is the heat entering into the system at temperature T. The notation δQ reflects the fact that this is an inexact (read path dependent) differential. If the process in question is reversible, then the integral on the LHS of the Clausius inequality is path independent, hence we can identify the state (i.e. path independent) variable

$$S(A) - S(B) = \int_B^A \frac{\delta Q_r}{T} \tag{1.2}$$

where δQ_r is the heat transfer during a reversible process from state B to state A. Eqn (1.2) is our

[1] In their statements Kelvin and Clausius were both motivated by the previous work of Sadi Carnot on heat engines.

[2] In fact, Claudius proved the case of equality in the case of reversible processes in his 1854 paper.[14]

first definition of entropy. [3] Although this is a correct definition of entropy for thermodynamic systems, it does not identify the statistical nature of the property. Moreover, being stated strictly in terms of thermodynamic quantities, it obscures the generality of the concept of entropy which has found application in many fields, including information theory,[13] biology,[6] and sociology.[8]

Equation 1.2 implies that

$$dS \coloneqq \frac{\delta Q}{T} \tag{1.3}$$

is in fact an exact differential, and upon comparison with the Clausius inequality one obtains

$$dS \geq 0, \tag{1.4}$$

which is perhaps the most familiar statement of the second law to the contemporary reader.

It was Boltzmann who first identified the quantity $\frac{\delta Q_r}{T}$ as an exact differential and formulated the foregoing definition of entropy. Boltzmann produced this first definition in the year 1866.[4,9] In the following years, from 1868 to 1872, Boltzmann produced several important papers. Among the most notable results obtained by him in this period is that the Maxwell distribution is an attractive stationary point for the velocity distribution of a body of spatially homogenous gas.[4,9].

At this point, the statistical nature of entropy is not yet clear. Later, through discussions with Josed Loschmidt, and through consideration of Maxwell's demon, Boltzmann went on to explore the entropy from an explicitly statistical point of view. However, we shall leave Boltzmann behind and explore this probabilistic point of view with a modern treatment in the next section.

1.2 Statistical Framework

In *"Statistical Field Theory"* Giorgio Parisi provides an excellent discussion of the statistical formulation of entropy and its equivalence to the thermodynamic definition. [10, see §1] Here we shall give a brief summary of his approach.

Consider a system with configuration space $\mathcal{X}$ and let μ be a measure on $\mathcal{X}$. Then the system has a "Hamiltonian" $H(x)$, $x \in \mathcal{X}$. The fundamental hypothesis of equilibrium statistical mechanics is that the equilibrium probability distribution for this system is given by the so called canonical distribution

find source to justify this

$$P_\beta(x) = \frac{e^{-\beta H(x)}}{Z} d\mu, \tag{1.5}$$

where the partition function Z is fixed by normalisation and is given by

$$Z = \int_{\mathcal{X}} e^{-\beta H} d\mu. \tag{1.6}$$

The ensemble average of an observable $A(x)$ at equilibrium is then given by

$$\langle A \rangle_\beta = \int_{\mathcal{X}} A(x) P_\beta(x) d\mu. \tag{1.7}$$

The canonical distribution (1.5) is characterised by the fact that it maximises the entropy functional

[3] Here the reader may note that the definition (1.2) only identifies changes in entropy and furthermore that the quantity described is not dimensionless and in fact has units of $\mathrm{J} \cdot \mathrm{K}^{-1}$. The latter issue is characteristic of entropy in classical systems. Since the multiplicity of a classical system is not well-defined, one must consider a certain region of phase space as an analogue of multiplicity. This region carries with it the units of action. See the book by Landau and Lifshitz for a detailed discussion.[5] This issue of a fixed zero for entropy is dealt with by the third law of thermodynamics which states that in the limit as the temperature of a crystal lattice approaches zero, its entropy goes to zero. But that is by the way.

$$S[P] := -\langle \log P\rangle_P = -\int_X P(x)\log P(x)d\mu, \tag{1.8}$$

which, if $\mathcal{X}$ is finite with M elements (and μ is the counting measure), is equivalent to

$$S[P] = -\sum_{i=1}^{M} P_i \log P_i \tag{1.9}$$

which is the familiar *Shannon entropy*,[4] after Claude E. Shannon who introduce the notion in his landmark 1948 paper *"A Mathematical Theory of Communication"*.[13] Let us further define the energy functional $E[P]$ and the free energy functional $\Phi[P]$,

$$U = E[P] := \langle H\rangle_P \tag{1.10}$$

$$\Phi[P] := E[P] - \frac{S[P]}{\beta} \tag{1.11}$$

and write

$$S_\beta := S[P_\beta] \tag{1.12}$$

for the equilibrium entropy of the system. We can show that $dS_\beta = dS$, where S is the thermodynamic entropy as defined in Eqn. 1.2. First of all note that for a thermodynamic system the configuration space $\mathcal{X}$ is precisely the phase space of N positions and N momenta. We will furthermore allow H to depend on a parameter λ so that we can do work on the system by varying λ. So $H = H(q, p, \lambda)$, where $q = (q_1, \dots, q_N)$ are the positions and $p = (p_1, \dots, p_N)$ are the momenta of the system. Hence the partition function is

$$Z = \int H(p, q, \lambda)P_\beta dpdq. \tag{1.13}$$

Moreover the first law of thermodynamics states

$$\delta Q = dU - \delta W, \tag{1.14}$$

hence, by Eqn. 1.3, $dS = (dU - \delta W)/T$. Hence if

$$\beta\delta Q = \beta(dU - \delta W) = dS_\beta, \tag{1.15}$$

then $\beta\delta Q$ is an exact differential, i.e. β is proportional to $1/T$. From this we could then conclude that $dS_\beta = dS$, and $S_\beta = S$ up to an additive constant. It remains to prove Eqn. 1.15. From Eqn. 1.13 we derive the differential

$$-d(\log Z) = Ud\beta + \beta\left\langle \frac{dH}{d\lambda}\right\rangle_\beta d\lambda = Ud\beta + \beta\delta W. \tag{1.16}$$

Moreover we observe that

$$\Phi[P_\beta] = -\frac{1}{\beta}\log Z = U - \frac{S_\beta}{\beta}. \tag{1.17}$$

Eqn. 1.15 now follows from Eqns. 1.16 and 1.17.

[4] Also sometimes called the Gibbs entropy.

1.3 The Present Work

1.3.1 Entropy Production

discuss each of these before

In equilibrium statistical mechanics, entropy can be interpreted as the disorder exhibited by a system in a certain statistical ensemble. The stationary state is then the distribution that maximises this disorder. For example, in the canonical ensemble, where the system is assumed to be at temperature T, the Boltzmann distribution maximises the entropy of the system. Hence the Boltzmann distribution is the statistical steady state for the canonical ensemble. The Shannon, Gibbs, and Boltzmann entropies are notions of steady-state entropy. They are concerned with the phase space of the system, Λ, and not with trajectories along the phase space. This space of phase space trajectories we shall henceforth denote by Ω.

Chapter 2

The Telegraph Process

2.1 Known Results

The telegraph process is the continuous-time, two-state Markov chain described by a transition matrix

$$W = \begin{pmatrix} -\alpha & \beta \\ \alpha & -\beta \end{pmatrix}. \tag{2.1}$$

add figure of the system

The telegraph process is a building block of more complicated processes which we will consider later in this report. For this reason it is helpful to review some known results about its entropy production rate. Although the entropy production of discrete state-space Markov chains is usually considered from the stand point of graph theory, here we will attempt to replicate these results using Gaspard's framework for entropy production. In this way we will develop some useful results by considering this simple system.

We will label the two states $|a\rangle$ and $|b\rangle$. We shall use the terms 'particle' and 'system' interchangeably, i.e. we may speak of the system or the particle being in state $|a\rangle$ etc. Given an initial condition $P(0) = P_0 = p\,|a\rangle + (1-p)\,|b\rangle$, the system will evolve according to $P(t) = e^{tW}P_0$, and it will eventually reach the equilibrium distribution P_∞, which is the normalised eigenvector of e^{tW} with eigenvalue equal to one. Indeed

$$P(t) = A(t)\,|a\rangle + B(t)\,|b\rangle = \frac{1}{\alpha+\beta}\left((\beta + re^{-(\alpha+\beta)t})\,|a\rangle + (\alpha - re^{-(\alpha+\beta)t})\,|b\rangle\right), \tag{2.2}$$

with $r = \alpha p - \beta(1-p)$. The internal and external entropy productions are then given by [2]

$$\dot{S}_i(t) = (\alpha A(t) - \beta B(t))\log\left(\frac{\alpha A(t)}{\beta B(t)}\right) = re^{-(\alpha+\beta)t}\log\left[\frac{1 + \frac{r}{\beta}e^{-(\alpha+\beta)t}}{1 - \frac{r}{\alpha}e^{-(\alpha+\beta)t}}\right] \tag{2.3}$$

$$\dot{S}_e(t) = -re^{-(\alpha+\beta)t}\log\left(\frac{\alpha}{\beta}\right). \tag{2.4}$$

The most immediate observation from this result is that both the internal and the external entropy production rates of the telegraph process converge exponentially to zero. This is consistent with the fact that at its stationary state the telegraph process obeys the detailed balance condition, so it does not allow any current. Markov processes with null current cannot produce any entropy. The exponential decay of the entropy production rate also reflects the exponential decay of the initial state to statistical equilibrium.

In the symmetric case where $\alpha = \beta$, we have

$$\dot{S}_e \equiv 0. \tag{2.5}$$

Since $\dot{S}_e$ is the contribution to entropy production due to the heat reservoir(s) coupled to the system [12], we conclude that in the symmetric case the interaction between the telegraph process and any reservoirs does not result in any entropy production.

2.2 Path probabilities

Consider the discrete time, two-state Markov chain that evolves according to the matrix

$$M = \mathbb{1} + \frac{t}{N}W, \quad N \in \mathbb{N}, \tag{2.6}$$

with W as in 2.1. Then, given that the system starts in state $|a\rangle$, the probability of finding it there again after time t/N, having made no jumps, is given by $\langle a| M |a\rangle$. Moreover, the probability of the system making no jumps in time $t = Nt/N$ is

$$\underbrace{\langle a| M |a\rangle \langle a| M |a\rangle \dots \langle a| M |a\rangle}_{N \text{ inner products}} = \left(1 + \frac{tW_{11}}{N}\right)^N \tag{2.7}$$

To find the probability of the constant path ω_0 (the unique path which begins and ends in $|a\rangle$, with no jumps in time t), we take the limit of $N \to \infty$ to find

$$\mathbb{P}(\omega_0) = \lim_{N\to\infty} \left(1 + \frac{tW_{11}}{N}\right)^N = e^{-\alpha t} \tag{2.8}$$

On the other hand, the probability of a path ω_1 that makes only one jump to state $|b\rangle$ and at time $t_1 > 0$ is given by

$$\mathbb{P}(\omega_1) = \lim_{N\to\infty} \left(\underbrace{\langle a| M |a\rangle \dots \langle a| M |a\rangle}_{N_1 \text{ inner products}} \overbrace{\langle a| M |b\rangle}^{=\frac{\alpha t}{N}} \underbrace{\langle b| M |b\rangle \dots \langle b| M |b\rangle}_{N-N_1 \text{ inner products}} \right) \tag{2.9}$$

In the above, $t_1 = N_1 t/N$. Now if let N go to infinity while fixing $N_1/N = t_1/t$, this will ensures that N_1 is large. Note also that

$$\lim_{N\to\infty} (1 + \frac{x}{N})^{cN} = (e^x)^c = e^{cx}$$

Here we must reckon with the term $\langle a| M |b\rangle = \frac{\alpha t}{N}$. Clearly, this term approaches zero as $N \to +\infty$. In relation to the particle's path, this is intuitive. Because the path space Ω is uncountably infinite (in fact it is infinite dimensional), in general a particular path may not have non-zero measure. In fact, the constant path ω_0 is the only path with non-zero probabilty. Let $\Omega_1 \subset \Omega$ be the subset of all paths with only one jump. Then the limit of Equation 2.9 is the density of $\mathbb{P}|_{\Omega_1}$ (with respect to the lebesgue measure, under the "canonica" random variable from $\Omega_1 \to \mathbb{R}$).

justify the differential limit

This allows us to evaluate the Limit 2.9 to find

$$\mathbb{P}(\omega \in \Omega_1) = \lim_{N\to\infty} \left[\left(1 + \frac{tW_{11}}{N}\right)^{\frac{Nt_1}{t}} \frac{\alpha t}{N} \left(1 + \frac{tW_{22}}{N}\right)^{\frac{N(t-t_1)}{t}} \right] \tag{2.10}$$

$$= \alpha e^{-\alpha t \frac{t_1}{t}} e^{-\beta t \frac{(t-t_1)}{t}} dt_1 = \alpha e^{-\alpha t_1} e^{-\beta(t-t_1)} dt_1 \tag{2.11}$$

Let now $\Omega_L \subset \Omega$ be the subset of paths that make L jumps at times t_i . We can similarly calculate the density of $\mathbb{P}$ on Ω_2 to be

$$\mathbb{P}(\omega \in \Omega_2) = \alpha\beta e^{-\alpha t_1} e^{-\beta(t_2-t_1)} e^{-\alpha(t-t_2)} dt_1 dt_2. \tag{2.12}$$

More generally, the density of $\mathbb{P}|_{\Omega_L}$ is

$$\mathbb{P}(\omega \in \Omega_L) = \alpha^n \beta^m \prod_{i=0,\text{ even}}^{L} e^{-\alpha(t_{i+1}-t_i)} dt_i \prod_{i \text{ odd}}^{L} e^{-\beta(t_{i+1}-t_i)} dt_i; \;\; t_0 = 0,\; t_{L+1} = t. \tag{2.13}$$

slight error in this statement

In Equation 2.13, n is the number of transitions of the kind $|A\rangle \to |B\rangle$, while m is the number of transitions of the kind $|B\rangle \to |A\rangle$. Clearly $n + m = L$. Notice that if L is even, then the particle ends up back at $|A\rangle$ by time t, hence $n - m = 0$. Otherwise if L is odd, then the system is in state $|B\rangle$ at time t, so $n - m = 1$. t_i being the time of the i-th jump, we must have $t_i \in (t, t_{i-1})$. Furthermore, since the particle's jumps occur instantaneously, and the lower bound on the time between two jumps is zero, the number of jumps does not impose any additional constraints on the t_i's.

It would be reasonable at this point to check that our description of $\mathbb{P}$ indeed matches that of a probability measure. We can calculate the probability of Ω_L to be

$$\mathbb{P}(\Omega_L) = \alpha^n \beta^m \int_0^t dt_1 \int_{t_1}^t dt_2 \dots \int_{t_{L-1}}^t dt_L \prod_{i=0,\text{ even}}^{L} e^{-\alpha(t_{i+1}-t_i)} \prod_{i \text{ odd}}^{L} e^{-\beta(t_{i+1}-t_i)} \tag{2.14}$$

Let us first examine this quantity in the case where $\alpha = \beta$. In such an instance Equation 2.14 takes the much simpler form

$$\mathbb{P}(\Omega_L | \alpha = \beta) = \alpha^L e^{-\alpha t} \int_0^t dt_1 \int_{t_1}^t dt_2 \dots \int_{t_{L-1}}^t dt_L = \frac{(\alpha t)^L}{L!} e^{-\alpha t}. \tag{2.15}$$

We find that in the case where $\alpha = \beta$ the system reduces to a simple Poisson counting process with parameter $\lambda = \alpha t$. To treat the general case, first notice that 2.14 may be written

$$\mathbb{P}(\Omega_L) = \alpha^n \beta^m e^{-\beta t} \int_0^t e^{(\beta-\alpha)t_1} dt_1 \int_{t_1}^t e^{(\alpha-\beta)t_2} dt_2 \dots \int_{t_{L-1}}^t e^{(\beta-\alpha)t_L} dt_L, \tag{2.16}$$

where we have assumed that L is odd. If L were even, then we would only replace the integrand of the right most integral with $e^{(\alpha-\beta)t_L}$ and also replace the prefactor $e^{-\beta t}$ with $e^{-\alpha t}$. Now we will need the following lemma.

Lemma 2.2.1 *For $n = 0, 1, 2 \dots$, Let $I_n(t, r)$ be the integral*

$$I_n(t, r) = \int_0^t dt_1 \int_{t_1}^t dt_2 \dots \int_{t_{L-1}}^t dt_n \prod_{i \text{ odd}}^{n} e^{rt_i} \prod_{i \text{ even}}^{n} e^{-rt_i}, \tag{2.17}$$

with the convention $I_0 = 1$. Then we have the recurrence relation

$$\dot{I}_{n+1}(t, r) = e^{-rt} I_n(t, r); \;\; I_{n+1}(0, r) = 0 \tag{2.18}$$

for odd n and

$$\dot{I}_{n+1}(t, r) = e^{rt} I_n(t, r); \;\; I_{n+1}(0, r) = 0 \tag{2.19}$$

for even n.

proof: We will show the odd n case. The case for even n is analogous. Note that

$$I_n(t,r) = \int_0^t e^{rt_1} dt_1 \int_{t_1}^t e^{-rt_2} dt_2 \dots \int_{t_{L-1}}^t e^{rt_n} dt_n \tag{2.20}$$

$$= \int_{[0,t]^n} \mathbb{1}_{\{0<t_1<\dots<t_n<t\}} e^{rt_1} e^{-rt_2} \dots e^{rt_n} dt_n \dots dt_1 \tag{2.21}$$

$$= \int_0^t e^{rt_n} dt_n \dots \int_0^{t_3} e^{-rt_2} dt_2 \int_0^{t_2} e^{rt_1} dt_1. \tag{2.22}$$

Then

$$I_{n+1}(t,r) = \int_0^t e^{-rs} ds \int_0^s e^{rt_n} dt_n \dots \int_0^{t_3} e^{-rt_2} dt_2 \int_0^{t_2} e^{rt_1} dt_1 = \int_0^t e^{-rs} I(s,r) ds. \tag{2.23}$$

Differentiating 2.23 gives the result.

Taking the Laplace transform of (2.18) and (2.19), we obtain

$$\begin{aligned} \tilde{I}_{n+1}(s) &= \frac{\tilde{I}_n(s+r)}{s}, \qquad n \text{ odd}, \\ \tilde{I}_{n+1}(s) &= \frac{\tilde{I}_n(s-r)}{s}, \qquad n \text{ even}. \end{aligned} \tag{2.24}$$

Here $\tilde{f}(s)$ denotes the Laplace transform of $f(t)$. Applying this relation recursively we have

$$\begin{aligned} \tilde{I}_2(s) &= \frac{\tilde{I}_0(s)}{s(s+r)} & \tilde{I}_3(s) &= \frac{\tilde{I}_1(s)}{s(s-r)} \\ \tilde{I}_4(s) &= \frac{\tilde{I}_0(s)}{s^2(s+r)^2} & \tilde{I}_5(s) &= \frac{\tilde{I}_1(s)}{s^2(s-r)^2} \\ \tilde{I}_6(s) &= \frac{\tilde{I}_0(s)}{s^3(s+r)^3} & \tilde{I}_7(s) &= \frac{\tilde{I}_1(s)}{s^3(s-r)^3} \\ &\vdots & &\vdots \end{aligned} \tag{2.25}$$

whence we derive the expressions

$$\begin{aligned} I_{2n}(t,r) &= \mathcal{L}^{-1}\left[\frac{\tilde{I}_0(s)}{s^n(s+r)^n}\right](t), \\ I_{2n+1}(t,r) &= \mathcal{L}^{-1}\left[\frac{\tilde{I}_1(s)}{s^n(s-r)^n}\right](t). \end{aligned} \tag{2.26}$$

Computationally, evaluating (2.26) is a matter of partial fraction decomposition. The required inverse Laplace transform can subsequently be read off from a look-up table. The simple form of (2.26) may also allow us to sum the $\mathbb{P}(\Omega_L)$ in Laplace space to recover the normalisation

$$\sum_{L=0}^{\infty} \mathbb{P}(\Omega_L) = 1. \tag{2.27}$$

Chapter 3

The Waiting Room

3.1 The System

The system depicted in Figure 3.1a is a Markov process with transition matrix

$$W = W_{i\to j} = \begin{pmatrix} -\alpha-\gamma & \gamma & 0 & \alpha \\ \gamma & -\alpha-\gamma & \alpha & 0 \\ 0 & \gamma & -\beta-\gamma & \beta \\ \beta & 0 & \gamma & -\beta-\gamma \end{pmatrix}. \tag{3.1}$$

Henceforth we shall take $\alpha > \beta >> \gamma$ and call the resulting system the granular 'waiting room' system. As in Section 2.2, it is possible to write down the density of paths around any given path by conditioning on the number of jumps made by a particle in the waiting room system. The added complication in this case is that the density will not only depend on the number of jumps, but also on the type of each jump (e.g. the density of paths making one jump of the type $|1\rangle \to |2\rangle$ is different to those making one jump of the type $|1\rangle \to |4\rangle$).

From this granular waiting room system we obtain the Coarse Grained Waiting Room (CGWR) system by collapsing the states $|2\rangle$ and $|4\rangle$ together as in Figure 3.1b. The combination state obtained by this procedure we shall call $|w\rangle$. This coarse grained system is no longer Markov. It is also intuitively clear that the coarse grained system has non-zero entropy production along its paths. See for example Figure 3.2 which shows a diagram of a sample path. Clearly, the statistics of the forward and adjoint path are not equivalent. Our goal is to calculate the expectation value for the Kullback-Leibler divergence of the coarse grained paths when the system is in its steady

give example to show non-Markov maybe?

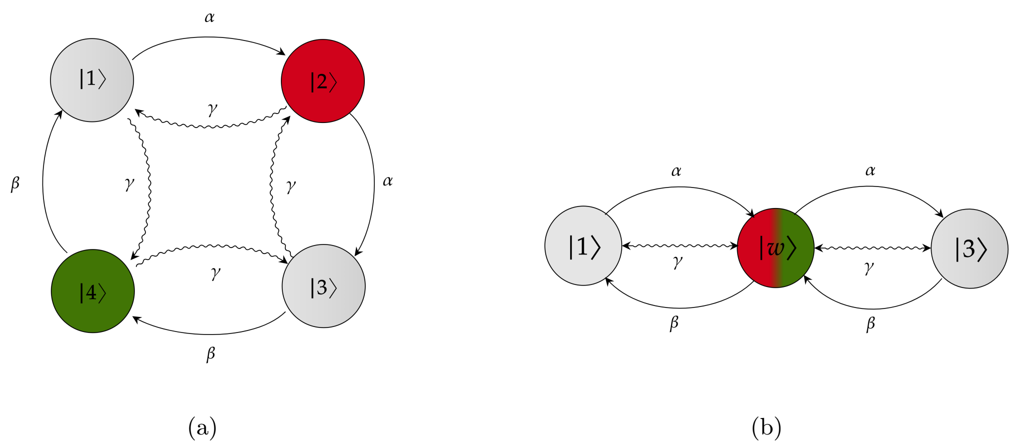

Figure 3.1: Subfigure (a) shows the granular waiting room system while (b) shows the coarse grained waiting room. We take $\alpha > \beta >> \gamma$.

state. Each state in the granular system has an exponential waiting-time distribution $\psi_{|i\rangle}(t)$. This is distinct from the distribution of a transition $|i\rangle \to |j\rangle$ which we denote by $\psi_{ij}(t)$. The probability that a transition away from $|i\rangle$ is of the type $|i\rangle \to |j\rangle$ is then

$$P_{ij} = \int_0^\infty \psi_{ij}(t)dt. \tag{3.2}$$

The waiting time distribution conditioned on a given transition is then the normalised distribution $\psi_{ij}(t)/P_{ij}$. For example, using the methodology from Section II, we derive for state $|1\rangle$

$$\begin{aligned} \psi_{|1\rangle}(t) &= (\alpha+\gamma)e^{-(\alpha+\gamma)t}, \\ \psi_{12} &= \alpha e^{-(\alpha+\gamma)t}, \\ \psi_{14} &= \gamma e^{-(\alpha+\gamma)t}, \end{aligned} \tag{3.3}$$

whereby $P_{12} = \frac{\alpha}{\alpha+\gamma}$, and $P_{14} = \frac{\gamma}{\alpha+\gamma}$ follow straightforwardly. Note that ψ_{12} and ψ_{14} are not normalised.

We stress that in the CGWR, the state $|w\rangle$ does not indicate a particular combination state of $|2\rangle$ and $|4\rangle$, but any convex linear combination state $a\,|2\rangle + b\,|4\rangle$. Moreover, if a state $a\,|2\rangle + b\,|4\rangle$ is measured at $t = 0$, and the system has made no jumps in time t, then it occupies the combination state *something*

3.2 Discrete Time Treatment

3.2.1 Path-Histories and Gluing Scheme

We will first treat this system in discrete time with time step τ. Write $M = \mathbb{1} + \tau W$ for the stochastic matrix of the discretised process. To avoid burdensome notation, we will write $(|i\rangle, m)$, for $i \in \{1, w, 3\}$ and $m \in \mathbb{N} \cup \{0\}$ to mean that the system was in state $|i\rangle$ for time $m\tau$. For any path ω beginning in $|1\rangle$ and ending in $|3\rangle$ the history of ω can be written

$$\{(|1\rangle, n_1), (|w\rangle, n_2), (|3\rangle, n_3), (|w\rangle, n_4), (|1\rangle, n_5), \ldots (|3\rangle, n_N)\}.$$

For example, if the path is given by $\omega = |1\rangle \xrightarrow{3} |w\rangle \xrightarrow{2} |1\rangle \xrightarrow{5} |w\rangle \xrightarrow{4} |3\rangle$ then we will have $n_1 = 3, n_2 = 2, n_3 = n_4 = 0, n_5 = 5, n_6 = 4$. Given a history in the form $\mathcal{P} = \{i_1, i_2, i_3, \ldots, i_N\}$ in the granular system, where the i_k are states, we know that the probability of the path described by $\mathcal{P}$ is given by

$$\mathbb{P}(\mathcal{P}) = \langle i_1| M |i_2\rangle \langle i_2| M |i_3\rangle \ldots |i_{N-1}\rangle \langle i_{N-1}| M |i_N\rangle\,.$$

Let now $\bar{\mathcal{P}} = \{(|i\rangle, 2), (|w\rangle, 2), (|j\rangle, 2)\}$, $i, j \in \{1, 3\}$ be a path in the coarse grained system. The particle jumping from $|i\rangle$ to $|w\rangle$ may have entered state $|2\rangle$ or $|4\rangle$, so the path probability for $\bar{\mathcal{P}}$ can be written

$$\mathbb{P}(\bar{\mathcal{P}}) = \langle i| M |i\rangle \langle i| M \Big(|2\rangle \langle 2| + |4\rangle \langle 4| \Big) M |j\rangle \langle j| M |j\rangle \tag{3.4}$$

To simplify the act of summing up over large subsets of paths, we would like to factorise the bracket on the RHS of the above to write

$$\langle i| M |i\rangle \langle i| M \big(|2\rangle + |4\rangle \big) \big(\langle 2| + \langle 4| \big) M |j\rangle \langle j| M |j\rangle \tag{3.5}$$

but note that (3.5) gives rise to invalid paths such as

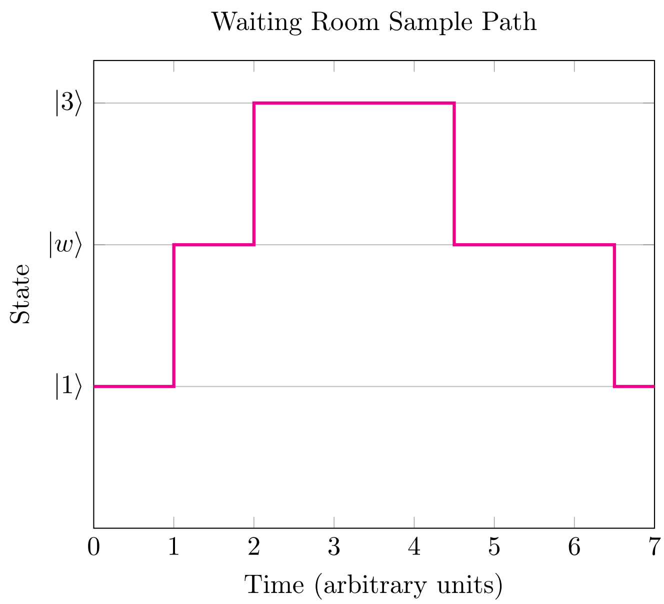

Figure 3.2: A typical path for the coarse grained system. It jumps from state $|1\rangle$ to $|w\rangle$ at $t = 1$ then to $|3\rangle$ at $t = 2$. It then climbs back down to $|w\rangle$ and eventually to $|1\rangle$. Since the system is much more likely to be in $|2\rangle$ if it has entered $|w\rangle$ from $|1\rangle$ (as it doest at time $t = 1$), the expected waiting time in this case is much smaller than the case where the system enters $|w\rangle$ from $|3\rangle$ (as it does at $t = 4.5$), in which case it is most likely to have entered state $|4\rangle$ in the granular description. It is clear from this picture that the forward and reverse paths have different statistics.

$$\langle i| M |i\rangle \langle i| M |2\rangle \langle 4| M |j\rangle \langle j| M |j\rangle .$$

This path is invalid because it ends the third time step in state $|2\rangle$ but begins the fourth time step in $|4\rangle$. This is not allowed since no time elapses between the end of one time step and the beginning of the next. To resolve this issue we shall use a 'gluing' scheme. Denoting by $\bar{z}$ the complex conjugate of z, We write

$$\mathbb{P}(\bar{\mathcal{P}}) = \frac{1}{2\pi} \int_{|z|\leq 1} d^2z \, \langle i| M |i\rangle \langle i| M \big(z\,|2\rangle + |4\rangle\,\bar{z}\big) \big(\langle 2|\,\bar{z} + \langle 4|\,z\big) M |j\rangle \langle j| M |j\rangle , \tag{3.6}$$

making use of the fact that

$$\int_{|z|\leq 1} d^2z \, z^2 = 0, \quad \int_{|z|\leq 1} d^2z \, |z|^2 = 2\pi. \tag{3.7}$$

If we now define

$$J_n := \frac{1}{(2\pi)^n} \int_{|z_1|\leq 1} d^2z_1 \int_{|z_1|\leq 1} d^2z_2 \ldots \int_{|z_n|\leq 1} d^2z_n \big(z_1\,|2\rangle + |4\rangle\,\bar{z}_1\big)\big(\langle 2|\,\bar{z}_1 + \langle 4|\,z_1\big) M \ldots \\ M\big(z_n\,|2\rangle + |4\rangle\,\bar{z}_n\big)\big(\langle 2|\,\bar{z}_n + \langle 4|\,z_n\big), \tag{3.8}$$

then the probability of any path can easily be written in terms of the J_n's. Considering again the path ω with history $\{(|1\rangle, n_1), (|w\rangle, n_2), (|3\rangle, n_3), (|w\rangle, n_4), (|1\rangle, n_5), \ldots (|3\rangle, n_N)\}$. We can now write

$$\mathbb{P}(\omega) = \langle 1| M |1\rangle^{n_1} \langle 1| M J_{n_2} M |3\rangle \langle 3| M |3\rangle^{n_3} \ldots \langle 1| M J_{n_{N-1}} M |3\rangle \langle 3| M |3\rangle^{n_N} \tag{3.9}$$

This is still not the form we will use to find the expectation value of the Kullback Leibler divergence, however the above expression does make clear a very important qualitative feature of the entropy production in the coarse grained waiting room system. Studying the RHS of Equation (3.9), we note that the only scalar terms which are not stable under time reversal are those of the form $\langle 1| MJ_{n_i}M |3\rangle$ and $\langle 3| MJ_{n_i}M |1\rangle$. Hence, writing ω^* for the time reversed path corresponding to ω, we have

$$\frac{\mathbb{P}(\omega)}{\mathbb{P}(\omega^*)} = \frac{\langle 1| MJ_{n_2}M |3\rangle}{\langle 3| MJ_{n_2}M |1\rangle} \frac{\langle 1| MJ_{n_4}M |3\rangle}{\langle 3| MJ_{n_4}M |1\rangle} \cdots \frac{\langle 1| MJ_{n_{N-1}}M |3\rangle}{\langle 3| MJ_{n_{N-1}}M |1\rangle}. \tag{3.10}$$

Taking the logarithm this is

$$\log\left(\frac{\mathbb{P}(\omega)}{\mathbb{P}(\omega^*)}\right) = \log\left(\frac{\langle 1| MJ_{n_2}M |3\rangle}{\langle 3| MJ_{n_2}M |1\rangle}\right) + \log\left(\frac{\langle 3| MJ_{n_4}M |1\rangle}{\langle 1| MJ_{n_4}M |3\rangle}\right) + \ldots \log\left(\frac{\langle 1| MJ_{n_{N-1}}M |3\rangle}{\langle 3| MJ_{n_{N-1}}M |1\rangle}\right). \tag{3.11}$$

Equation 3.11 says that the entropy produced along ω is precisely the entropy produced along the sections of ω where the system travels from $|1\rangle$ to $|3\rangle$ (or vice versa) through $|w\rangle$. In particular, no terms of the form $\langle 1| MJ_{n_i}M |1\rangle$ or $\langle 3| MJ_{n_i}M |3\rangle$ appear in the logarithm, which is to say that if an excursion away from state $|1\rangle$ (respectively $|3\rangle$) ends before visting state $|3\rangle$ (respectively $|1\rangle$) then *it will not contribute to the entropy production of the path.*

Another important implication is that the expectation $\mathbb{E}\left[\log\left(\frac{\mathbb{P}(\omega)}{\mathbb{P}(\omega^*)}\right)\right]$ breaks down into (relatively) simple sums. Letting Ω be the space of all paths for the coarse-grained system, We have

$$\begin{aligned}
\mathbb{E}\left[\log\left(\frac{\mathbb{P}(\omega)}{\mathbb{P}(\omega^*)}\right)\right] &= \sum_\Omega \mathbb{P}(\omega)\log\left(\frac{\mathbb{P}(\omega)}{\mathbb{P}(\omega^*)}\right) \\
&= \sum_\Omega \mathbb{P}(\omega)\log\left(\frac{\langle 1| MJ_{n_4}M |3\rangle}{\langle 3| MJ_{n_4}M |1\rangle}\right) + \sum_\Omega \mathbb{P}(\omega)\log\left(\frac{\langle 1| MJ_{n_2}M |3\rangle}{\langle 3| MJ_{n_2}M |1\rangle}\right) + \ldots \\
&\quad + \sum_\Omega \mathbb{P}(\omega)\log\left(\frac{\langle 1| MJ_{n_{N-1}}M |3\rangle}{\langle 3| MJ_{n_{N-1}}M |1\rangle}\right).
\end{aligned} \tag{3.12}$$

This decomposition motivates the splitting of paths into sections that travel from $|1\rangle$ to $|3\rangle$ and vice-versa. Such sections we shall call half cycles.

3.2.2 Half Cycles

Let us prepare the system in $|1\rangle$ and define the stopping times

$$\begin{aligned}
T_3 &= \inf\{n > 0 : |x_n\rangle = |3\rangle\}, \\
T_1 &= \inf\{n > 0 : |x_n\rangle = |1\rangle\}.
\end{aligned} \tag{3.13}$$

Certainly, $\mathbb{E}(T_3 \mid |x_0\rangle = |1\rangle) < \infty$. So, we may ask *what is the expected entropy production, $\mathcal{S}_{1\to 3}$, up to time T_3?*

The probability for any path from $t = 0$ to T_3 can be written

$$\mathbb{P}(\omega) = \langle x_0 = 1| m_1 |1\rangle \langle 1| MJ_{n_1}M |1\rangle \langle 1| m_2 |1\rangle \ldots \langle 1| m_N |1\rangle \langle 1| MJ_{n_N}M |3\rangle. \tag{3.14}$$

Here the m_i and the n_i are the duration, in time steps, of the i-th visit to $|1\rangle$ and $|w\rangle$ respectively. $\langle 1| m |1\rangle$ denotes the path segment

$$\langle 1| m |1\rangle = \underbrace{\langle 1| M |1\rangle \langle 1| M \ldots M |1\rangle}_{m \text{ steps}}. \tag{3.15}$$

Now, since T_1 is a stopping time of the granular process which is accessible in the coarse-grained setting, $m_i > 0$ and $n_i > 0$ are independent. Eqn. (3.14) is the probability of a path which takes N excursion from $|1\rangle$ before visiting $|3\rangle$. Let us define the state

flesh out this explanation

$$|w_1\rangle = \frac{\alpha}{\alpha+\gamma}|2\rangle + \frac{\gamma}{\alpha+\gamma}|4\rangle . \tag{3.16}$$

$|w_1\rangle$ is the state of the system immediately after a jump $|1\rangle \to |w\rangle$. Beginning from $|w_1\rangle$ the probability that the next jump is to $|3\rangle$ is given by

$$\begin{aligned} P(|w_1\rangle \to |3\rangle) &= \frac{\alpha}{\alpha+\gamma}P(|2\rangle \to |3\rangle) + \frac{\gamma}{\alpha+\gamma}P(|4\rangle \to |2\rangle) \\ &= \frac{\alpha^2}{(\alpha+\gamma)^2} + \frac{\gamma^2}{(\alpha+\gamma)(\beta+\gamma)} =: p \end{aligned} \tag{3.17}$$

Let $\mathcal{N}$ be the random variable indicating the number of excursions made from $|1\rangle$ before $|3\rangle$ is reached. Then $\mathcal{N}$ is a geometric r.v. such that

$$\begin{aligned} P(\mathcal{N} = N) &= p(1-p)^{N-1} \\ \mathbb{E}\mathcal{N} &= 1/p. \end{aligned} \tag{3.18}$$

Observe that

$$\sum_{n_i,m_i}^{\infty} \langle 1|\, m_1\, |1\rangle \langle 1|\, MJ_{n_1}M\, |1\rangle \dots \langle 1|\, m_N\, |1\rangle \langle 1|\, MJ_{n_N}M\, |3\rangle = P(\mathcal{N} = N). \tag{3.19}$$

Then we will have for the entropy production up to time T_3,

$$\begin{aligned} \mathcal{S}_{1\to 3} &= \sum_{\Omega} \mathbb{P}(\omega) \log\left(\frac{\langle 1|\, MJ_{n_N}M\, |3\rangle}{\langle 3|\, MJ_{n_N}M\, |1\rangle}\right) \\ &= \sum_{N=1}^{\infty} \sum_{m_i,n_i}^{\infty} \langle 1|\, m_1\, |1\rangle \langle 1|\, MJ_{n_1}M\, |1\rangle \dots \langle 1|\, m_N\, |1\rangle \langle 1|\, MJ_{n_N}M\, |3\rangle \log\left(\frac{\langle 1|\, MJ_{n_N}M\, |3\rangle}{\langle 3|\, MJ_{n_N}M\, |1\rangle}\right) \\ &= \left[\sum_{N=1}^{\infty} \sum_{m_i,n_i} \langle 1|\, m_i\, |1\rangle \dots \langle 1|\, m_{N-1}\, |1\rangle \langle 1|\, MJ_{n_{N-1}}M\, |1\rangle\right] \sum_{m_N,n_N}^{\infty} \langle 1|\, m_N\, |1\rangle \langle 1|\, MJ_{n_N}M\, |3\rangle \log\left(\frac{\langle 1|\, MJ_{n_N}M\, |3\rangle}{\langle 3|\, MJ_{n_N}M\, |1\rangle}\right) \\ &= \left(\sum_{N=1} P(\mathcal{N} \geq N)\right) \sum_{m,n}^{\infty} \langle 1|\, m\, |1\rangle \langle 1|\, MJ_nM\, |3\rangle \log\left(\frac{\langle 1|\, MJ_nM\, |3\rangle}{\langle 3|\, MJ_nM\, |1\rangle}\right) \\ &= \mathbb{E}\mathcal{N} \sum_{m,n}^{\infty} \langle 1|\, m\, |1\rangle \langle 1|\, MJ_nM\, |3\rangle \log\left(\frac{\langle 1|\, MJ_nM\, |3\rangle}{\langle 3|\, MJ_nM\, |1\rangle}\right) \\ &= \frac{1}{p} \sum_{m,n}^{\infty} \langle 1|\, m\, |1\rangle \langle 1|\, MJ_nM\, |3\rangle \log\left(\frac{\langle 1|\, MJ_nM\, |3\rangle}{\langle 3|\, MJ_nM\, |1\rangle}\right). \end{aligned} \tag{3.20}$$

Now, using the same methodology as in Section II, we obtain

$$\langle 1|\, m\, |1\rangle \langle 1|\, MJ_nM\, |3\rangle \xrightarrow{\tau\to 0} \mathrm{d}t_1\mathrm{d}t_2\alpha^2 e^{-(\alpha+\gamma)t_1}\left(e^{-(\alpha+\gamma)(t_2-t_1)} + \frac{\gamma^2}{\alpha^2}e^{-(\beta+\gamma)(t_2-t_1)}\right). \tag{3.21}$$

Moreover, using the definition of J_n we have

$$\langle 1| MJ_nM |3\rangle = \alpha\tau \langle 2| MJ_{n-1}M |3\rangle + \gamma\tau \langle 4| MJ_{n-1}M |3\rangle, \tag{3.22}$$

and similar for $\langle 3| MJ_nM |1\rangle$ such that in the continuous-time limit (3.20) becomes

$$\mathcal{S}_{1\to 3} = \frac{1}{p}\int_0^T \mathrm{d}t_2 \int_0^{t_2} \mathrm{d}t_1 \alpha^2 e^{-(\alpha+\gamma)t_1} \left(e^{-(\alpha+\gamma)(t_2-t_1)} + \frac{\gamma^2}{\alpha^2} e^{-(\beta+\gamma)(t_2-t_1)}\right) \log\left(\frac{\alpha^2}{\beta^2}\frac{e^{-(\alpha+\gamma)(t_2-t1)} + \gamma^2/\alpha^2 e^{-(\beta+\gamma)(t_2-t_1)}}{e^{-(\beta+\gamma)(t_2-t_1)} + \gamma^2/\beta^2 e^{-(\alpha+\gamma)(t_2-t_1)}}\right). \tag{3.23}$$

By symmetry, $\mathcal{S}_{3\to 1}$ is given by the integral

$$\mathcal{S}_{3\to 1} = \frac{1}{p^*}\int_0^T \mathrm{d}t_2 \int_0^{t_2} \mathrm{d}t_1 \beta^2 e^{-(\beta+\gamma)t_1} \left(e^{-(\beta+\gamma)(t_2-t_1)} + \frac{\gamma^2}{\beta^2} e^{-(\alpha+\gamma)(t_2-t_1)}\right) \log\left(\frac{\beta^2}{\alpha^2}\frac{e^{-(\beta+\gamma)(t_2-t1)} + \gamma^2/\beta^2 e^{-(\alpha+\gamma)(t_2-t_1)}}{e^{-(\alpha+\gamma)(t_2-t_1)} + \gamma^2/\alpha^2 e^{-(\beta+\gamma)(t_2-t_1)}}\right), \tag{3.24}$$

where

$$p^* = \frac{\beta^2}{(\beta+\gamma)^2} + \frac{\gamma^2}{(\beta+\gamma)(\alpha+\gamma)} \tag{3.25}$$

is the probability that a particular excursion from $|3\rangle$ visits $|1\rangle$ before returning to $|3\rangle$. The entropy production rate of the CG waiting room is given by

$$\dot{\mathcal{S}}^C = \frac{S_{1\to 3} + S_{3\to 1}}{\mathbb{E}_1 T_3 + \mathbb{E}_3 T_1}. \tag{3.26}$$

3.3 Mechanism of Entropy Production

The dynamic entropy production of a continuous-time semi-Markov chain along non-equilibrium trajectories, $\dot{\mathcal{S}}_{\text{dyn}}$ can be split into two contributions [3]

$$\dot{\mathcal{S}}_{\text{dyn}} = \dot{\mathcal{S}}_{\text{aff}} + \dot{\mathcal{S}}_{\text{wtd}}. \tag{3.27}$$

$\dot{\mathcal{S}}_{\text{aff}}$ results from what are called the cycle affinities of the system. If Γ is the incidence graph of a semi-Markov chain, and $\mathcal{C}$ is a closed loop on Γ, then we define the cycle affinity, $\mathcal{A}_{\mathcal{C}}$, as

$$\mathcal{A}_{\mathcal{C}} = \log \prod_{(i\to j)\in\mathcal{C}} \frac{a_{ij}}{a_{ji}}, \tag{3.28}$$

where a_{ij} is the transition rate for $i \to j$ and we are are using the convention that $\log\left(\frac{0}{0}\right) = 0$. By dynamic reversibility, $a_{ij} > 0$ implies $a_{ji} > 0$, so $\mathcal{A}_{\mathcal{C}}$ is well-defined. Letting $r_{\mathcal{C}}$ be the rate at which the cycle $\mathcal{C}$ is observed in the long-time limit, we have [1]

$$\dot{\mathcal{S}}_{\text{aff}} = \sum_{\mathcal{C}} r_{\mathcal{C}} \mathcal{A}_{\mathcal{C}} \tag{3.29}$$

$\dot{\mathcal{S}}_{\text{wtd}}$ arises from asymmetric waiting-time distributions in semi-Markov systems. These waiting-time distributions carry information about the history of the system. For example, as the CG waiting room system transitions from $|3\rangle$ to $|w\rangle$, it enters the combination state

$$|w_3\rangle = \frac{\beta}{\beta+\gamma}|4\rangle + \frac{\gamma}{\beta+\gamma}|2\rangle\,, \tag{3.30}$$

which is distinct from the state $|w_1\rangle$ defined in (3.16). The states $|w_1\rangle$ and $|w_3\rangle$ have different waiting-time distributions according to their time evolutions,

$$\begin{aligned}
|w_1(t)\rangle &= \frac{1}{\alpha e^{-(\alpha+\gamma)t}+\gamma e^{-(\beta+\gamma)t}}\left(\alpha e^{-(\alpha+\gamma)t}|2\rangle + \gamma e^{-(\beta+\gamma)t}|4\rangle\right)\\
|w_3(t)\rangle &= \frac{1}{\beta e^{-(\beta+\gamma)t}+\gamma e^{-(\alpha+\gamma)t}}\left(\gamma e^{-(\alpha+\gamma)t}|2\rangle + \beta e^{-(\beta+\gamma)}|4\rangle\right)
\end{aligned} \tag{3.31}$$

An observer viewing the process in reverse will observe state $|w_3\rangle$ when $|w_1\rangle$ occurs in the forward process. The distinct waiting-time distributions then give rise to entropy production which is captured in Eqns. (3.23) & (3.24). In fact, $\dot{\mathcal{S}}_{\text{wtd}}$ is the only contribution to the entropy production of the CG waiting room process. Cycles such as $\mathcal{C}_1 = |1\rangle \to |w\rangle \to |1\rangle$, which do not navigate the whole phase space, will not contribute to entropy production because they have zero cycle affinity. We have rates $\alpha+\gamma$ for the transition $|1\rangle \to |w\rangle$ and $\beta+\gamma$ for the transition $|w\rangle \to |1\rangle$, hence,

$$\mathcal{A}_{\mathcal{C}_1} = \log\left(\frac{\alpha+\gamma}{\beta+\gamma}\frac{\beta+\gamma}{\alpha+\gamma}\right) = 0. \tag{3.32}$$

and likewise for any other cycle that does not explore the entire phase space. A similar calculation will show that in fact *all* cycles of the CG waiting room system have zero affinity. The underlying physical intuition is that if a cycle includes an edge $(i \to j)$ then it necessarily includes the reversed edge $(j \to i)$. Hence, the forward and reverse trajectories include the same transitions, albeit in a different order. For the granular waiting room the situation is reversed.

In the granular waiting room system, there is no entropy production due to waiting-time distributions and

$$\dot{\mathcal{S}}_{\text{dyn}} = \dot{\mathcal{S}}_{\text{aff}}. \tag{3.33}$$

Consider a clock-wise cycle of the granular system shown in Figure 3.1a, beginning from $|1\rangle$ and making transitions

$$|1\rangle \to |2\rangle \xrightarrow{t_2} |3\rangle \xrightarrow{t_3} |4\rangle \xrightarrow{t_4} |1\rangle\,, \tag{3.34}$$

where t_i is the time spent occupying a state before transitioning away from it. The transition $|1\rangle \to |2\rangle$ occurs at $t=0$ and $|4\rangle \to |1\rangle$ at the final observation time $t=T$. In reverse time, the observed sequence of states is,

$$|1\rangle \to |4\rangle \xrightarrow{t_4} |3\rangle \xrightarrow{t_3} |2\rangle \xrightarrow{t_2} |1\rangle\,. \tag{3.35}$$

In reverse time, the observed sequence of transitions is different, but the time interval spent in each state remains unchanged. Moreover, the waiting-time distribution for each state is independent of the particle's history. For these reasons, we have

$$\dot{\mathcal{S}}_{\text{wtd}} = 0 \tag{3.36}$$

in the granular system.

Bibliography

[1] Jann van der Meer, Benjamin Ertel, and Udo Seifert. "Thermodynamic inference in partially accessible Markov networks: A unifying perspective from transition-based waiting time distributions". In: *arXiv preprint arXiv:2203.12020* (2022).

[2] Luca Cocconi et al. "Entropy production in exactly solvable systems". In: *Entropy* 22.11 (2020), p. 1252.

[3] Ignacio A Martı ́nez et al. "Inferring broken detailed balance in the absence of observable currents". In: *Nature communications* 10.1 (2019), pp. 1–10.

[4] Jos Uffink. "Boltzmann's Work in Statistical Physics". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Spring 2017. Metaphysics Research Lab, Stanford University, 2017.

[5] Lev Davidovich Landau and Evgenii Mikhailovich Lifshitz. *Statistical Physics: Volume 5*. Vol. 5. Elsevier, 2013.

[6] E. Schrodinger. *What is Life?: With Mind and Matter and Autobiographical Sketches*. Cambridge paperbacks. Cambridge University Press, 2012. ISBN: 9781107604667. URL: https://books.google.co.uk/books?id=hP9-WIEyv8cC.

[7] S. Blundell and K.M. Blundell. *Concepts in Thermal Physics*. Oxford University Press, 2008. URL: https://books.google.co.uk/books?id=SikbtAEACAAJ.

[8] Kenneth D Bailey. "Living systems theory and social entropy theory". eng. In: *Systems Research and Behavioral Science* 23.4 (2006), pp. 291–300. ISSN: 1099-1735.

[9] Carlo Cercignani. *Ludwig Boltzmann: The Man Who Trusted Atoms*. Oxford University Press UK, 1998.

[10] G. Parisi. *Statistical Field Theory*. Avalon Publishing, 1998. ISBN: 9780738200514. URL: https://books.google.co.uk/books?id=bivTswEACAAJ.

[11] William H. Cropper. "Rudolf Clausius and the road to entropy". In: *American Journal of Physics* 54.12 (1986), pp. 1068–1074. DOI: 10.1119/1.14740. eprint: https://doi.org/10.1119/1.14740. URL: https://doi.org/10.1119/1.14740.

[12] J. Schnakenberg. "Network theory of microscopic and macroscopic behavior of master equation systems". In: *Rev. Mod. Phys.* 48 (4 Oct. 1976), pp. 571–585. DOI: 10.1103/RevModPhys.48.571. URL: https://link.aps.org/doi/10.1103/RevModPhys.48.571.

[13] C. E. Shannon. "A mathematical theory of communication". In: *The Bell System Technical Journal* 27.3 (1948), pp. 379–423. DOI: 10.1002/j.1538-7305.1948.tb01338.x.

[14] Rudolf Clausius. "X. On a modified form of the second fundamental theorem in the mechanical theory of heat". In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 12.77 (1856), pp. 81–98.

[15] William Thomson. "On the dynamical theory of heat: with a numerical result deduced from Mr Joule's equivalent of a thermal unit, and M. Regnault's observations on steam". In: *Proceedings of the Royal Society of Edinburgh* 3 (1851), pp. 48–52.